Executive Functioning and the Emergence of Conversational Understanding.

Clair Pond

A thesis submitted for the degree of Doctor of Philosophy

Department of Psychology

September 2009

Abstract

Conversational understanding (CU) refers to the ability to accurately interpret the meaning of discourse-embedded language. Siegal and Surian (2007) have suggested that the developmental expression of CU often reflects the unmasking of a cognitive competence obscured by processing biases associated with a shortage of computational resources. This thesis sought to investigate the relationship between developing computational resources (conceptualized in terms of executive functioning (EF)) and CU in 3- to 6-year-old children. According to Miyake et al. (2000), EF can be considered to reflect three main skills – inhibition and updating of mental representations, and mental set shifting. This thesis investigated the independent contribution of these three aspects of EF to CU.

Experiments 1 and 2 investigated whether a relationship exists between the three components of EF and a narrow, scalar implicature (SI) measure of CU. Updating ability was found to demonstrate significant relations with SI scores. Experiment 3 revealed that shifting ability is significantly related to a broad, conversational violations task (CVT) measure of CU. Experiment 4 used training to significantly enhance children's shifting ability. However, this did not lead to a corresponding improvement in CVT performance.

Experiment 5 investigated the relationship between EF and CU in bilingual children. Bilinguals demonstrated a non-significant shifting advantage relative to monolingual controls, but did not demonstrate a CU superiority. Experiment 6 provided a more detailed analysis of the relationship between EF and CU by presenting monolingual children with both the narrow and broad CU measures previously presented, along with the updating, inhibitory and shifting EF measures used in Experiments 2-5 and three new measures of these EF components. All three EF components were found to demonstrate relations with CVT performance after the effects of age and verbal intelligence had been controlled for.

The consistent demonstration of significant relationships between EF components and CU revealed in this thesis provides partial support for Siegal and Surian's masked competence model of the developmental expression of CU.

Acknowledgements

I would like to thank my primary supervisor, Professor Michael Siegal, for providing direction and criticism where required and helping me to expand my understanding of the field focused upon in this thesis. I am also grateful for the support provided by my second supervisor, Professor Richard Eiser.

My family and friends also deserve mention for their ceaseless patience and encouragement.

I would also like to express gratitude to the ESRC who provided the funding which enabled me to conduct the research reported in this thesis and all the children who took part in my studies along with their teachers, schools and parents.

Table of Contents

Table of Tables

Table of Figures

Chapter 1: Conversational understanding in young children

1.1 Focus of this thesis

"Conversational understanding" (CU) refers to the ability to accurately interpret the meaning of discourse-embedded language. It denotes the ability to appreciate the functional use of language – the way in which it is actually adapted to communicate in the world. This can be contrasted with a formalistic appreciation of language which is concerned with understanding the grammatical rules constraining the way linguistic elements are combined.

When communicating, speakers adjust the language they use to accommodate the linguistic, physical and general knowledge context in which it is given (Clark, & Carlson, 1981; Clark, & Marshall, 1981). For example, if a speaker has just spoken to their conversational partner about someone called Jack, they might continue to talk about Jack using the referring expression "he". The replacement of the expression "Jack" with "he" would be based on the assumption that the partner would take the preceding linguistic context into account and thus identify "he" as Jack rather than some other male. Similarly, a speaker conversing with someone who has just missed hitting the ball during a game of tennis might tell that person to "try harder". The speaker would assume that the listener would take the physical tennis-game context into account and so realise that they should try harder to hit the ball rather than try harder to achieve some other aim such as learning another language or writing poetry. Further, a speaker might tell their partner that they had frostbite and assume that the partner would take the general knowledge that frostbite only occurs when someone is exposed to very cold conditions, into account. The partner might thus be expected to realise that the speaker had recently been somewhere very cold.

As illustrated above, CU requires integration of information provided by the logical/literal meaning of the words used in a message, with contextual information. Integration of information from the two sources (logic and context) enables more accurate inference of the intended meaning behind the communication.

Although most adults would seem to have mastered the ability to accurately interpret language in context, studies have consistently reported that young children find integration of information from the domains of language used and the context used in, difficult. The purpose of this thesis is to examine the relation between CU and aspects of executive functioning (EF) involving the ability to focus, inhibit and update information, the absence of which may "mask" children's competence on many cognitive tasks.

1.2 Conversational understanding in childhood

Some investigators have observed that early on, children demonstrate the presence of a lexical bias evidenced by a concentrated focus on the words used in a message. Friend and Bryant (2000) presented English-speaking 4-, 7- and 10-year-olds with communications in which the words used indicated positive affect: e.g. "You're doing a great job" but the vocal expression used to transmit the message indicated negative affect: e.g. the message was delivered using an angry tone of voice. Friend and Bryant then asked the children to indicate whether they thought the speaker was happy or angry by pointing to facial stimuli depicting happy or angry expressions. 4- and 7-year-old's ratings of speaker emotions were found to more closely reflect the affect expressed lexically than the affect expressed vocally, whereas the ratings of the 10-year-olds more closely reflected the vocal expressions. The younger children appeared to concentrate on the words used to the neglect of vocal cues to affect.

Further evidence of a bias to focus on the words used in communication was found by Lee, Torrance, and Olson, (2001) in a nursery rhyme context. English-speaking 3- to 7-year-olds were presented with excerpts from nursery rhymes e.g. "Hickory dickory dock, the mouse ran up the clock", and asked to judge Teddy's description of what was happening in the rhyme. Although Teddy was told he did not have to use the same words as the rhyme, children below 6-years-old judged descriptions he gave using true paraphrases such as "the mouse raced up the clock", negatively, rewarding only those descriptions which used the same words as the rhyme i.e. "the mouse ran up the clock". Young children seemed unable to detach from the specific words used in the nursery rhyme context.

However, it would not seem that children are simply bound to a lexical focus. Olson and Hildyard (1981) presented English-speaking 4- to 8-year-olds with a story depicting a communication between characters: children were told about a boy who was not happy with the amount of popcorn he had, and who consequently said to his friend "You've got more than me". Olson and Hildyard found that when they asked the children what the boy in the story had said, 4- to 5-year-olds tended to respond with the boy's intended meaning i.e. give me some of your popcorn, rather than his actual statement. This contrasts with the 7- to 8-year-olds who reported both the boy's statement and his intended meaning.

Robinson, Goelman and Olson (1983) presented children with a referential communication game in which they were presented with ambiguous instructions and in which their ambiguous instructions were acted upon. An example of an ambiguous instruction was the command "Pick up the red flower" when both a big red flower and a small red flower were available. The speaker and the listener had independent sets of the same items to choose from and would both choose an item following the instruction and compare selections. Robinson, Goelman and Olson found that when questioned as to what had actually been said following a mismatch between the selections of the speaker and listener, in which the speaker had picked up the big red flower, English-speaking 5-year-olds correctly rejected incorrect paraphrases such as "Did you/I say to pick up the blue flower?" and correctly accepted verbatim repetitions of ambiguous sentences such as "Did you/I say to pick up the red flower?" However the children were nevertheless significantly less likely to correctly reject disambiguated interpretations of ambiguous instructions such as "Did you/I say to pick up the big red flower?" Knowledge of speaker's intentions appeared to influence children's representations of messages.

Beal and Flavell (1984) reported a similar finding. They presented English-speaking 6- to 8-year-olds with ambiguous messages purporting to identify the pictures that a character had selected from a variety presented. For example, children might be presented with pictures of a blue circle, a blue triangle and a red triangle and given the identifying message "The blue one". The experimenter then indicated to half of the children the picture that the character had intended them to identify. The other half remained ignorant. All children were then asked whether the message could identify each of the pictures shown. Beal and Flavell found that children who knew the intended referent were less likely to acknowledge that the ambiguous message could identify multiple referents, favouring the intended referent. Children with knowledge of the speaker's intention seemed unable to focus on the literal message itself.

Further, Lee, Torrance and Olson (2001) found that presenting an utterance in a narrative/story context, e.g. Big Bird and Snuffy go to Maria's for lunch. Big Bird says "I want some food", made it difficult for children to children to focus on the exact words used in a message. Lee, Torrance and Olson asked English-speaking 3- to 7-year-olds to judge Teddy's description of the utterance in the narrative and instructed Teddy to repeat exactly what the character had said, using the same words. Nevertheless, children below 6-years-old rewarded Teddy for true paraphrases such as "I want something to eat" which used different words from the original utterance. Young children seemed unable to focus on the specific words used in the narrative context.

In contrast to the lexical bias studies, the reports above indicate that children are able to consider factors external to the utterance itself. Nevertheless, as with the lexical studies, these investigations again indicate the tendency of children to demonstrate an inappropriate focus when processing communication: in these cases towards contextually determined intentions at the expense of lexical considerations.

However, other studies have reported a lack of bias. Dews et al., (1996) presented English-speaking 5- and 6-year-olds with clips from popular cartoons showing characters delivering ironic criticisms in which the logical/literal meaning of their communication was positive but the intended meaning was negative. For example, a character described someone who refused to aid them as "helpful". Dews et al then presented the children with two possible meanings for each ironic criticism, one positive and one negative, and asked the children to select which one the character had intended. 5-year-olds were found to perform at chance levels on this task indicating they were neither biased towards the literal message content nor the conflicting situational cues to ironic meaning.

More recently, Filippova and Astington (2008) provided an in-depth investigation of the irony appreciation of 5-, 7- and 9-year-olds. Filippova and Astington used narratives accompanied by pictures to present ironic comments (e.g. a story depicted a character who kept missing goals in a football match and who was subsequently described as "a great scorer.") Children's ability to understand different aspects of irony was also investigated through the use of questions probing the meaning of the ironic utterance given, the belief of the speaker giving the utterance, the communicative intention of the speaker and the speaker's attitude. Filippova and Astington found that children demonstrated appreciation of these different aspects of irony in a set order: the order in which they are mentioned above, but that appreciation of all aspects was not revealed until 7-years-of-age, indicating that children do not clearly appreciate irony until 7 years.

It appears that young children are not bound to demonstrate an inappropriate focus when interpreting communication. However, all of the above studies appear to indicate that young children are unable to consider the logical elements of a message in light of contextual factors.

1.3 The Gricean framework and the Relevance approach

Grice (1975) proposed that appreciation of certain conversational rules/maxims relating the message to its communicative context, provided the foundation for CU. He depicted communication as a cooperative exchange and argued that the speaker could be expected to adhere, where possible, to maxims governing the informativeness (Maxims of First and Second Quantity), truthfulness (Maxim of Quality) and relevance (Maxim of Relation) of their contribution. That is to say that the speaker could be counted upon to make their contribution as informative as the listener required but not more informative than necessary, to be truthful and to make their contribution relevant to the context in which it was given. Appreciation of these maxims was expected to prompt the listener to draw information from both the message itself and the context it was presented in. This was thought to enable assessment of the communication in terms of its environment, and so allow accurate inference of the speaker's intended meaning. Apparent maxim violations produced within this cooperatively construed exchange, were thought to provoke listener inference of implied meanings. These meanings are not apparent from sole focus on either the message, or the context of the communication, but emerge from consideration of the message in light of its context and the Gricean maxims relating the two.

For example, a speaker's production of the message "You are being really helpful", in a context in which their conversational partner is not aiding them in any way, can be considered with regard to the Maxim of Quality. The message appears to be in direct conflict with the maxim concerning the truthfulness of communication, as the physical context indicates that the partner is not being at all helpful. Neither sole consideration of the message, nor of the absence of aid context, tells the partner that the speaker wants to be helped. However consideration of the message in light of the context and Maxim of Quality emphasises the fact that the partner is not helping. The partner's perception that the speaker is trying to draw attention to the fact that they are not helping leads the partner to infer that the speaker is wanting their help.

The direction of the message "You might want to look where you're going" to an adult who has just bumped into the speaker, could be considered with regard to the Maxim of Second Quantity. The message appears to be in direct conflict with the maxim demanding that the speaker refrain from providing too much information, as the physical context indicates that the listener is an adult, who as such would be expected to already be aware of the advised instruction. Neither sole consideration of the message, nor of the bumping context, tells us about the speaker's opinion about the other adult. However consideration of the message in light of the context and Maxim of Second Quantity indicates that speaker thinks the other adult is acting as if he does not possess the knowledge expected of him, and so leads the other listener to infer that the speaker has a poor opinion of them.

A speaker's production of the message "People" in a context in which they have been asked who they met at a function, can be considered with regard to the Maxim of First Quantity. The message appears to be in direct conflict with the maxim demanding that the speaker give as much information as is necessary as the preceding linguistic context: "Who did you meet?" indicates that the enquirer was already aware that people were met. The message is thus unlikely to have provided the enquirer with the information they were seeking. Neither sole consideration of the reply message, nor of the particular question context, tells the enquirer that the speaker wants to be unhelpful. However consideration of the message in light of the context and Maxim of First Quantity indicates that the speaker is trying to avoid answering the question, and so leads the partner to infer that the speaker is deliberately trying to be unhelpful.

The delivery of the message "Her mother is nice" in a context in which the speaker has been asked what they think of someone called Chloe, can be considered with regard to the Maxim of Relevance. The message appears to be in direct conflict with the maxim requiring that communication should be relevant to the current context, as a comment concerning Chloe's mother is not directly relevant to the linguistic context which seeks an opinion of Chloe. Neither sole consideration of the message, nor of the particular question context, informs the enquirer that the speaker does not think Chloe is nice. However consideration of the message in light of the context and Maxim of Relevance emphasises the fact that the speaker doesn't feel able to say talk about Chloe in a similar manner, and so leads the partner to infer that the speaker does not like Chloe.

In addition to his main maxims, Grice (1975) also speculated about the need to involve others such as a Maxim of Politeness requiring speakers to make their communication respectful. Consider production of this message "The food was horrible!" in a context in which the listener has made the speaker's dinner. The message appears to be in direct conflict with the maxim requiring that communication should be respectful, as it is disrespectful to refer to tell someone that food is horrible in a physical context in which the person listening has been responsible for preparing that food. Neither sole consideration of the message, nor of the physical context, tells the listener that the speaker is trying to be rude to them. However consideration of the message in light of the context and Maxim of Politeness emphasises the fact that the speaker does not want to be polite, and so leads the listener to infer that the speaker is trying to be rude to them.

Sperber and Wilson's (1986) Relevance Theory account of CU, focuses on the maxim Grice conceived as restricting the relevance of the speaker's communication to the context in which it is given. Relevance theory maintains that all other conversational maxims ultimately derive from this "Relevance" maxim. According to the Relevance account, the listener determines the speaker's intended meaning by searching via inference for the relevance of the utterance meaning to the context in which it is presented. Search criteria are defined in terms of the relative processing effort involved, with effort calculated in terms of the cognitive effects gained i.e. knowledge, given the processing cost borne i.e. mental energy expended searching. Sperber and Wilson (1986) suggest that the search starts from the basis of an initial, logical utterance interpretation, and that the extent of processing resources (i.e.

working memory) available, determines the relative processing effort deployed. This is thought to establish the degree of search motivation experienced, ascertaining the level of contextual enrichment achieved through successively more extensive searches for cognitive effects/further information. According to Sperber and Wilson's account, interpretation of the phrase "Can you pass the salt?" in the dinner-time context, will start with an initial logical reading: Do I possess the physical ability to move the salt? However the possession of sufficient processing resources will enable inference concerned with the relevance of the question in the context it's given: the speaker wants to know because they would like to have the salt and so would like me to pass it to them if I am able. Such inferences enable enrichment of the message interpretation.

Studies with adults provide support for the Relevance account of CU. Bott and Noveck (2004) found that for French speakers, the production of requested, contextually enriched message interpretations took longer and was less reliable than the production of requested decontextualised interpretations. Moreover, they found that limiting the amount of time in which interpretations could be made, increased the proportion of decontextualised interpretations made.

Noveck and Posada (2003) reported analysis of ERP data for French adults in line with this in their investigation of immediate reactions provoked by underinformative sentences such as "Some elephants have trunks" (English translation). Such sentences are true in their decontextualised/ logical form but give rise to false contextually enriched/CU interpretations e.g. that not all elephants have trunks. The enriched interpretation emerges out of the conversational convention that the use of weak terms such as "some" implies the inapplicability of stronger, related terms, such as "all" in this case. This conversational technique is known as a scalar implicature (SI) and can be explained in relation to Grice's Maxim of First Quantity which demands that speakers say as much as is necessary. The listener can use this maxim to infer that if a stronger, more informative term applied, the speaker would have used it. It can therefore be deduced that the speaker does not consider that a stronger term applies.

As well as presenting adult participants with underinformative SIs, Noveck and Posada (2003) also presented SI sentences which were true on both decontextualised and contextually enriched levels such as "Some houses have bricks" (English translation) and SI sentences which were false on both decontextualised and contextually enriched levels such as "Some crows have radios" (English translation). In their analysis of ERP data, Noveck and Posada focused on the N400 wave which typically peaks approximately 400 msec after the presentation of contextually inappropriate words. They found that not only were the N400 waves for the underinformative sentences flatter than those produced for the sentences which were true or false on both decontextualised and contextualised levels, but that this was the case whether or not participants produced contextually enriched interpretations of the underinformative sentences, indicated by whether they agreed or disagreed with the underinformative sentences. This suggests that the weak SI term was not processed as inappropriate immediately, but only became recognised as such later on. Therefore, it would seem that the inference that the stronger related term "all" did not apply/contextualised enrichment, was a time-consuming procedure. Indeed in correspondence, the reaction time data that Noveck and Posada collected revealed that the processing of contextually enriched message interpretations took longer than the processing of decontextualised interpretations. The time demanding nature of arriving at contextually enriched interpretations, supports the suggestion that effortful processing is involved in deriving such meanings, rather than that they become directly accessible following a conceptual change.

However, even stronger evidence in favour of the effortful, Relevance account is provided by a report that depleting the amount of available processing resources decreased the proportion of contextually enriched message interpretations produced. De Neys and Schaeken (2007) gave Dutch-speaking participants a spatial storage task: the Dot Memory task, which required memorization of patterns of dots within a 2D matrix to enable later pattern replication. De Neys and Schaeken found that when working memory resources were engaged in this storage task, participants gave less contextually enriched interpretations and provided more decontextualised readings.

1.4 Cognitive development and the conceptual change approach

In line with Relevance Theory, work with adults appears to support the proposal that CU reflects the availability of processing resources (working memory). It thus seems plausible that the emergence of CU reflects the maturation of processing abilities. However, early research with young children seemed to support a conceptual change model of CU.

Jean Piaget advanced a theory that cognitive development is underpinned by a number of processing revolutions in which the nature of the mechanisms involved in cognition drastically alters as a result of conceptual change. His ideas led the way for much conceptual change research, a substantial amount of which provided apparent support for such an account. Piaget (1975) proposed that cognitive development could be divided into four main stages. He suggested that a dominant factor characterised each stage: understanding through action, understanding through perception, logical thought restricted to reality and logical thought unbound by reality. Piaget held that transition from one stage to another occurred in a set sequence, and resulted from a dramatic restructuring of thoughts/concepts of the world: a conceptual change.

Studies were conducted to investigate whether children did indeed demonstrate dominant patterns of thought at various stages in development. A not B error research seemed to indicate that young infants failed to conceive of objects existing independently of their own behaviour, consistent with the view that their understanding was based on their action on the world (e.g. Butterworth, 1977; Landers, 1971; Piaget, 1954). The traditional A not B error paradigm involves infants observing an object being hidden in Location A multiple times, and being allowed to recover the object each time, before observing the object being hidden in Location B. Below around 11 months, infants typically search in Location A for the object in Location B. The standard Piagetian interpretation was that the infants conceptualised the object in terms of their actions at Location A. Older infants however, typically search correctly in Location B, which was taken to indicate that they had restructured their thoughts and now conceived of objects independently of their actions at past locations.

Conservation research appeared to indicate that superficial factors constrained the way young children conceptualised the world, consistent with the view that they understood the world in terms of their perception (Dodwell, 1960; Elkind, 1961; Hood, 1962; Piaget, 1952; Piaget & Inhelder, 1969; Smedslund, 1964). One variety of task presents children with a set amount of substance e.g. water in a jar, buttons set out in a row, whose presentation is subsequently altered: the water would be poured into a taller, narrower jar, the buttons would be spread out in the row. Below around 7 or 8 years of age, children typically declare that there is more water in the taller, narrower jar than in the original jar despite witnessing that only the water from the original jar is poured into the taller jar, and that there are more buttons in the spaced out row than in the unspaced row despite having witnessing the experimenter merely spreading the buttons in the spaced out row. Older children however, typically declare correctly, that the amount of water and number of buttons is unaffected by their altered presentation. This was taken to indicate that the children had restructured their thoughts so that they were no longer completely bound by perception.

In line with Piaget's account of cognitive development, Beal and Flavell (1984) and Robinson and Whittaker (1986) put forward a conceptual change account of CU. Young children's apparent inability to consider the logical elements of a message in light of contextual factors was used to support the suggestion that young children fail to differentiate between that which is said: the logical/literal meaning (LM) and that which is meant: the intended meaning (IM). It was argued that children conceived of LM as equivalent to IM, with the consequence that if they were provided with one of these types of meanings, they thought it a reflection, and so an acceptable expression, of the other. Thus it could be argued children who had been made aware of the intended referent of the speaker's ambiguous instructions in Beal and Flavell (1984) consequently conceptualised the instructions in terms of this intended referent, and so could not conceive of the instructions as pertaining to identify any other referent. Similarly, it could be argued that the 4- and 7-year-olds who more heavily relied upon the words the speaker used than the tone of voice in which the words were produced to identify the speaker's emotion in Friend and Bryant, (2000), were conceptualising the intent to express these emotions with the words used, and so could not appreciate the expression of any other emotions. Apparent inappropriate preference for one over the other could be explained in terms of the comparative explicitness of the presentation of the two meanings within a particular study. Grice (1975) had argued that the use of conversational maxims hinged on the ability to apply the maxims to LM to draw inference about IM. Since this would be an impossible feat for children failing to distinguish the two concepts of meaning, adherence to the conceptual change account entailed that CU emerged as a result of a radical shift in thinking, one that resulted in the conceptual differentiation of LM and IM.

Robinson and Whittaker (1986) noted the similarity between apparent difficulties with the LM-IM distinction and problems reported with appearance-reality (AR) understanding. Comprehension of the AR distinction indicates an appreciation that identity is not determined by what something looks like just as the intended meaning of communication is not determined by the words used to communicate. Furthermore, mastery of the AR distinction indicates acceptance that it is quite possible for something to look like one thing, but actually be something else, knowledge vital to the development of accurate means of identification. This can be seen as a parallel to the mastery of the LM-IM distinction which enables appreciation that it is acceptable for communication to have a certain LM, but actually express a different IM, knowledge central to the development of CU.

Pioneered by John Flavell, traditional tasks assessing appreciation of the AR distinction present participants with deceptive objects (such as a sponge that looks like a rock). Flavell (1988) noted that in these conventional tasks, prior to object presentation, participants would be given brief instruction on the meaning of the AR distinction and explanation of the terms used to express the distinction in the task questions: "looks like" relating to appearance and "really and truly" relating to reality. A deceptive object would then be presented and the participant asked questions in a counter-balanced order concerning the object's appearance: "When you look at this with your eyes right now, does it look like a rock or does it look like a sponge?" and its identity: "What is this really and truly – is it really and truly a sponge or is it really and truly a rock?" On such tasks 4-year-olds tend to succeed whereas 3-year-olds typically demonstrate considerable difficulty, a behaviour differentiation which Flavell et al. (1986) attributed to the emergence of a conceptual change: the appreciation of representational mental states. This capacity was thought to enable children to differentiate between the nature of an object/reality and their perceptual representation of the object/object appearance.

Robinson and Whittaker (1986) suggested that the early difficulty observed with the LM-IM distinction, arose from trouble conceptualising the nature of the relationship between the LM and IM elements. Robinson and Whittaker (1986) argued that a message needed to be understood "as a stimulus to be interpreted in its own right and also as a representation" (p43) of an IM, providing only indirect access, the accuracy of which could be variable. They likened such a relationship to that of the partnership of appearance and reality. Portrayed in such a way, appearance would be conceived both as an independent element, and as a representation of reality (enabling indirect and thus potentially inadequate access). Difficulties in appreciating the AR distinction could thus be understood in terms of trouble conceptualising the differentiated relationship between appearance and reality, and the acquisition of competency in this area characterised in terms of a conceptual shift reflecting appreciation of this distinction.

However, evidence has been produced more recently which suggests that poor performance on AR assessments cannot be attributed to difficulty in conceptual differentiation, but seems instead representative of the child's developing acquisition of discourse. Gauvain and Greene (1994) used a modified AR task, in which children were presented with standard test questions, but were additionally permitted to express identity knowledge/reality appreciation through illustration of item function in a show-and-tell task. For example children were presented with a pencil shaped like a toothbrush: deceptive item, alongside a real toothbrush: true item, and questioned as to what they thought each item looked like and what they thought each item really was. They were also asked to describe the function of each item. Children were then presented with suitable props to support function demonstration: a piece of blank paper for writing and a doll with exposed teeth for teeth brushing, and asked to show what each item could be used for. Gauvain and Greene reported that English-speaking children as young as 2 years, who demonstrated poor performance on standard AR task questions: more specifically problems were encountered with object identity/reality questions, displayed improved performance in the modified task. Having revealed appreciation of appearance during the standard AR task questions, they were able to express a differentiated conception of reality during the function demonstration phase.

Sapp, Lee, and Muir, (2000) produced concordant findings. However they asked children to demonstrate both the function and the appearance of deceptive objects to indicate that children had similar representations of both properties. In the Gauvain and Greene (1994) study, knowledge of the properties of appearance and reality was ascertained through different means: children were only asked to demonstrate functional properties relating to identity, knowledge of appearance was assessed via verbal response. Sapp, Lee and Muir presented English-speaking children with an array of objects including a deceptive object such as a candle which looked like a crayon and asked the children to pass them objects which looked like particular items. The initial request was for a non-deceptive item, but following this, the deceptive item was requested e.g. "I want to take a picture of Teddy with something that looks like a crayon. Can you help me?" Following this children were asked to pass items which served particular functions. Again, the initial request was for a non-deceptive item, but following this, the deceptive item was requested e.g. "I want a candle to put on Bart's birthday cake. Can you help me?" Sapp, Lee and Muir found that 3-year-olds who were performing poorly on standard AR tasks performed competently on the demonstration trials.

Similarly, Rice et al. (1997) found that adaptations to task methodology improved performance. In one condition AR questions were posed in the context of a deceptive scenario: tricking the other experimenter, to emphasise the role of representational mental states in object perception. In another condition, a deceptive object such as a sponge that looked like a rock was presented alongside an exemplar item of appearance e.g. a rock and an exemplar item of reality e.g. a sponge, to reduce the need for independent mental representation of the different dimensions. Both manipulations were found to have a positive effect on the production of correct AR task responses by English-speaking 3-year-olds.

Deak, Ray, and Brenneman (2003) noted the similarity of response patterns for questions in a standard AR task to patterns for questions in a control task presented within a similar discourse structure but not focused upon the AR distinction. English-speaking 3-, 4- and 5-year old children were presented with a traditional AR task in which they examined a deceptive object and answered the standard AR questions "Does it look like a rock or does it look like a sponge?" and "Is this really and truly a sponge or really and truly a rock?" However, a control task was also delivered, in which children viewed pictures of animals wearing or holding familiar items e.g. a monkey was shown holding a cookie, and asked questions in the following format: "Does it look like a monkey or does it look like cookies?" "Does it have a monkey or does it have cookies?" Deak, Ray, and Brenneman found that children's tendency to perseverate when responding to AR task questions i.e. answer "rock" to both appearance and reality questions, could be predicted from their tendency to peserverate when responding to control task questions, along with their vocabulary level. This prompted the suggestion that the poor performance of young children on standard AR tasks might actually be reflecting their inability to recognise that successive forced choice questions providing identical response options require independent evaluation. Thus the findings of Deak, Ray, and Brenneman provide support for the suggestion that the standard AR task might be more representative of the ability to identify speaker intentions than knowledge of an AR distinction. This mirrors claims made across the developmental literature.

In such disparate areas of proficiency as number conservation, false-belief appreciation and cosmology, investigators have successfully shown that adapting tasks to make the experimenter's intentions more explicit, enhances children's performance on a range of cognitive developmental tests. For example, by refraining from question repetition (Rose & Blank, 1974) and using a "naughty" teddy to transform task materials (McGarrigle & Donaldson, 1974), investigators have managed to substantially improve performance in Piagetian number-conservation tasks, though age differences still persist. It has been argued that such task manipulations prevent children from introducing mistaken interpretations concerning the experimenter's question intentions. For example, a child might decide that they are being asked the same question a second time because their first response to the question had been wrong, indicating that they should now change their response, or they might think that the question was being asked a second time because something <u>had</u> now changed, so conclude again that they should alter their response.

In a like manner, performance on tests gauging false-belief understanding has been found to improve following the incorporation of a minor modification to task presentation. False-belief tasks assess appreciation that someone can possess a mental representation of a situation (a belief) which conflicts with the reality of the circumstances. In an "unexpected-locations" false-belief task, a character places an object in one location (A) and is unaware of its subsequent transfer to another location (B). They should thus hold a false belief regarding the whereabouts of the object, which should be manifest in a search for the object in its original (and now vacated) location. Siegal and Beattie (1991), and more recently Yazdi et al., (2006), have demonstrated that altering the standard question in this task, from "Where will [*the character*] look for ......?" to "Where will [*the character*] look first for ....?", substantially improves children's performance. It is thought that the "look first" instruction clarifies the experimenter's IM in the task by denying the interpretation "Where should she look to find .....?" The inability of young children to derive the intended interpretation in the original task can be considered attributable to a failure to recognise the Gricean dictum that speakers are obliged to not say more than is necessary. Using the Gricean maxim to relate the experimenter's instructions to the task context should result in an emphasis of the character's ignorance and so direct interpretation towards the restricted look first reading. However, the response of young children appears to be more context-focused, leading them to interpret the instructions in terms of where the character should look to find the object.

More explicit framing of the IM behind questions used to assess children's cosmological knowledge has also been found to enhance performance. Siegal, Butterworth, and Newcombe, (2004) investigated the knowledge that children possessed concerning the shape of the earth and the day-night cycle. In their second experiment they presented English speaking 4- and 5-year-olds with two sets of questioning formats, one of which presented children with generative questions such as "What is the shape of the earth? Where is the sun at night?" and the other of which delivered forced-choice questions such as "Is the world round or flat? When the sun shines on this part of the earth, is it day or night on the other part?" Siegal, Butterworth, and Newcombe found that reducing the potential ambiguity surrounding the desired response promoted the demonstration of accurate knowledge.

In line with the above findings highlighting the impact of discourse structure on the task responses obtained from children, Fritzley and Lee (2003) report the unsuitability of yes-no questions such as "Is this red?" in reference to a red cup, for children below the age of 4 years, based on data from English-speaking 3-, 4- and 5-year-olds. 4- and 5-year-olds failed to demonstrate any response bias to such questions when they were found to be comprehensible i.e. when they did not incorporate nonsense words, although a negative response bias was demonstrated for incomprehensible questions such as "Is this yint?" However, 2-year-olds were found to have an affirmation bias (tendency to respond "yes") to both comprehensible and incomprehensible questions and 3-year-olds were shown to be inconsistent and often biased in their response frame, varying the bias when demonstrated, between affirmation and negation.

Additionally, a recent though controversial analysis has claimed that even infants' performance on A-not B tasks is powerfully influenced by pragmatic factors (Topál, et al., 2008). It would seem that performance on cognitive developmental measures is to some substantial extent reflecting children's ability to determine IM rather than the conceptual deficits the tasks are seeking to assess.

If, as the evidence suggests, developmental tasks are misrepresenting conceptual competence, and children don't in fact experience difficulty distinguishing appearance from reality, it is plausible that they are actually able to conceive of a message as a stimulus to be interpreted in its own right and also as a representation of the speaker's IM. This conflicts with Robinson and Whittaker's (1986) proposal that children possess confused representations in which LM and IM are indistinct.

1.5 Recent directions in conversational understanding research

1.5.1 Explicit vs. implicit conversational understanding

More recent research into the emergence of CU has indicated that young children in fact do possess early LM-IM differentiation. However, appreciation of this distinction seems to be restricted with regard to the contexts and response modalities through which it's demonstrated, suggesting that the understanding is less likely to be the potentially conscious, explicit appreciation of adults, but rather, an unconscious, implicit knowledge. Ackerman (1993) reports evidence suggesting that, when provided with contextual information indicating the IM of a message, English-speaking children as young as 5 years can take this information on board and represent it separately from information gleaned from a message: LM. This is indicated by the fact that they can appreciate the sufficiency of their knowledge and that of others as shown by judging their ability to identify a story referent, despite recognising the insufficiency of an actual utterance to provide such knowledge as revealed by denying that a story character's utterance had sufficiently provided this knowledge.

For example, in one experiment Ackerman (1993) presented a third of children with an informative utterance version of a story in which Nancy's utterance enabled Janice to identify which type of apples she wanted: "I want the small yellow apples." Another third of the children were given an ambiguous utterance version of the story in which Nancy's utterance helped Janice to narrow down the selection of apple varieties she could choose from but did not enable her to identify the exact type of apple wanted: "I want the small apples" indicated that Nancy didn't want large apples but didn't reveal which colour apples she wanted The remaining third of children received a contextually informative utterance version in which information provided in the context of the story: reference of the recipe to yellow apples, complimented ambiguous information provided in the utterance "I want the small apples" to enable identification of the small yellow apples Nancy wanted.

In one condition the experimenter identified the type of apple Nancy wanted, but in another condition the type of apple was not indicated. Ackerman (1993) then asked the children whether they knew exactly which type of apple was right (children in the experimenter identification condition were asked if they had known before they had been shown) and whether what Nancy had said at the very end of the story was enough alone to know which type of apple was right.

Children as young as 5-years-old were significantly more likely to claim knowledge of the type of apple Nancy wanted when presented with the informative and contextually informative utterance versions of the story than when presented with the ambiguous utterance version. Furthermore, analyses indicated that children were not significantly likely to attribute this knowledge to the utterance alone in the contextually informative and ambiguous versions of the story. Such reasoning would seem to require clear distinction of IM from LM.

Moreover, Ackerman (1993) demonstrated that misattribution of the source of referent knowledge to the utterance occurred significantly more frequently following experimenter identification of the correct meaning/object referents. He interpreted this result to indicate that the confusion in development lies not in differentiating IM from LM, but in relating the message to reality. He suggested that early on, children tend to treat messages as reflective of IM, and to judge the consistency of messages and referents rather than their sufficiency, so that a message which precludes specific identification of, but nonetheless allows interpretation in terms of, its intended referent (i.e. an ambiguous utterance), is deemed adequate. However, he proposed that the focus on message and referent consistency can be disrupted, so that the absence of independent referent identification or the use of a deceptive context can act to direct attention away from the external referent to an internal representation. He proposed that the relation between internal representations and messages are judged in terms of sufficiency, thereby revealing children's conceptual competence.

The suggestion that children possess early LM-IM differentiation, sits well with Reid's (1996) report on the investigation of ambiguity sensitivity in English-speaking 5- to 6-year-olds. Recognition of ambiguity requires appreciation that the meaning of a message is consistent with multiple IMs. It thus requires differentiation of LM and IM. Reid argued that emphasising the need to evaluate messages, rather than to merely base judgements on them, improved ambiguity detection, suggesting that past task demands had previously obscured the true extent of competency children possessed. Reid provided evidence supporting the proposal that children possess an early appreciation of ambiguity. Participants were presented with a selection of stimuli: pictures of clown faces, differing with relation to some attributes: emotion expression, hat-shape, presence of freckles, and consistent with respect to others: possession of one nose and two eyes. They were asked to select a message referent if it could be identified from the message, or to pick up a card with a large

"X" on it if the referent could not be determined. If, for example, the set from which participants were asked to select consisted of a happy clown with a square hat and no freckles, a sad clown with a square hat and freckles and a happy clown with a round hat and no freckles, a message referring to "A happy clown with a square hat" would be unambiguous, consistent with only one of the clown faces. However a message referring to " a clown with a nose and two eyes" would be ambiguous, consistent with all of the clown faces.

Reid (1996) found that despite demonstrating difficulty in determining the adequacy of ambiguous messages, indicated by whether they attempted to identify referents, children took longer to respond and did so with less certainty when a message was ambiguous than when it was unambiguous. This was the case even when the intended referent was in fact known to them. These findings suggest, as do those of Ackerman (1993) that children are, to a degree, aware of the LM-IM distinction.

Sekerina, Stromswold, and Hestvik (2004) drew similar conclusions. They investigated picture selection in an ambiguous pronoun task. The following sentence provides an example of an ambiguous pronoun: "The boy has placed the box behind him". In this example, the term "him" could either refer back to the sentence subject: interpreted as an internal referent -the boy, or to an unidentified agent: denoting an external referent – another character. Sekerina et al., created pairs of pictures, so that in each couplet one scene favoured interpretation in terms of the sentence-internal referent: An image depicted two agents, a boy and a man and a box could be seen behind the boy, and the other scene promoted interpretation in terms of an external referent: An image depicted two agents, a boy and a man, and a box could be seen behind the man.

Sekerina, Stromswold, and Hestvik (2004) found that when using messages containing such ambiguous pronouns to select between picture pairs: an explicit response, English-speaking 4- to 7-year-olds failed to demonstrate appreciation of alternative meanings. The overwhelming selection of pictures representing the sentence-internal referent (93% responses), was only countered by a selection of pictures representing the sentence-external response equal to that of the error response rate in the control sentence-internal referent selection task. The control task was identical to the task just described, except that the ambiguous pronouns used in the messages were replaced with reflexives (such as himself or herself), in this way providing unambiguous identification of the sentence-internal referent. However, when Sekerina et al., used patterns of eye gaze to investigate the allocation of attention to the two referents in the pronoun task, they found a substantial increase in looks to the external referent (relative to the internal referent reflexive control task), indicating the presence of implicit (non-conscious) sensitivity to alternative interpretations and so the LM-IM distinction.

Eskritt, Whalen and Lee (2008) provide even further evidence of early implicit appreciation of the meaning differentiation. As mentioned earlier, Grice (1975) held that an understanding of his conversational maxims required application of the maxims to LM to draw inference about IM. It thus necessitates appreciation of the distinction between the two meanings. Eskritt, et al., presented English-speaking 3-, 4- and 5-year-olds with a conversational violations task (CVT) consisting of a series of trials in which they were asked to locate a sticker hidden under one of four pots. Each time they could seek helpful information from one puppet regarding the whereabouts of the sticker, and each time they were presented with the same two puppets to choose from. One of the puppets always provided appropriate information regarding the location of the sticker e.g. if the sticker was under the orange cup, this puppet would tell the child that the sticker was under the orange cup. However, the other puppet consistently produced a response which violated a specific conversational maxim. The conversational maxim violated, varied across participants, and was concerned either with the truthfulness, informativeness or relevance of communication. To illustrate, if the puppet was violating the maxim concerned with the relevance of communication, and the child asked for help locating the sticker, this puppet would produce a response such as "I like these cups".

If children had appreciation of the conversational maxims, they should recognise the maxim violations, and over the course of the trials, associate them with the particular puppet who was consistently producing them. They should then develop a preference for requesting help from the other non-violating puppet. Eskritt, Whalen and Lee (2008) reported that from the age of 3 years, children developed a preference for the non-violating puppet when violations pertained to the maxim concerned with the relevance of communication. From the age of 4 years, they found that children also demonstrated preferences for the non-violating puppet when violations pertained to the maxims concerned with the informativeness and truthfulness of communication. Recognition of these maxims provides support for the argument that children have appreciation of the LM-IM distinction at this early age.

It could be argued that children might have simply learnt that asking for help from the violating puppet did not aid them in locating the sticker, and that it did not require conclusions to be drawn regarding intended meanings such as deception, purposeful vagueness or deliberate irrelevance. If this was true, success on the task would not require and so reflect appreciation of the meaning distinction. Such a criticism certainly limits the conclusions which can be drawn from success on this task. However, even the weakest interpretation of the task supports the conclusion that children of this age are capable of learning about redundancy in communication, which provides the basis for drawing implicatures about intended as distinct from logical meanings.

1.5.2 The role of informational processing

It could be that information processing limitations prevent implicit knowledge from being exploited during explicit (conscious) response consideration. Such information-processing limitations may be masking implicitly revealed CU to preclude explicit demonstration, in line with a performance account of CU development (Surian, 1995). Corroboration of this analysis appears to be provided by Musolino and Lidz (2003). These authors investigated the phenomenon of isomorphism - the tendency to interpret ambiguous sentences with negation and quantified noun phrases, such as "every horse didn't jump over the fence", in a manner consistent with their superficial syntactic structure. Such a strategy is typically employed by young children, in that they will use the surface syntactic position of elements to determine the scope of their application. Thus in the example given above, the quantified expression "every", in preceding "didn't", will take scope over this negation, producing the "wider" interpretation that no horse jumped over the fence. This contrasts with the non-isomorphic interpretation of such sentences typically given by adults. They will normally attribute scope to the negation "didn't" over the quantified expression "every" resulting in the "narrower" reading that not every horse jumped over the fence so giving rise to the implication that some in fact did.

Musolino and Lidz (2003) demonstrated that the isomorphic tendency of children actually reflects an attenuated processing preference present in adults rather than a conceptually immature representation. They achieved this by revealing that adults can form isomorphic interpretations in certain circumstances, and by further establishing that the incorporation of a manipulation to reduce processing demands, reduced the extent of isomorphic construal, replicating the effect reported with children (Musolino & Lidz 2006).

Musolino and Lidz (2003) established that English-speaking adults overwhelmingly interpret the ambiguous phrase "Two frogs didn't jump over the rock" isomorphically, in accordance with its syntactic structure. In such an interpretation, the quantification term "two" takes scope over the negation term "didn't", giving rise to the "wider" reading that two (particular/definite) frogs didn't jump over the rock. The wider interpretation can be contrasted with an alternative interpretation in which the negation term "didn't" takes scope over the quantification term "two". This interpretation would give rise to the "narrower" reading that it was not the case that two frogs jumped, and hence the implication that less than two frogs jumped. Musolino and Lidz then demonstrated that the creation of a positive expectation against which the negative quantified noun phrase could be starkly contrasted, as in "Two frogs jumped over the fence but two frogs didn't jump over the rock" vastly enhanced the extent of non-isomorphic interpretation. The use of an affirmative statement with which to contrast negation is thought to reduce the demands incurred in processing the negation (Horn, 1989). Evidence that processing resource manipulation affects the isomorphic interpretations of adults in a manner consistent with its effects upon the isomorphic interpretations of children, would seem to provide strong support for a continuity account of CU competency, indicating early conceptual masking induced by processing demands.

Investigations focusing on the appreciation of SIs also indicate that reducing processing requirements lowers the age at which CU is demonstrated. Noveck (2001) studied the appreciation that English-speaking children between the ages of 5- and 10-years-old and adults have of the SIs arising from the weak scalar terms "might" and "some". To examine whether children understand that the conversational use of "might" implies "not necessarily", Noveck presented participants with 2 open boxes, one of which contained a parrot and a bear and one of which contained just a parrot, and a closed box which they were told had the same contents as one of the open boxes. A puppet then proceeded to deliver a series of statements to the children concerning the contents of the closed box and the children were asked to evaluate these statements. The critical statement was "There might be a parrot in the box". If children derive the implicature that "might" entails "not necessarily" they should negatively evaluate the puppet's statement as there did have to be a parrot in the closed box. However, if they fail to compute the SI, they should interpret "might" logically to be consistent with the reading that there must be a parrot, and so positively evaluate the puppet's statement. Noveck found that despite demonstrating accurate evaluation of control statements provided by the puppet, 7- and 9-year-olds produced significantly more logical interpretations of the critical SI statement than adults.

Noveck (2001) also investigated the ability of French speaking children to understand that the conversational use of the French term for "some": certains, implies "not all": pas tous. Children were presented with a series of control assertions such as the phrases "All dogs have spots" and "Some birds live in cages" (English translations) which varied in terms of their logical truth. However, underinformative statements which were logically true but which gave rise to false SIs were also presented, for example "Some giraffes have long necks" (English translation). Noveck discovered that despite demonstrating accurate evaluation of control statements, 10-year-olds were significantly more likely than adults to incorrectly accept the underinformative statements indicating that they were interpreting the term "some" logically. Thus, Noveck provided evidence of a clear preference for the logical interpretation of weak scalar terms across languages and terms.

However Chierchia et al., (2001), and more recently, Papafragou and Musolino (2003), Papafragou and Tantalou (2004), Feeney, Scrafton, Duckworth, and Handley (2004) and Guasti et al., (2005) have reported that adapting tasks to lessen processing requirements, substantially enhances the ability of children, some of whom are as young as 3 years of age, to compute SIs.

Chierchia et al. found that presenting a description containing a SI alongside a description without the SI enabled English-speaking children between the ages of 3- and 6-years old to compute the SI that "or" implies "not and". Chierchia et al. presented children with a scenario for which the strong descriptive term "and" was appropriate: A story was told about some farmers who wanted to clean their animals, and who eventually decided to clean both a horse and a rabbit. The children then listened as 2 puppets provided descriptions of the farmers' actions, one of which gave rise to a false SI: Every farmer cleaned a horse or a rabbit, and the other of which did not evoke an SI: Every farmer cleaned a horse and a rabbit. Children were found to display a strong preference for the description which did not evoke the false SI. It appeared that presenting children with a choice of statements to choose between, reducing the need for independent representation of a contrasting interpretation to the

SI, promoted children's ability to demonstrate SI appreciation through avoidance of the false SI.

Papafragou and Musolino (2003) found that emphasising the relevance of contextually driven meanings by presenting SIs as descriptions of goal attempts, and using a training phase to highlight experimenter intentions, also helped young children to compute SIs. Papafragou and Musolino provided Greek-speaking 5-year-olds with training trials which emphasised the need to focus on the felicity/appropriateness of statements rather than their logical truth. Children watched the experimenter ask a puppet to identify items in front of them and then heard the puppet give responses which were either both logically true and felicitous/appropriate e.g."This is an elephant" in reference to an elephant, or both logically true but infelicitous/inappropriate e.g. "This is a little animal with four legs" in reference to a dog. After each response, children were asked if the puppet answered well, and if there was a better answer. If children failed to correct a puppet's inappropriate response, the Experimenter told them that the puppet did not answer well and indicated what a better response would have been, e.g. "This is a dog" in reference to the dog.

Children were then presented with test trials in which they were again asked to evaluate puppet statements which this time contained SIs. The relevance of SIs was emphasised by indicating that story characters were attempting to achieve goals e.g. Mickey was challenged to put hoops on a pole, and that the puppet statements were performance descriptions which should reflect whether or not the goal had been achieved. A typical trial consisted of Mickey being shown to place all 3 available hoops on a pole and the puppet describing Mickey as having put 2 hoops on the pole. Papafragou and Musolino found that 5-year-olds were significantly more competent at computing the "some implies not all", "start implies did not finish" and "two implies not three" SIs when given the felicity training phase and presented with a high relevance SI context than when asked to compute the SIs in a less relevant context without training.

Papafragou and Tantalou (2004) built on the findings of Papafragou and Musolino by enhancing the ability of young children to compute SIs. Although Papafragou and Musolino had found that training and a high SI relevant context enhanced SI computation, SI computation for two of the scalar terms did not rise above chance-levels with 5-year-olds only computing the "some-all" and "start-not finish" SIs half of the time. However, Papafragou and Tantalou heightened the accessibility of contextually driven meanings, by presenting SI statements as responses to questions inducing glaring expectations regarding situational notification. Rather than just forming descriptions of what happened, as in Papafragou and Musolino, the SI statements in Papafragou and Tantalou's study were produced as answers to queries regarding the satisfaction of specific criteria e.g. Did you eat the sandwich? I ate the cheese. Papafragou and Tantalou's SI task was also more naturalistic because children were encouraged to derive implicatures instinctively, as a sincere response to hearing directed communication, rather than in the explicit context of evaluating utterance felicity. The sincerity of SI computation was attained by asking the children to assign prizes on the basis of achievement as indicated by answers to queries regarding the satisfaction of specific criteria. Computation of an SI that for example, "not all" of a task was achieved, should thus have directly impacted on assignment of a reward.

Papafragou and Tantalou (2004) investigated computation of the "some-all" quantificational SI, encyclopaedic SIs requiring world knowledge e.g. that a sandwich comprises more than just cheese as relevant to the SI mentioned in the preceding paragraph, and ad hoc SIs requiring situational knowledge e.g. that there are 2 gifts comprising a parrot and a doll to be wrapped in relation to the SI: Did you wrap the presents? I wrapped the parrot. Greek-speaking 4- and 5-year-olds were found to compute all the SIs at significantly above chance levels on Papafragou and Tantalou's more naturalistic and direct task, in support of the suggestion SI computation can be promoted by reducing the processing load introduced through incorporation of unnecessary task demands.

Feeney et al. (2004) also reported that increasing the accessibility/relevance of enriched contextual meanings improved children's SI performance. Feeney et al. contrasted the ability of English-speaking 8-year-olds to compute "some-all" SIs when the SIs were presented in English translations of the contextually thin statements used by Noveck (2001) Experiment 2 e.g. "Some giraffes have long necks" with their ability when the SIs were delivered via contextually thick storyboards in which pictures and text were used to depict the actions and interactions relevant to the SIs. For example, one storyboard presented a series of photographs and text showing a child discovering and then consuming, in a one by one fashion, all of the sweets she had found. The child was then shown facing an angry mother who asks "Charlotte what have you been doing with the sweets?" Charlotte is shown to reply "I've eaten some of them". Feeney et al found that the 8-year-olds performed well in the contextually thick condition, computing 79% of SIs and produced significantly more SIs in the contextually thick condition than in the contextually thin condition.

Guasti et al. (2005) shed further light on factors affecting children's ability to compute the some-all SI by considering the effects of felicity training separately from the effects of enhanced accessibility/relevance of enriched contextual meanings. In one experiment, felicity training was introduced which consisted of drawing children's attention to the fact that logically true descriptions are not necessarily the best. Italian-speaking children saw a set of objects and were asked to choose between two labels for each object e.g. a grape was presented and children were asked to identify the better description from the Italian translations of the terms "grape" and "fruit". SI appreciation was then assessed using Italian translations of the test statements used by Noveck (2001) Experiment 2. In a separate experiment, the accessibility of enriched contextual meanings was enhanced by embedding SIs in descriptions of events provided by an observing puppet, inline with the SI test presentation of Papafragou and Musolino (2003). For example, a puppet watched five soldiers discuss whether they would go in search of treasure by motorbike or horse, and then all choose to travel by horse. The puppet was then asked what was happening in the story and replied using the description "Some soldiers are riding a horse".

Guasti et al. found that training improved the ability of Italian-speaking 7-year-olds to compute SIs, although this enhanced ability was still below adult levels and did not persist over time (1 week) to a second testing session. However, enhancing the accessibility of enriched contextual meanings was found to raise performance to adult levels. It is noticeable that Papafragou and Musolino reported that the combined effects of training and enhancing the accessibility of enriched contextual meanings only raised performance to chance levels. However, the children tested by Papafragou and Musolino were younger by 2 years. This indicates that there is more to the emergence of SI appreciation than increased access to enriched contextual meanings, at least to the extent that it was enhanced in this study and by Papafragou and Musolino. This additional factor could perhaps relate to the ability to independently represent contrasting interpretations to SIs, which Chierchia et al. (2001) found led to strong SI appreciation among children as young as 3- to 6-years old.

The report of Pouscoulous, Noveck, Politzer, and Bastide (2007), lends additional weight to the suggestion that processing costs mask early appreciation of SIs. Pouscoulous et al. found that the complexity of linguistic material incorporated and context and nature of the task used, impacted on children's computation of SIs. In an initial experiment, French-speaking 9-year-olds were presented with a set of puppet statements as descriptions of a scene in front of them. The scene comprised a set of boxes which were surrounded by toy elephants: no elephant was inside a box, and which contained toy turtles: no turtle was outside a box. Additionally, toy dolphins were scattered both in and around boxes. The critical false SI statement was a French translation of the phrase "Some turtles are in the boxes" using the term "certains" to express the meaning "some". As expected, the 9-year-olds incorrectly agreed with the critical SI statement more than adults, revealing a greater tendency to interpret SIs logically, despite demonstrating adult-like performance on control items such as the translated "Some dolphins are in the boxes".

However, the processing demands of the task were reduced in a subsequent experiment. The linguistically complex French term "certains" was replaced with the more simplistic French term "quelques". Furthermore, distractor items: attention-attracting items that are irrelevant to the SI statement, were removed. Participants were presented with a basic scene comprising boxes containing tokens, in which no token was placed outside a box. Additionally, rather than truth evaluation, Pouscoulous et al. (2007) used an action response to ascertain SI computation. Rather than asking participants to agree or disagree with a scene description, Pouscoulous et al. asked them to make sure a scene satisfied a statement such as the translated "I want all the boxes to contain a token", modifying the scene if required. Such changes were found to promote the ability of young children to compute SIs, so that only 32% of 4-year-olds, 27% of 5-year-olds and 17% of 7-year-olds responded logically to an SI. That is to say that as well as demonstrating competence in restraining themselves from making inappropriate scene modifications in response to control statements, e.g. in response to the translated statement "I would like some boxes to contain a token" given during a scene in which some, but not all of the boxes contained a token, the majority of children correctly modified a scene in which all the boxes contained a token upon hearing the translated SI statement "I want some boxes to contain a token". The low rate of logical response in this experiment can be contrasted with the rate of 91% produced by the 9-year-olds in Pouscoulous et. al.'s initial experiment.

Pouscoulous et al. also considered the effect of computing SIs in a negative context e.g. "I want some boxes NOT to have a token". Negative contexts are more difficult to process than their corresponding positive contexts i.e. "I want some boxes to have a token" (Horn, 1989, Prado & Noveck in press). In line with the masked competence account, the authors found that presenting SIs in more processing-intensive negative contexts appeared to make them significantly more difficult for children to compute. In contrast to the 17% of 7-year-olds responding appropriately to SIs in the standard positive context described above, 40% of 7-year-olds incorrectly neglected to modify a scene in which no boxes contained tokens in response to the negative context SI statement translated as "I would like some boxes to not contain a token". This was the case, despite demonstrating appropriate responses to the same statement in control scenes e.g. in which all the boxes contained a token.

The analysis of personal reference pronoun reversals conducted by Dale and Crain-Thoreson (1993) lends further support to the suggestion that children possess early proficiency in deriving pragmatic interpretations in discourse, but that this is susceptible to masking by processing demands. Personal reference pronouns are special in that their object of reference is not fixed, so that the term "you" can be used to refer to different people in different contexts and even during simple conversational exchanges in which the speaker and listener exchange roles. Correct production of personal reference pronouns such as "I", "You", "he" or "she" is performed in accordance with Grice's maxims of Quantity. However, sometimes reversals occur, in which the pronouns are incorrectly applied so that the speaker might use the term "I" to refer to the listener or "you" to refer to him/herself.

Dale and Crain-Thoreson (1993) established that reversal production within their sample of English-speaking 20-month-old children was unrelated to cognitive abilities such as visuo-spatial skills (indicating they were not attributable to difficulties in perspective switching) and negatively associated with proportion of referential vocabulary (suggesting they were not a consequence of a bias to attribute speaker independent referential meanings). They also discounted effects of grammar (indicating reversals were not a consequence of lack of relevant linguistic knowledge) and imitation (suggesting they were not a consequence of blindly copying the application of others). This enabled rejection of an explanation in terms of conceptual difficulty. However, they observed that reversers failed to produce reversals consistently, and that the majority occurred in the context of relatively long or complex utterances. It was also noted that reversals were more likely to occur in imitation contexts and to involve reversal of the second person pronoun "you".

Bearing such description in mind, but also taking into account the fact that those children who actually produced reversals infact tended to produce more correct second person pronouns than non-reversers, Dale and Crain-Thoreson (1993) propose a processing complexity explanation of reversals. They suggest that the referential shift required to alternate pronoun direction "presents a substantial processing load" (Dale & Crain-Thoreson, 1993, p. 585), susceptible to impediment from long complex utterances and salient pronoun representations formed in memory during imitation episodes. Furthermore, the very fact that reversers produced more *correct* second person pronouns than non-reversers despite producing more incorrect pronoun reversals than the other group, is used to support a conception of reversers as competent "risk takers", taking chances with their application of pronouns rather than employing the alternative formulaic (thank-you), generic (non-specific application of a reference term to identify whomever appropriate in the circumstances), nominal (Daddy read) and third person (go away) reference strategies of non-reversers. This pattern of results strengthens the case for a theory of proficiency concealed by limited processing resources.

Meroni and Crain's (2003) investigation of the "garden-path" effect in young children lends additional support to the theory that conversational understanding is present from an early age but is masked by a lack of processing power. These authors investigated the performance of children on "garden-path" sentences (in which transitory referential ambiguity can lead to ascription to an interpretation later discredited by analysis of further input). An example of a garden path sentence would be "Whilst Myrtle ate the chicken laid an egg". The referential ambiguity in this example is due to the phrase "the chicken", containing the second noun of the sentence. This noun could be interpreted as the object of the verb "ate" (indicating that the chicken was eaten) or read as the subject of the verb "laid" (indicating that the chicken laid an egg). The "garden path" description of this sentence, is supported by evidence concerning the interpretation of similarly structured phrases, in which processing behaviour is consistent with the presence of an initial tendency to process the noun immediately following a verb as its object, superseded following the processing of the second clause of these sentences, by the need to assign a clausal subject (Mitchell, 1987; Van Gompel and Pickering, 2001).

However, the idea that some sentences are inherently misleading presupposes the initial centrality of grammatical judgements to ambiguity resolution and language processing more generally. Although evidence has supported such a claim under restricted conditions (Britt, 1994), many findings support the suggestion that other factors such as semantic plausibility and referential context exert influence during the earliest stages of processing (Britt, 1994; Garnsey et al., 1997; Trueswell, 1996). Nonetheless, when Trueswell et al. (1999) analysed eye-movements during sentence processing their findings appeared to indicate that unlike adults, 5-year-old children failed to use context to identify appropriate sentence referents, indicating a lack of CU proficiency in this area. However, Meroni and Crain (2003) reported that when given an opportunity to plan task responses before being allowed to initiate them, extending the processing period, English-speaking children between the ages of 3- and 6-years-old produced more contextually appropriate responses, failing to be led down the garden path.

Meroni and Crain (2003) asked children to "Put the frog on the red napkin into the box". The instruction is logically ambiguous and could be read to require either moving a frog which is already on a red napkin directly into a box or to require moving a frog onto a red napkin and then into a box. However, children were presented with a context comprising one frog sat on a blue napkin, another frog sat on a red napkin, a further "empty" red napkin and a box. Consideration of the scene should prompt interpretation of the phrase "on the red napkin" as a modifier of the sentence subject "the frog" and so lead children to select the frog sat on the red napkin and to move it directly into the box. However, when children were allowed to respond to the instruction immediately, they were equally likely to move the frog sat on the red napkin and the frog sat on the blue napkin. This suggests that they were unable to integrate contextual information in the scene with the logical information provided in the instruction. Nevertheless, when children experienced an enforced delay between hearing the instruction and responding, introduced by asking children to turn away from the scene whilst listening to the instruction and thus requiring them to turn back to face the scene before responding, children correctly selected the frog sat on the red napkin in 92% of the trials. A lengthier processing period appeared to improve contextually appropriate interpretations of instructions, providing further evidence that children possess CU masked by processing limitations.

More recently, Beck, Robinson and Freeth (2007) investigated children's ability to delay interpretations to ambiguous instructions and provided evidence which indicates that apparent early difficulties with ambiguity detection might be more reflective of difficulties in decision-making. Beck, Robinson and Freeth compared the performance of English-speaking 5- to 6-year-olds and 7- to 8-year-olds on a task requiring children to make a choice between responding to ambiguous instructions or delaying an interpretation until further information had been provided. A doll's toy was hidden in 1 of 3 envelopes placed on a table and children were told that they would be provided with 2 clues to help the doll identify the envelope containing the toy. Children were told that if they could identify which envelope contained the toy after the first clue was placed on the table, they could move the doll to stand by that envelope. However, children were told that if they wanted to wait until they heard the second clue before identifying the envelope, they could move the doll to stand next to the first clue. To assess children's ability to explicitly ascertain their knowledge of the toy's whereabouts, separate trials were given in which following the presentation of a clue, children were asked "Do you really know where it is or don't you really know?"

Beck, Robinson and Freeth (2007) found that when given ambiguous clues regarding the toy's whereabouts e.g. if presented with 3 spotty envelopes and told "the toy is in the spotty envelope", 7- to 8-year olds found it easier to delay their identification of an envelope, indicating implicit recognition of the redundancy of the clue, than to explicitly acknowledge their ignorance. 5- to 6-year-olds however, were equally poor at delaying their identification and acknowledging their ignorance in response to ambiguous clues. Nevertheless, 5- to 6-year-olds appeared to demonstrate greater proficiency in delaying their response to ambiguous information when not required to make an explicit decision to do so. In a separate experiment, Beck, Robinson and Freeth presented English-speaking 5- to 6-year-olds with a continuous stream of information and simply asked children to indicate the identification of a picture which was slowly being uncovered, from a possible set of 4 options as soon as they could determine it. Some pictures could be identified early on, but some pictures revealed ambiguous details early on and so could only be identified following further unmasking. Beck, Robinson and Freeth reported that children demonstrated competence in delaying identification of a picture for which only ambiguous details had so far been revealed: 87% responded correctly on 5 or 6 out of 6 trials despite demonstrating poor ability to acknowledge their ignorance of the picture's identity:

only 39% children responded correctly on 5 or 6 out of 6 trials. This prompted Beck, Robinson and Freeth to suggest that it is the reduced decision-making ability of young children rather than an insensitivity to ambiguity which leads them to demonstrate poor performance in CU tasks. This is inline with a masked competence account of CU demonstrated.

Furthermore, Nilsen and Graham (2008) found that children's ability to inhibit pre-potent responses was negatively related to their failure to take their conversational partner's perspective into account, providing even more direct evidence that children's processing ability is associated with their ability to demonstrate CU. Nilsen and Graham presented English-speaking 3-, 4- and 5-year-olds with a referential communication task in which the children had access to common ground information which was also available to an experimenter, and privileged ground information which was not shared by the experimenter. The information concerned the location of objects in a display case placed inbetween the children and the experimenter. Common ground information concerned objects which were visible from both sides of the display case and so visually accessible to both the children and the experimenter. Privileged ground information concerned objects which were only visible to the child due to an opaque screen blocking the experimenter's visual access to certain objects. Children were also presented with information processing tasks assessing their ability to inhibit pre-potent responses, their ability to manipulate information in memory and their ability to demonstrate cognitive flexibility (shift between alternative representations of the world).

Nilsen and Graham (2008) compared children's use of adjectives in instructions provided for the experimenter in the common ground condition, with their use of adjectives in instructions provided in the privileged ground condition.

Comparison focused upon instructions to pick one of two similar objects differing on one dimension e.g. the small duck in a display case containing both a large and a small duck, when both objects were visible to the child and the experimenter in the common ground condition, but only the object being described was visible to the experimenter in the privileged ground condition. Nilsen and Graham also assessed children's ability to identify one of two similar objects differing on one dimension from instructions provided by the experimenter which lacked reference to the critical dimension, in the privileged ground condition when both objects were visible to the child but only one of the objects was visible to the experimenter.

Children's choice of, and looks to, the referential alternative i.e. the object that was hidden from the experimenter in the privileged ground condition and thus not relevant to her instructions, were found to negatively correlate with their ability to inhibit pre-potent responses when the effects of age and verbal ability had been taken into account. More specifically, children who were better able to perform correctly in an inhibitory control task requiring children to name dogs drawn in blue ink as "Red" and dogs drawn in red ink as "Blue", and who thus demonstrated their ability to inhibit the pre-potent tendency to use colour words to refer to visual colours, were more likely to take the speakers visual context into account when interpreting communication. This provides further evidence that information processing ability is associated with children's capacity to demonstrate CU, inline with the masked competence account of CU.

A considerable amount of evidence thus appears to support the proposal that the emergence of conversational understanding is not so much a consequence of conceptual change, as a reflection of maturing processing abilities. This account is in line with Relevance Theory and the work with adults discussed in Section 1.3.

1.6 Theory of mind and conversational understanding

As well as concentrating on more sensitive measures of children's appreciation of the message meaning-intended meaning distinction, recent research has also focused on the role of "theory of mind" (ToM) reasoning in CU. It is widely agreed that CU requires appreciation of the speaker's intended meaning: communicative intentions. Intentions are a type of mental state, so a listener's recognition of communicative intentions necessitates conceptualisation of the mental world, commonly referred to as ToM reasoning. However, whilst early research on CU had neglected to investigate whether ToM ability was actually involved, more recent research produced evidence of a relationship between the two factors.

Many individuals with autism appear to suffer from a ToM deficit and also demonstrate impaired CU. Their CU difficulties are evidenced in a failure to take account of context when interpreting ambiguous terms and sentences, and to make bridging inferences which enable relation of communication to the context in which it is presented (Joliffe & Baron-Cohen, 1999). Problems are also encountered with understanding metaphors and producing speech acts (Dennis, Lazenby, & Lockyer, 2001). Metaphors involve identification of one subject with another e.g. "You are a real rock", enabling implication that pertinent characteristics of the second subject can be seen in the first. Contextual inferencing is necessary for this to enable decisions regarding which characteristics are referred to. Speech acts are produced when communication is used to achieve a goal. For example, "Mary, this is Lisa, Robert's wife" is a speech act of introduction. Contextual knowledge is required to appreciate this as it enables identification of situation-appropriate speech goals. The poor CU of individuals with autism is also manifest in the problems they display in grasping word connotations (Frith, 1989) i.e. the emotions and attitudes associated with the use of one particular term over another to express a common basic meaning. For example, the use of the term "paraded" rather than the term "walked" when discussing how someone moved along the street, has connotations of flamboyance and attention-attraction. Appreciation of such connotations requires contextual knowledge to enables identification of these alternative attributions to words with the same basic meanings.

Noveck, Guelminger, Georgieff, and Labruyere (2007) reported that like French-speaking 4-year-old children, French-speaking adults with autism failed to demonstrate the non-isomorphic tendency of typically developed adults, when interpreting ambiguous sentences with negation and quantified noun phrases, such as "every horse didn't jump over the fence" (see Section 1.5.2). Neither adults with autism, nor the children tested, were found to demonstrate a preference for attributing scope to the negation "didn't" over the quantified expression "every". Parsing in such a way, would result in the "narrower" reading that not every horse jumped over the fence, giving rise to the implication that some in fact did. However typically developed adults were found to demonstrate exactly this narrow parsing preference, indicating that the context warranted the inference of a narrower reading. A third of the participants with autism taking part in this study, were found to fail a standard first-order false-belief assessment of ToM, despite possessing chronological and mental ages considerably greater than those at which typically developing children usually pass these tasks. That is to say that a third of the participants with autism tested, demonstrated difficulty recognising that someone could hold a false belief about the physical world. This suggests that the autistic adult's lack of ToM might have played a role in the restriction of their CU.

In further accord with the proposal that CU requires mental state appreciation, Sodian and Frith (1992) reported that English- and German-speaking children with autism between the ages of 6- and 19-years-old demonstrated poor performance on tasks tapping deceptive ability, despite displaying competent performance on tests of sabotage. Deception involves intentionally communicating incorrect information to someone, requiring consideration of the message in light of its context, and thus reflecting CU. However sabotage merely requires the physical impediment of another's plans. The ability to conduct sabotage in the absence of ToM, indicates that the problem experienced was specific to the obstruction of another's goals through the use of communication and not a more general goal impediment difficulty. Moreover, the deception task performance of the children with autism was predicted by their performance on a first-order false-belief task assessment of ToM. This provides strong support for the argument that mental state reasoning is required for CU.

Surian, Baron-Cohen, and van der Lely (1996) also found a direct relation between level of ToM and level of CU. Surian, Baron-Cohen, and van der Lely found that English-speaking children with autism with a mean chronological age of 12 years 11 months performed at a much lower level than verbal IQ matched children with Specific Language Impairment with a mean chronological age of 11 years 10 months and verbal IQ matched typically developing children with a mean chronological age of 6 years 7 months, on a CVT task assessing appreciation of Gricean conversational maxims. The CVT presents a scenario in which a doll: Lucy, poses questions e.g. "Where do you live?" and 2 dolls: Tom and Jane, each provide a response e.g. Tom says "I live on the moon" and Jane says "I live in a town". One response to each question violates a Gricean maxim. In this example, Tom's response "I live on the moon" violates the Maxim of Quality as the answer is obviously untrue. Task participants are told that for each question, one of the dolls will give a silly answer, and each time are asked to indicate which doll has given the silly answer. More significantly, the CVT performance of the children with autism was related to their performance on a first-order false-belief task assessment of ToM. This provides further support for the conception of CU in terms of the recognition of intended meaning.

Additionally, Tedoldi, Surian and Siegal (2005) found a relationship between CVT proficiency and performance on a first-order false-belief task assessment of ToM in a sample of deaf children, comprising both native and late-signers. Late signing deaf children have been found to demonstrate similar difficulties with false-belief tasks to those produced by children with autism, whereas native signing deaf children appear to perform at a level comparable with typically developing hearing children (Woolfe, Want, & Siegal 2002). These results lend further weight to the proposal that CU requires intention recognition.

Additional support for the conceptualisation of CU in terms of the appreciation of intended meanings, is lent by Happé (1993). Happé considered the performance of typically developing English-speaking 4- and 5-year-olds on second-order ToM tasks which assess the ability to appreciate that people can form false beliefs about other people's beliefs about the physical world. Second-order ToM tasks require more comprehensive mental world understanding than first order tasks which only assess the ability to appreciate that people can form false beliefs about the physical world. Happé found that the performance of typically developing 4- and 5-year-olds on second-order ToM tasks predicted their scores on a task assessing appreciation of irony. Irony arises when the speaker intends for the listener to recognise their intention to create a situation in which information drawn from the context of a message is inconsistent with the message itself, with a view to generating humour. Thus, appreciation of irony requires consideration of the message in light of its context, and so reflects CU. To illustrate irony, Happé presented children with a story in which a boy: David, is left to make a cake on his own. However, when David adds the eggs to the rest of the ingredients he doesn't break the shells, adding the eggs to the ingredients along with their shells. Children are told that when David's father comes by and sees what David has done, he says "What a clever boy you are, David!" Children's appreciation of irony was examined by asking them "What does David's father mean? Does he mean David is clever or silly?"

Martin and McDonald (2004) reported similar findings for English-speaking youths with Aspergers Syndrome (generally regarded as suffering from a ToM deficit). Martin and McDonald also reported that second order ToM appreciation predicted recognition of irony. Further, Langdon, Davies and Coltheart (2002) reported a relationship between first order ToM and the ability to appreciate irony in English-speaking patients with schizophrenia. They found that first order ToM predicted appreciation of irony, even when other factors such as difficulties with inhibition, were taken into account.

The role of ToM in CU is also indicated by the findings of Winner et al. (1998). They investigated the ability of English-speaking patients with right hemisphere damage (ToM impairments are common in this population), to distinguish lies from ironic jokes. As outlined above, ironic jokes arise when the speaker intends for the listener to recognise their intention to create a situation in which information drawn from the context of a message is inconsistent with the message itself. However, lies occur when the speaker doesn't intend for the listener to recognise their intention to create a situation in which information drawn from the context of a message is inconsistent with the message. Appreciation of ironic jokes and lies, and the difference between them, thus requires consideration of messages in light of their context, and so reflects CU. Winner et al. found that their patients with right hemisphere damage indeed demonstrated impaired ToM ability, and that their performance on second order ToM tasks predicted their ability to distinguish ironic jokes from lies. The relationship between second order ToM and the ability to distinguish ironic jokes from lies was also demonstrated by the typically developed controls employed in this task.

However, Hadwin et al. (1997) found that training English-speaking children with autism who were between the ages of 4- and 14-years-old, to pass tests assessing ToM, did not improve CU. CU was gauged in terms of story narration ability, which requires consideration of the previous story context in which current storylines are being presented, or the conversational use of mental state terms which requires consideration of the context in which the terms are being presented. However, it is likely that the training scheme used, bestowed only a shallow, narrow focus of mental state understanding, insufficient to supply the insight into communicative intentions required for CU.

Sperber and Wilson (2002) propose that that the part of the brain responsible for processing the meaning of communication is actually a sub-module of the area in the brain concerned with ToM. They point out that the demands communication makes on ToM are different to those made by other types of human behaviour, noting that the range of possible intentions behind communication is vast compared to the scope for physical actions, the latter restricted by practical constraints on activities which do not apply to communication. This leads Sperber and Wilson to suggest that there is a submodule of the area of the brain concerned with ToM, dedicated to the specific inference of communicative intentions, employing unique processing mechanisms to achieve this aim.

Happé and Loth (2002) point out that the suggestion that ToM works differently in the communicative domain to elsewhere, sits well with findings indicating its early demonstration in communicative contexts (e.g. Baldwin, 1993, O Neill, 1996) despite its apparent absence in other domains (children standardly fail false-belief tasks before 4 years of age (Wellman, Cross, & Watson, 2001)). Indeed, Happé and Loth's discovery that English-speaking children failing a standard Sally-Anne unexpected location false-belief task (Baron-Cohen, Leslie, & Frith, 1985) were demonstrating false belief understanding in the context of establishing effective communication, lends further support to Sperber and Wilson's account. The idea that ToM might come online in the communicative domain prior to its appearance elsewhere, would also concur with a longitudinal study of conversational perspective-taking (CPT) and first-order false belief understanding.

Bernard and Deleau (2007) presented French-speaking 3- and 4-year-olds with a CPT task assessing their ability to match promissive speech acts e.g. "I will ....", directive speech acts e.g. "You will ...." and assertive speech acts e.g. "This is....", with social relationships. For example, children were required to indicate the assertive speech act "My drawing is nicer than yours" (English translation) would be spoken by a baby to an older child, by a girl to a boy or by a woman to a man. Children were also given a CPT task examining their ability to relate communication to the physical context from which it emerged. Scenes were presented which provided details of a previous event and the characters involved e.g. 3 boys playing football. A further scene was then presented in which 1 of the characters involved in the previous scenes produced a communication e.g. "Why don't you want to play with us anymore"

(English translation) in a situation in which either 1 more character from the previous scene was present and a further 2 new characters were shown, as for this example, or 2 more characters from the previous scene were present and a further 1 new character shown, as for an example where 3 children had been reading, 1 of whom later said "It's a pity you didn't come and read with us" (English translation). Children were required to identify which of the characters present in the communication scene, the communication had been directed at. The CVT was used as a further CPT task.

Bernard and Deleau (2007) found that CPT scores summed across the 3 tasks, predicted false belief appreciation, but that the reverse relation did not hold. It could well be, that activity in a relatively autonomous communication submodule of the region of the brain concerned with ToM, nevertheless promotes activation elsewhere in the surrounding region.

1.7 Summary

CU refers to the ability to appreciate how language is adapted in accordance with the context in which it is used. However, whilst adults appear proficient in this area, children below around 7 years of age fail to spontaneously demonstrate CU competency. Early research supported the idea that CU emerged as a result of a radical shift in the way that language was viewed – a conceptual change. However, more recent findings have supported the proposal that children in fact possess CU from early on, with some studies suggesting that children as young as 3-years-old possess CU, but that this proficiency is masked by restricted information processing resources. Recent research has also provided evidence to endorse the suggestion that ToM plays a role in CU. This supports the conceptualisation of CU as a process involving the appreciation of communicative intentions.

Chapter 2: Executive functions and conversational understanding

2.1 Elucidating the concept of "executive functioning"

The previous chapter examined evidence suggesting that even young children possess considerable CU. However, their proficiency appears to be masked by restricted information processing resources which require investigation. Information processing resources are commonly conceptualised in terms of cognitive mechanisms involved in the integration, manipulation and interpretation of input from the environment. However. such mechanisms are dependent upon executive functioning (EF) to provide processing direction. EF can be broadly defined as a group of skills which "control and regulate thought and action" (Friedman et al., 2006). It is therefore plausible that limitations in EF are responsible for the "masking" of early CU.

EF is an "umbrella" term, used to refer to a multitude of abilities related by their common goal of enabling flexible, goal-directed behaviour (Hughes & Graham, 2002). Zelazo and Muller (2002) highlight a number of different approaches to the precise conceptualisation of EF. For example, it has been portrayed as a collection of multiple sensory-driven abilities impaired by damage to the prefrontal cortex (Tranel, Anderson, & Benton, 1994), but also as a unitary sensory-driven ability from which all apparent EF deficits can be thought to derive (Carlson, Moses, & Hix, 1998). EF has additionally been conceptualised as a higher order (i.e. involving conscious reasoning) processing ability (Baddeley, 1996; Norman & Shallice, 1986) which Miyake et al. (2000) point out had "at least in the early stages of development… a unitary flavor" (p. 51). However, EF has also been considered as a problem solving function in which differentiable EF abilities are conceptualised not as independent faculties separable from the context in which they are applied, but rather in terms of their contribution to the acquisition of problem solving goals (Zelazo, Carter, Reznik

& Frye, 1997).

Miyake et al. (2000, p. 49) focus on the debate within the literature concerning the "unity and diversity" of EF. They note support for the proposal that there is a common element underlying EF, such as goal neglect (Duncan et al., 1997), diminished associations among working memory representations of objectives, the items/agents involved and relevant knowledge (Kimberg & Farah, 1993) or even an area of the brain such as the pre-frontal cortex (Engle, Kane, & Tuholski, 1999). However to contrast this position, they highlight neuropsychological and correlational evidence pointing towards the contrary view that EF is fractionated rather than unitary, a stance also considered by Zelazo and Muller (2002).

Studying the underlying neural basis of cognitive abilities can shed some light on the constitution of the core concepts involved. Dissociations manifest in the neural correlates/representations of cognitive processes, are thought to reflect core concept dissociations. The majority of investigations into the neural correlates of EF have focused on the prefrontal cortex area, a direct consequence of the fact that the very construction of the concept of EF appears to have emerged from the study of the behavioural effects of damage to this region. However, Zelazo and Muller (2002) draw attention to evidence indicating that EF is not entirely independent of areas outside the prefrontal cortex. They note that EF also seems reliant on the normal functioning of additional regions such as those involved in the limbic system. Further, they observe that some people with damaged prefrontal cortex regions fail to demonstrate impaired EF, whilst conversely other people display executive deficits in the absence of injury to this area. Such findings indicate that studies focusing on the prefrontal cortex cannot provide a fully comprehensive account of EF. Additionally, it has to be recognised that the assumption that neural organisation is reflective of cognitive structure, which provides the rationale for applications of dissociationist methodology to investigations of cognition, is not uncontested (Ellis & Young 1988). Therefore, caution is encouraged in drawing conclusions from neuropsychological findings centring on the prefrontal cortex. Nonetheless, such data could still prove useful in suggesting possible conceptual distinctions.

In support of a non-unitary/multifactorial conception of EF, Zelazo and Muller (2002) suggest that neuropsychological studies focusing on the functions of two different regions within the prefrontal cortex - the orbitofrontal cortex and the dorsolateral prefrontal cortex, would appear to necessitate appreciation of a conceptual distinction between "hot" versus "cool" EF. The authors contrast findings which seem to indicate that damage to the dorsolateral prefrontal cortex impairs the ability to demonstrate adequate EF in traditional, unnatural experimental scenarios they conceive of as "cool"/detached, with data which appears to suggest that damage to the orbitofrontal cortex impairs the ability to demonstrate adequate EF in more involved, naturalistic circumstances which they conceive of as "hot"/emotionally-engaging. More specifically Zelazo and Muller (2002, p.455) propose that "whereas cool EF is more likely to be elicited by relatively abstract, decontextualised problems, hot EF is required for problems that are characterised by high affective involvement or demand flexible appraisals of the affective significance of stimuli".

Additional support for a compound view of EF would appear to be provided by observations highlighted by Miyake et al. (2000), of patients with brain-damage. These studies appear to indicate an independence of performance on alternative EF tasks in that the same individual will perform poorly on one task thought to tap EF, whilst performing well on a different task also thought to tap EF (Godefroy, Cabaret, Petit-Chenal, Pruvo, & Rousseaux, 1999; Shallice, 1988). Such task independence would appear to indicate the tapping of alternative differentiable EF skills.

The conception of a multi-faceted EF structure would seem to be lent further credence by the small degree of correlation typically found between different EF tasks, when presented to participants in batteries (Miyake et al. 2000). Low correlations would appear to indicate a lack of common function involved across the various EF tasks. Building on the correlational data, further support for a compound view of EF is provided by the application of factor analytical statistical procedures (Miyake et al., 2000; Zelazo & Muller, 2002). Such techniques exploit inter-task correlations. Zelazo and Muller note that developmental studies employing such procedures (e.g. Levin et al., 1991; Welsh, Pennington, & Grossier, 1991; Hughes, 1998a; Espy et al., 1999; Carlson & Moses, 2001; Pennington, 1997) typically identify three to four distinguishable abilities with various labels such as Motor Inhibition, Set Shifting, Verbal Working Memory (Pennington, 1997) or Attentional Flexibility, Inhibitory Control, Working Memory (Hughes, 1998a).

However, like the neuropsychological method, the application of the factor analytic approach to the study of EF is subject to criticism. The validity of the inter-task correlational relationships, on which factor analysis depends, has been called into question. Concerns have been raised about the confounding role of performance factors such as language and visuo-spatial abilities (Miyake et al., 2000; Zelazo & Muller, 2002). For example, it could be that low correlations between performance on EF tasks are more reflective of differences between the language demands of the tasks rather than differences in the nature of EF required by these tasks. Further, Miyake et al. point out that many EF tasks demonstrate low reliability, a factor they suggest might arise as a result of participants employing alternative tactics to solve tasks on different occasions, or as a reflection of the finding that EF is invoked more strongly in unfamiliar contexts. Miyake et al point out that the significance of this lies in the fact that variables which demonstrate poor reliabilities necessarily yield low correlations, since any relationship with another variable will be inconsistent and so cannot be strong. Additionally, questions have been raised regarding the conceptualisation of the constructs emerging through the process of factor analysis (Miyake et al., 2000; Zelazo & Muller, 2002). For example, since consensus is lacking regarding the abilities tapped by the individual EF tasks employed in the batteries in the first place, from whose data the factors are abstracted, it would seem that there can be little sense in which the factors abstracted can be noncontentiously identified.

To overcome some of these problems, Miyake et al. (2000) focused on latent variables in their factor analysis of EF abilities. The individual tasks employed in the EF battery used for their study were carefully selected with 3 literature-informed, theoretically distinct EF abilities in mind: the ability to inhibit dominant or pre-potent responses, the ability to shift between mental representations and the ability to update mental representations. The commonality among the collection of tasks selected to reflect/measure each ability was then extracted as a latent variable, and the relation of each latent variable to the other latent variables examined to investigate variable/factor differentiation. By opting for a number of battery tasks which reflect the same specific pre-selected ability, and extracting the commonality amongst them, Miyake et al. reduced, although by no means removed, the extent of performance factor confounding. Further, participants were also presented with additional complex EF tasks in which the specific EF abilities being tapped was less transparent. Examination of the relationships between the latent variables and performance on these complex EF tasks, enabled Miyake et al. to provide some clarification regarding the abilities tapped by these more controversial EF tasks.

Miyake et al. reported that inhibition, shifting and updating EF abilities correlated reasonably with each other, supporting the idea of an underlying unification of EF. However, in support of the multifactorial view of EF, confirmatory factor analyses indicated that these abilities were plainly distinguishable. A CFA model which did not constrain the correlations between the 3 EF latent variables was found to provide a better fit for the data than models which assumed perfect correlations among 2 or 3 of these variables. Furthermore, the differential contribution of these variables to the more controversial, complex EF tasks presented, lends additional support to a compound EF construct. Friedman et al. (2006) replicated Miyake et al.'s findings, although they neglected to investigate the contribution of the 3 latent EF variables to more contentious EF tasks. Thus, strong support has been provided for a model of EF reflecting both a common and a differentiated structure. However, much of the research mentioned so far has focused on the adult rather than the child population.

2.2 Development of executive functioning in young children

Hughes and Graham (2002) suggest that the traditional adult-focus of EF research can be attributed to: initial beliefs concerning the late maturation of the EF-associated prefrontal cortex, original indications that the deficits resulting from childhood prefrontal lesions did not become evident until adulthood, and the complexity of standard adult tasks resulting in floor performance during childhood. However, they note that reports that EF deficits are associated with developmental disorders such as Autism and Attention Deficit Hyperactivity Disorder (ADHD) have encouraged investigation into EF in atypically developing child populations, which have in turn spurred EF research in typically developing children. Hughes and Graham also point out that investigation of EF in young children reduces the effects of some of the confounds thought to affect research in the adult population. They argue that the problem of performance factors which reflects the fact that many adult EF tasks present considerable non-EF demands is diminished by the task simplification, necessary for presentation to a younger, less competent child audience. Furthermore, Hughes and Graham suggest that the issue of low task reliability, which it is thought might reflect the fact that EF is invoked more strongly in unfamiliar contexts, is less likely to be a problem for children. This is because children tend to take longer to become accustomed to task frameworks and so will usually require more presentations of a task than adults before they cease to classify the task as novel/unfamiliar. The development of EF during childhood is thus currently a thriving research topic (Zelazo, Carlson & Kesek, in press). However, Hughes and Graham acknowledge that research with young children presents its own set of problems, such as those stemming from language limitations.

EF research with young children has often employed simplified versions of adult EF tasks such as the Stroop task and the Wisconsin Card Sort task. The Stroop task (Stroop, 1935) requires suppression of an automatic response, so that response instructions can be obeyed. This enables assessment of the inhibition EF component, which is indicated by response error and response time. When the Stroop task is used with adults, the automatic response that is commonly evoked involves the reading of word stimuli, whilst response instructions require labelling of the ink colours in which the stimuli are printed (e.g. McKenna & Sharma, 1995). Correct responses are thus thought to require inhibition of automatically accessed lexical information. However, by evoking an alternative automatic response, this task has been adapted for use with young children, many of whom cannot read. Wright et al. (2003) have modified the task by presenting animal stimuli in which the body and head either belong together (e.g. a cow's body with a cow's head) or are mismatched (e.g. a cow's body with a duck's head). The use of facial information in identification is thought to be the automatic response evoked (Quinn & Eimas, 1996), whilst the response instructions require identification on the basis of the body presented. Correct responses are thus thought to require inhibition of automatically processed facial information.

The Wisconsin Card Sort task (Grant & Berg, 1948) requires moving between different mental representations of the same stimuli. This task has been used to assess what has come to be termed as the shifting component of EF. The standard adult task presents participants with a selection of target cards depicting stimuli differing in dimensions e.g. colour, shape and quantity (as in 1 green star, 2 yellow circles, 3 red squares). Participants are required to sort a set of cards depicting the same dimensions but in conflicting arrangements (e.g. 1 yellow square, 2 red stars, 3 green circles), in accordance with the target cards. They are not explicitly told the sorting rule (e.g. shape, colour or number), and so must infer it from feedback on previous sortings. After a set number of correct sorts (e.g. 10), the sorting rule is changed and the participant must attempt to infer the new sorting rule. Rule change is not announced, but must be inferred from negative feedback in response to sorting in line with the previous rule. Sorting rules are changed a number of times (e.g. colour to shape to number to colour to number to shape), to enable assessment of the ease with which participants can shift their focus from one sorting dimension to another. Shifting ability is indicated by the number of perseverations following change of the sorting rules. One commonly used adaptation of this task for children is the Dimensional Change Card Sort task (DCCS) of Zelazo et al., (1996). Children given the DCCS are explicitly told a sorting rule and are alerted to a rule change. They are still required to shift between their mental representations of the stimuli (in terms of shape, colour or number), so the adapted task remains a measure of shifting ability. However, explicit instruction of the sorting rules reduces the inferential demands of the task, to accommodate childhood limitations in inferential processing (Casteel & Simpson, 1991).

However, other tasks used to assess EF in childhood are often created expressly for this younger population. The A not B error task (Piaget, 1954), which is now commonly regarded as one of the earliest measures of EF (Lehto et al., 2003), was created specifically to address a problem with object permanence demonstrated by infants between 7 and 12 months of age. In the standard version of this task, infants watch whilst an object is hidden under an opaque cover at Location A. Following a short delay, they are encouraged to retrieve the object. The hiding process is repeated, a number of times, with infants retrieving the object each time from its hiding place at Location A. Following several successful retrievals from Location A, the object is then hidden under an opaque cover at Location B.

Piaget originally designed the task to demonstrate that young infants fail to create adequate mental representations of objects. He suggested that infants failing to locate the object in Location B were doing so because their reaching to Location A formed part of their mental representation of the object. However, looking preference data indicates that young infants can locate the object in Location B when the reaching motor response requirement is removed (Baillargeon et al., 1989). Such findings suggest that the standard reaching task does not accurately assess children's ability to form adequate object representations. Nevertheless, the task would appear to tap the updating and inhibition EF components as it requires infants to update their mental representation of the object's changing location, and inhibit their pre-potent Location A expectations.

The Freeze Frame task (Holmboe et al., 2008) was created with the specific intention of measuring inhibitory ability in 9-month-olds. In this task, infants have to learn to selectively inhibit attention to distractors in their peripheral vision, to retain a stimulus in the centre of their vision. Inhibition must be selective, and so must not applied when boring stimuli are presented in the centre of the infant's visual field, to ensure that it is self-controlled rather than a passive habituation effect. Furthermore, inhibition must be learned across the course of trials, rather than demonstrated in the early trials, to indicate that it is self-controlled rather than the result of an automatic competition between the central stimulus and distractors. Performance on the Freeze Frame task at 9 months is correlated with that on another inhibitory measure at 24 months: the Spatial Conflict task. This result suggests that the Freeze Frame task is tapping the inhibition EF component.

EF performance appears to improve with age during childhood (Gerstadt, Hong, & Diamond, 1994; Zelazo, Frye, & Rapus, 1996; Luciana & Nelson, 1998; Rueda et al., 2005), as one might expect given the development of the EF associated prefrontal cortex occurring during this stage of life (Huttenlocher, 1979; Huttenlocher & Dabholkar, 1997). However the nature of this development is open to dispute. Whilst evidence in the adult population appears to support a fractionated view of EF, it could be that EF in early childhood is a unified construct, only becoming dissociable into components at a later stage. Furthermore, if EF does appear to consist of independent components in childhood, the separable abilities might mature at the same rate or follow independent trajectories.

There is considerable evidence to support the conceptualisation of EF as a fractionated construct during childhood. Factor analytic studies have used patterns of shared variance within data drawn from task batteries to identify separable conceptual components. Espy, Kaufman, McDiarmid and Glisky (1999) employed an exploratory factor analysis to construct an EF model, with children aged between 2- and 5 and 1/2-years-old. Their findings led them to conclude that EF could best be conceptualised as a differentiable 4 factor construct at this stage in development. Following Miyake et. al (2000), Lehto, Juujarvi, Kooistra and Pulkkinen (2003) investigated childhood EF using a confirmatory latent variable factor analytic approach. An exploratory factor analysis was used initially to establish a 3 factor model of EF. The shared task variance reflecting each factor formed latent variables which were then investigated using a confirmatory factor analysis. The analysis compared the fit of the 3 latent variables/factors to the data, with alternative EF models. Lehto et al. concluded that EF in children aged between 8- and 13-years-old could best be viewed as a separable but interrelated 3 factor construct resembling the updating, inhibition and shifting model proposed by Miyake et al. (2000).

Furthermore, the suggestion that separable EF components may follow diverse developmental trajectories provides support for the view that EF manifests a differentiated structure during childhood. Hongwanishi, Happaney, Lee and Zelazo (2005) found evidence of the distinction between "hot" affective and "cool" decontextualised EF, found in adults, in 3- to 5-year-olds. Hongwanishi, et al. presented children with tasks selected to tap either "cool" or "hot" EF. Cool EF tasks comprised the DCCS sorting task and a Self-Ordered Pointing task requiring children to avoid pointing to stimuli they had previously pointed to. Hot EF tasks comprised a delay of gratification task assessing children's ability to delay receiving an immediate reward to achieve a greater reward at a later time-point and a gambling task measuring their ability to recognise that long-term prosperity was more likely to follow from smaller short-term gains and losses than greater immediate gains and losses. Simple correlational analyses revealed that performance on "cool" EF tasks was strongly positively related to verbal and performance mental age. The relationship with verbal mental age remained even when chronological age was partialled out. Furthermore, performance on "cool" tasks appeared to be related to temperament. However, performance on the "hot" EF tasks did not appear to be related to mental age or temperament. The difference in the relationships that the two EF components demonstrated with mental age and temperament, supports the conceptual differentiation of "hot" and "cool" EF. Moreover, the finding that only "cool" abilities were related to mental age, suggests that "hot" abilities might emerge earlier than "cool" abilities.

Espy and Bull (2005) also used simple correlational analyses to investigate a different potential EF division in 3- to 6-year-olds: the relationship between the ability to inhibit prepotent responses and the ability to shift between active and inactive mental rules/response sets. Espy and Bull reported that performance on tasks reflecting the ability to inhibit was unrelated to working memory span, whereas performance on tasks tapping the ability to shift was better in children with greater memory spans. Since memory span is known to improve with age (e.g. Dempster, 1981) Espy and Bull's findings suggest that the ability to inhibit prepotent responses might developmentally precede the ability to shift mentally (Blair, Zelazo & Greenberg, 2005).

Huizinga, Dolan and van der Molen (2006) subjected childhood EF to a confirmatory latent variable factor analysis. The results of Huizinga et al.'s investigation indicate that not only can childhood EF be differentiated into 2 separable constructs corresponding to updating and shifting but that these component parts mature at different rates. Huizinga et al. examined EF in a sample aged from 7- to 21-years-old. They found that whilst the ability to shift peaked during the teenage years, updating ability continued to improve past this age, with the updating ability of the 21-year-olds eclipsing that of the 15-year-olds.

However, a more recent latent variable factor analysis has indicated that contrary to the findings reported above, childhood EF can best be modelled by a single factor construct. Wiebe, Espy, and Charak (2008) examined EF in children aged between 2 and 6 years. They found that multifactor models did not account for significantly more variance in EF score, than a univariate model. Preference for parsimony led them to conclude that a univariate model provides the best account of childhood EF. Latent variable factor analyses provide more stringent assessments of conceptual structure than standard correlational or factor analyses, providing a possible explanation for the discrepant findings of Espy et al. (1999), Hongwanishi, et al. (2005) and Espy and Bull (2005). However, this explanation cannot extend to the work of both Lehto et al. (2003) and Huizinga et al. (2006) who employed latent variable factor analyses. A possible explanation for the discrepancy between the models of Wiebe et al. and Huizinga et al. is that the children in the Huizinga et al. study were considerably older: between 7 and 21 years, than were the children in Wiebe et al., who were between 2 and 6 years of age. It could thus be that EF initially emerges as a unified construct that only later fractionates following the selective taxation of particular functions. Wiebe et al. suggest that Lehto et al.'s use of the same data set to both derive and confirm the conceptual structure of EF severely weakens the validation of their 3 factor model, providing a plausible account for the discrepancy between the models of Wiebe et al. and Lehto et. al.

Thus, although it has been established that children's performance on EF tasks improves with age, the nature of the construct emerging is very much a current matter of debate, with both fractionated and unified stances receiving strong support. Indeed, support for a fractionated account provokes an additional concern regarding the independent development of the component parts. In their review of research into childhood EF, Garon, Bryson, and Smith (2008) investigated the development of EF components based upon the 3-factor framework presented by Miyake et al. (2000). Garon et al. concluded that there were two main stages of EF development. The first stage was seen to comprise an initial period before 3-years-of-age, during which the skills reflected in the component parts (inhibition, working memory and shifting) emerged. The second stage, after 3-years-of-age, was thought reflect a period of integration, during which the different component parts begin to establish connections between themselves, allowing children to coordinate the differentiated abilities.

Furthermore, the driving force behind development, also requires consideration. Rueda et al. (2005) looked at the effects of age, genetics and training on EF in children aged between 4 and 6 years. Cheek swabs were used to collect DNA samples which were examined with regard to the length of the alleles present in the DAT1 gene. Training was conducted over the course of 5 days and comprised computer tasks honing children's ability to anticipate, discriminate stimuli and resolve conflict. Results confirmed that EF improves naturally with age, and can be improved by using specialised training techniques, a procedure employed in Experiment 4 of this thesis. However the results also revealed that genetic factors played a prominent role in determining the progression of EF. Children whose DAT1 genes were made up of 2 long alleles were better able to resolve conflict that those whose DAT1 genes comprised 1 long and 1 short allele or 2 short alleles.

2.3 Executive functioning and theory of mind

It has been proposed that the development of EF might account for the emerging expression of CU in childhood. However, there is also evidence that EF development is linked to the expression of ToM ability, which may be seen as integral to CU. Many studies appear to support the suggestion that ToM is typically related to EF (Carlson & Moses, 2001; Davis & Pratt, 1995; Frye et al., 1995; Hala, Hug, & Henderson, 2003; Hughes, 1998b). Although, evidence that impairments in one ability do not inevitably lead to impairments in the other has cast doubt on the necessity of the association (Bach, Happé, Fleminger, & Powell, 2000; Rowe, Bullock, Polkey, & Morris, 2001; Pickup, 2008). Hughes and Graham (2002) have proposed that research supporting the EF-ToM relationship, can be divided into three main categories. One such category comprises studies indicating that populations who suffer a deficit in either EF or ToM tend to be impaired in the other. For example, Ozonoff et al. (1991) and Zelazo et al. (2002) reported this type of association in cases of children with autism. Another category consists of studies which show that typically developing children demonstrate improvements in both EF and ToM within the same time period: between 3 and 5 years of age (e.g. Diamond & Taylor, 1996; Frye et al., 1995; Wellman, Cross, & Watson, 2001). A final category put forward, consists of studies in which individual differences on measures of EF and ToM appear to be correlated (e.g. Carlson & Moses, 2001; Davis & Pratt, 1995, Hughes, 1998a).

However, as Hughes and Graham (2002) point out, there is considerable contention regarding the causal player in the EF-ToM relationship. One major argument is that EF promotes ToM. This view is supported by evidence that reducing the salience of reality, and so the inhibition required to overcome it, improves ToM in three-year-olds (Mitchell & Lacohee, 1991). Such a stance is also lent support by studies demonstrating that increasing the inhibitory demands of false belief tasks, dramatically reduces the percentage of 4-year-olds passing them (Leslie & Polizzi, 1998), and that the inhibitory ability of 3- and 4-year-olds predicts their ToM a year later, but the reverse relation does not hold (Hughes, 1998b). However, Hughes and Graham point out that implicit ToM, as evidenced by anticipatory eye movements concordant with expectations of false belief (Clements & Perner, 1994), and demonstrated by children as young as 15-months-old (Onishi & Baillargeon, 2005), proves problematic for an account which holds that inhibition is first applied to deliberate action and verbal description (Perner & Lang, 1999). Nevertheless, they concede that inhibition may well initially develop at a more basic motor level, or may even have a more indirect impact on ToM by allowing access to situations which support ToM development such as pretend play.

Another view discussed by Hughes and Graham (2002), holds that ToM promotes EF. According to this account, the appreciation of one's own mental states afforded by ToM, promotes greater self-awareness and so enables better self-control. Such a position is used to account for the EF deficits observed in populations associated with impaired ToM such as the stereotypical behaviours displayed by people with autism (Carruthers, 1996), and the lack of self-control demonstrated by people with schizophrenia (Frith, 1992). However, such a proposal is weakened by the reported failure of early ToM to predict later EF. Hughes, (1998b) presented EF and ToM tasks to 3- and 4-year-olds at Time 1 and then represented the tasks to the same children a year later at Time 2, along with additional measures of EF and ToM.

EF at Time 1 was found to predict ToM at Time 2, suggesting that EF affects ToM development. However, ToM at Time 1 did not predict EF at time 2, indicating that ToM does not play a role in EF development.

Hughes and Graham (2002) note that a further account proposes that a third, external factor, explains the link between EF and ToM. According to this view, EF and ToM are similarly bound by restricted access to embedded rules of reasoning, in line with cognitive complexity and control theory (Zelazo, 1999), as both abilities require appreciation of "if-then" reasoning. For example, IF Sally is looking for her displaced marble in a false belief task, the displacement of which she failed to witness, THEN she will look in the location where she originally left it. Similarly, IF a lexical stimulus is seen in a stroop task, THEN the colour of the ink it is printed in should be identified, but the word spelt out by the stimulus should not be read.

Three recent studies sought to test the role of EF in ToM reasoning. Sabbagh, Shiverick and Moses (2006) compared the relation of EF to performance on tasks assessing appreciation of false beliefs, false photographs and false signs. Measures of false belief appreciation reflect an understanding of mental representation and so ToM proficiency. Tasks assessing appreciation of false photographs are framed in a similar presentational format to false belief tasks, but require only an understanding of photographic and so non-mental representation. Thus, false photograph tasks do not reflect ToM proficiency. Tasks assessing appreciation of false signs are also framed in a similar presentational format to false belief tasks. Furthermore, like the false photograph task, false sign tasks require only appreciation of non-mental representation and thus do not reflect ToM proficiency. However, like false belief tasks but unlike false photograph tasks, measures assessing appreciation of false signs are concerned with representations which purport to reflect the current state of reality. EF was found to predict performance on tasks assessing appreciation of false beliefs and false signs but not on tasks assessing appreciation of false photographs. This was taken to indicate that the relationship between EF and ToM was not exclusive but reflected a more general role which EF plays in reasoning about representations which purport to reflect the current state of reality.

Sabbagh, Xu, Carlson, Moses and Lee (2006) used samples of American and Chinese children to investigate the predictive accuracy of EF in relation to ToM. Batteries of both EF and false-belief tasks were presented to the children. Sabbagh et al. found that EF predicted false-belief appreciation within the American and Chinese samples, but failed to predict performance across groups. Most interestingly, Chinese children were found to display superior executive abilities compared to American children. However, although Chinese 3- and 1/2-year-olds demonstrated the same level of EF as that shown by American 4-year-olds who performed competently on the false-belief battery, these younger Chinese children failed to demonstrate a similar false-belief proficiency. Such results appear to indicate that whilst important and seemingly necessary for mature false belief reasoning, accounting for the within group association, EF is not sufficient. The across group disparity would seem to indicate that additional input is required to supplement ToM development. Sabbagh et. al. point out that the early experience of conflicting mental states thought to be promoted by sibling interactions, could form such additional input, given the contrasting prevalence of sibling experiences across America and China.

Such an account might also help to explain the findings of Oh and Lewis (2008), who investigated the relationship between EF and ToM in Korean 3- and 4-year-olds. Oh and Lewis found that Korean children performed at ceiling on EF tasks assessing inhibitory and switching ability, demonstrating greater competence than displayed by Western children of the same age. However, despite this, the working memory of the Korean children i.e. their ability to update and store mental representations, and their ToM performance were found to be on par with that of Western children of the same age. In other words the ToM performance of the Korean children was lower than expected given their inhibitory and shifting superiority. Closer examination of the relationship between the different components of EF and ToM, confirmed the failure to reveal strong EF-ToM links in the Korean sample. No EF-ToM relationship remained significant once language and age were accounted for. However, a significant correlation between ToM and shifting ability: indexed by performance on the DCCS task, remained after language and age had been taken into account, in the English comparison sample. It might well be that with adequate supplementary input such as sufficient experience of mental state conflicts, EF promotes ToM, as suggested by the EF-ToM relationship found in Sabbagh et al.'s Chinese and English samples and Oh and Lewis's English sample. However, it could be that without this input, the role of EF becomes redundant as in Oh and Lewis's Korean sample. In such instances, the emergence of ToM might be more dependent on alternative resources such as mental state language.

2.4 Conversational understanding and executive functioning: A working model

Relationships between CU and ToM, CU and EF and EF and ToM have been discussed, but it remains to present a unifying framework. It could be that a true appreciation of the masking effects of processing resources on CU during childhood, requires consideration of an interplay between the ToM skill of intention appreciation and EF. In young children, processing restrictions might affect the ability to reason about the mind, impairing children's ability to follow speakers' intentions. However, with increasing age, these restrictions could be overcome by the development of EF abilities. Indeed, Scholl and Leslie (1999) conceive of ToM as the product of a specialised module built into the brain which confers processing biases that predispose children to express a ToM understanding. However, they do not propose that children immediately display ToM proficiency. Rather, they argue that the mechanism used to identify the appropriate module-generated behaviour interpretation: the Selection Processor, establishes a default preference for true-belief interpretations. This is quite possibly because of the lower processing load associated with true belief interpretations: true beliefs directly correspond to reality and so can be represented in terms of reality, however false beliefs do not correspond to reality and so must be represented separately (Apperley, Back, Samson, & France, 2008). The development of EF might enable children to overcome the true-belief preference and so demonstrate false-belief appreciation.

Siegal and Surian (2004, 2007) suggest that something similar might be happening in the case of interpretation of speaker intentions. It might be that the evolutionary pressure for rapid language development has created a processing bias to select logical, linguistically determined speaker intentions in situations where words are made salient, such as when presented with directed communication (Friend & Bryant, 2000) or a rhyming context is used (Lee, Torrance & Olson, 2001). However, there might be an alternative bias to focus on contextual information in other instances where the context becomes salient, such as in story presentations (Olson & Hildyard, 1981; Lee, Torrance & Olson, 2001) or when the intended meaning is revealed directly (Robinson, Goelman & Olson 1983; Beal & Flavell, 1984). CU tasks in which children perform at chance levels, neither demonstrating CU nor an interpretation bias (e.g. Dews et al. 1996) might be explained in terms of an equal salience of words and context, provoking switches in bias between the two components of communication. Both biases are thought to be overcome as processing resources increase with age. More specifically, Siegal and Surian suggest like Leslie, that it is the development of an EF system that enables the child to overcome such processing biases. Such a resource is thought to enable the child to inhibit default interpretations, to allow access to contextually enriched but logically supported interpretations.

Chapter 3: Experiment 1 – Investigating the relationships between executive functioning, a narrow scalar implicature measure of conversational understanding and theory of mind.

3.1 Introduction

The literature review in Chapters 1 and 2 considered the CU of young children and focused on two conflicting models of emergence proposing to explain the expression of such knowledge. According to one account, the expression of CU results from a conceptual change reflecting the development of a metarepresentational ability: the capacity to represent messages as representations of a speaker's communicative intentions (Beal & Flavell, 1984; Robinson & Whittaker, 1986). This metarepresentational ability is thought to enable differentiation of the logical meaning (LM) of the message from the intended meaning (IM). Such differentiation is considered to give rise to the appreciation that the LM does not provide direct access to the IM and so prompt consideration of the LM in light of the context in which it is presented. This is a competence account, because it attributes the expression of CU to the development of a conceptual competence: metarepresentational ability.

However, studies indicate that reducing processing demands improves children's performance on tasks assessing CU (e.g. Papafragou, & Musolino, 2003; Beck, Robinson, & Freeth, 2008). This pattern of results supports a masked competence account that children possess the necessary conceptual competence for CU, i.e. the LM-IM distinction, from an earlier age than they typically demonstrate this ability. This account attributes the expression of CU to the application of EF abilities. EF is thought to enable the overpowering of processing restriction induced interpretation biases focusing on either LM or IM. This is a performance account, because it suggests that children have already acquired the conceptual competence necessary for CU, but that the expression of CU is attributable to a performance-enhancing factor: EF, permitting access to this knowledge. This account proposes that removing interpretation biases allows consideration of the logical information in communication, in light of the context in which it is presented.

To assess evidence for these alternative accounts, Experiment 1 investigated the relationship between developing EF abilities as conceptualised according to the structure proposed by Miyake et al. (2000), and a measure of CU. Furthermore, since both competence and performance accounts conceptualise CU in terms of the recognition of communicative intentions, invoking a role for ToM, a ToM measure was also included.

Nilsen and Graham (2009) found that children's ability to inhibit prepotent responses was related to their CU, but did not find a relationship between CU and their cognitive flexibility. Furthermore, the aspect of CU investigated by Nilsen and Graham was communicative perspective taking. Experiment 1 examined the EF-CU relationship further by considering the relationship between EF and another aspect of CU: the appreciation of SIs, and by using alternative EF tasks to examine the relationship between CU and cognitive flexibility and the ability to inhibit prepotent responses.

Experiment 1 was concerned with establishing the nature of the relationship between a developing awareness of SIs, appreciation of false beliefs, and proficiency in inhibiting attention and shifting between alternative mental representations. These variables represent a measure of CU, the predominant gauge of ToM ability, and two of Miyake et al's three EF components, respectively. In line with the general consensus regarding the development of competencies during childhood, it was expected that all abilities would improve with age. In accordance with a masked competence/performance account of CU, it was also hypothesised that appreciation of false beliefs, and both inhibitory and shifting abilities, would be related to awareness of SIs. Indeed, a relationship was expected between appreciation of false beliefs and both EF abilities themselves (Carlson & Moses, 2001; Perner, Lang, & Kloo, 2002). Furthermore, it was expected that the relationship between awareness of SIs and both inhibitory and shifting EF abilities, would be mediated by appreciation of false beliefs, as argued in the masked competence account. According to this account, the application of EF contributes to children's ToM performance as required for the recognition of others' mental states. It is thus through its effect on ToM that EF promotes CU. The use of two separate EF measures assessing the skill to inhibit and shift between mental representations, was undertaken to allow more refined conclusions to be drawn regarding the role of EF abilities in the development of both CU, and more general ToM appreciation.

Both the ability to inhibit, and the ability to shift between mental representations, could be implicated in more general ToM reasoning, with default true belief representations of situations being inhibited where necessary and shifting between true and false belief representations occurring when required. Similarly, CU could also draw on both skills, to enable the inhibition of default literal or context-led readings of the communication when needed and shifts between alternative interpretations as required (Siegal & Surian, 2004). Therefore, performance on both EF measures was expected to be related to false-belief appreciation and to awareness of SIs.

By adding new control conditions to the SI trials used in Papafragou and Musolino's (2003) Experiment 2, this study was also able to examine the validity of the SI trials employed in the earlier study. In the SI trials of the second experiment in

Papafragou and Musolino's study, children were introduced to a puppet, and watched with the puppet as a character completed a task such as a bear putting all of the available hoops on a pole and a girl putting all four pieces of a jigsaw puzzle together. The puppet then described the character as having completed "some" of the task, e.g., "the bear put some of the hoops on the pole", or "the girl put some of the pieces in the puzzle" and children had to evaluate the puppet's descriptions. The reasoning behind this task was that if children were computing the SI "some = not all", they should have negatively evaluated the puppets' descriptions in these SI trials, because each time the character completed all of the task, but each time the puppet only described them as having completed some of the task. However, if children were failing to compute this pragmatic implicature, and were interpreting the meaning of "some" as compatible with the meaning of "all", they would have been expected to have positively evaluated the puppets' descriptions.

Apparent CU competency in the SI trials used in this task could result from a negative response bias. Thus there might have been children who were failing to compute SIs, but who appeared to demonstrate conversational competency, merely due to a tendency to respond negatively in such contexts. Fritzley and Lee (2003) suggest that when presented with comprehensible statements, 4- and 5-year-olds fail to demonstrate a response bias. However they do contend that, when faced with incomprehensible queries, 4- and 5-year-olds in fact demonstrate a negative response bias. Although one would normally expect children of this age to understand the puppet descriptions given in Papafragou and Musolino's (2003) task (and indeed the findings in Papafragou and Musolino (2003), Experiment 1 support such a suggestion), the specific sample of children who participated in their Experiment 2 may have experienced difficulty in comprehending the SI descriptions. It might be that the sample in Papafragou and Musolino's Experiment 2 experienced an enhanced difficulty comprehending the SI descriptions compared to the sample in Papafragou and Musolino's Experiment 1. If so, a negative response bias rather than the difference in the presentation context, may have resulted in negative evaluations of the puppets' descriptions.

Papafragou and Musolino (2003) attempted to control for the possibility of a negative response tendency by employing control descriptions which required positive evaluation. For example, a puppet would describe a situation in which a little girl jumped over a fence, with the statement "The little girl jumped over the fence". Such an approach can highlight response biases, because many factors are held constant across SI and control trials. Thus if children are simply responding negatively to SI descriptions due to an irrelevant aspect of the task (e.g. the fact that the agent giving the descriptions is a puppet), they might also respond negatively and so inappropriately in the control trials, highlighting the irrelevance of their responses to the task purpose. Papafragou and Musolino found that children were able to demonstrate response flexibility by evaluating negatively in the SI trials as required, but also positively in the control trials as appropriate. They argued that this supported their claim that performance on the implicature trials reflected a genuine ability to compute SIs.

However the control trials differed from the SI trials in Papafragou and Musolino's (2003) second experiment in their linguistic content. Whilst the SI trials used the implicature-relevant term "some", the control trials of the second experiment neglected to incorporate this term. It could have been that something specific to the term "some" was provoking children to evaluate the puppet's description negatively. Papafragou and Musolino incorporated the term "some" in the control conditions for the SI trials in their first experiment, and found children able to evaluate felicitous descriptions using the term "some", appropriately. However the context in which the puppet descriptions were presented was modified for Experiment 2, to promote attention to SIs, and this dramatically increased children's negative evaluation of descriptions incorporating the term "some" during the SI trials. It is feasible that the change of presentation context could have had a similar impact on the evaluation of descriptions incorporating the term "some" during control trials as the children may have had some awareness of the unusual nature of the SI terms as highlighted in the new presentation context. It might have been that the SI nature of the "some" term was prompting children to evaluate its use negatively, regardless of the context in which it was used.

Experiment 1 sought to rule out this possibility by employing control trials which required children to positively evaluate descriptions incorporating the term "some". These trials presented scenarios in which a character only completed some of their task, for example a bear put only three of five available hoops on a pole, and the puppet described the character as having "put some of the hoops on the pole." Positive evaluation of such control descriptions, in the face of negative evaluations of statements given during scalar trials, would support Papafragou and Musolino's suggestion that children rejecting "some" descriptions in the scalar trials, were actually doing so on the basis of having drawn implicatures which made their meaning inconsistent with the situation they were supposed to be reflecting.

Additional sets of control trials were incorporated, which required positive and negative evaluation of descriptions using the term "all" instead of "some". For example, the description "The bear put all of the hoops on the pole" was used in a situation in which the bear had put three of five available hoops on the pole and also in a situation in which the bear had put all five available hoops on the pole. These were employed to ascertain children's understanding of the term implicated as not applicable in the descriptions of the scalar trials – the term "all". By adding these additional control trials, children's understanding of both quantity terms relevant to the "some implies not all" SIs employed in this study was investigated. In addition, this procedure allowed for an examination of response flexibility so that children receiving the SI test trials and three sets of control trials were presented with two sets of trials requiring positive evaluations and two sets requiring negative evaluations. Furthermore, if performance on the SI trials was really reflective of children's ability to compute SIs, and not instead indicative of the demands of more general processing costs, children could be expected to perform significantly better in all of the control trials than in the SI trials. This is because the scalar trial processing demands attributable to the linguistic content of puppet descriptions, the situations they were supposed to be reflecting and the children's evaluative responses were shared with the control trials. The only notable difference between the scalar and control trials arose from the cross-matching of the term "some" with a context in which the term "all" would have been more appropriate, which emphasised the significance of drawing the "some-not all" SI.

3.2 Method

Participants. These were 75 children aged between 4- and 6-years-old split into three age groups: 25 4-year-olds (M = 54.88 months, range = 48-59 months), 27 5-year-olds (M = 66.44 months, range =60-71 months) and 23 6-year-olds (M = 75.13 months, range = 72-78 months). These children were predominantly white and middle class, and were recruited from a nursery and a primary school in Sheffield. All children spoke English as their first language, and none were known to have any specific language impairment.

Procedure. Children were tested individually, in a quiet area of their school, and all took part in five tasks. Testing was conducted over the course of two sessions (each lasting approximately 15-20 minutes), and the order of task presentation fixed. The first session commenced with the Dimensional Change Card Sort (DCCS) Shifting EF task followed by the Day-Night, Inhibitory EF task and two Unexpected Location False-Belief ToM tasks. The second session consisted of the presentation of the SI task.

DCCS Task. The DCCS Task described by Zelazo (2006) was used to assess the ability to shift between mental representations, conceptualised as an EF ability by Miyake et al. (2000). This task required the child to alter their focus between the object depicted on a set of cards and the colour of ink used to depict it, in accordance with sorting instructions. Focus alternation is thought to enable the appropriate re-categorisation of cards thereby permitting correct sorting in accordance with instruction. Two trays, each displaying a target card cross-matched with the test cards, were used for this task. One target card displayed a blue bird and the other, a red car (see Appendix Ia).

Fourteen test cards were employed in this task (see Appendix Ib), two of which were used in a demonstration phase. Seven of the test cards depicted a blue car and seven depicted a red bird. Preliminary questioning requiring animal and colour labelling was used to establish that the children could appropriately distinguish between the two colours and the two objects depicted. In the demonstration phase, the experimenter introduced the task by labelling the target cards "Here is a blue bird and here is a red car." The experimenter then proceeded to encourage the child to sort either by colour or by shape, the choice of which was counterbalanced across children. The experimenter told the child, "Now we're going to play a card game. This is the colour/shape game. In the colour/shape game all the blue ones/birds go here" (pointing to the tray with the blue bird target card) and all the red ones/cars go here" (pointing to the tray with the red car target card). Only the relevant dimension of the picture on the card was labelled. The experimenter then proceeded to sort one type of test card, e.g., a card displaying a blue car, by the relevant dimension saying "See, here is a blue one/car. So it goes here", placing the card face down in the tray with the blue/car target card. The rules were then repeated. The child was then shown the other type of test card, e.g., a card displaying a red bird, and the experimenter said, "Now here is a red one/a bird. Where does this one go?" If the child took the card and sorted it correctly, or simply indicated the correct tray by pointing, the experimenter gave positive feedback. If they pointed, the experimenter asked, "Can you help me put this red one/bird down?" The experimenter ensured that the card was placed face down in the appropriate tray, turning the card over if necessary. If the child sorted incorrectly, the experimenter gave negative feedback, and again ensured that the card was placed face down in the appropriate tray. The experimenter then proceeded to the pre-switch phase.

On the first pre-switch trial, the child was told it was their turn to sort, and the previous rules were repeated. The experimenter then randomly selected a test card (e.g. displaying a red bird), showed it to the child, and labelled it by the relevant dimension only. The child was asked to place the card in the appropriate tray. No feedback was given following sorting and the experimenter moved onto the next pre-switch trial. There were six pre-switch trials in total. Cards were selected for sorting, ensuring that two identical test cards were not presented on more than two consecutive trials. When the pre-switch trials had been completed, the experimenter told the child that they were going to play a new game, and introduced them to the rules of the game they hadn't play before. "We are not going to play the colour/shape game anymore. We are going to play the shape/colour game. In the shape/colour game, all the birds/blue ones go here" (pointing to the tray with the blue bird target card) "and all the cars/red ones go here" (pointing to the tray with the red car target card). The child was then presented with cards which the experimenter identified using only the label relevant to the new sorting rules, and asked to sort cards which had been again selected to ensure that the same type of test card was not presented on more than two consecutive trials. There were six post-switch trials, and as in the ore-switch trials, no feedback was given following sorting.

Children needed to sort correctly in at least five pre-switch trials in order for their post-switch performance to be included (all children achieved this). The total number of cards correctly sorted in the post-switch phase was recorded as the score to be used in later data analysis with a possible range of 0-6.

Day-Night Task (D-N inhib). The Day-Night Task of Gerstadt, Hong and Diamond (1994) was used to assess attentional inhibition. This was also conceptualised as an EF ability by Miyake et al (2000). This task required inhibition of the standard labelling of pictures depicting the sun as "day", and pictures depicting the moon as "night", to allow converse labelling i.e. a picture of the sun labelled as "night". A set of two training cards and sixteen testing cards was used for this task. Half of these cards displayed a white crescent moon and stars on a black background, and half displayed a yellow sun and white clouds against a light blue background (see Appendix II). Children were shown a card with the moon on and told, "We're going to play a funny game. When you see this card I want you to say "day"," and asked to repeat the word "day". They were then shown a card with the sun on and told, "When you see this card I want you to say "night"," and asked to repeat the word "night". Following this, children were shown a card with a sun on and if they fail to respond appropriately, prompted with "What do you say when you see this card?" If children responded correctly they were praised and presented with a card with the moon on it. If children responded correctly again, they were again praised and these two trials were counted as trials 1 and 2 of the test phase, and testing continued from there.

The test phase involved presentation of cards in the following pseudorandom order sun(s), moon (m), m,s,m,s,s,m,m,s,m,s,s,m,s,m. No feedback was given during the testing phase. If a child got either of the first two trials wrong or failed to respond, these two trials were counted as practice and the child reminded of the rules and then presented with a sun card followed by a moon card. If the child responded correctly, these were then counted as trials 1 and 2 of the test phase, and testing continued from there; otherwise these trials were counted as further practice trials and the child reminded of the rules again. The total numbers of correct "day" or "night" responses were recorded as the score to be used in later data analysis with a possible range of 2-16.

Unexpected locations false belief tasks. Two unexpected locations false belief tasks were used to assess ToM based on the Sally-Anne task of Baron-Cohen, Leslie, and Frith (1985). The tasks were presented in a fixed order (the Sally-Anne task followed by the Max-Tom task). Each task involved two characters, two potential object locations and one object to be hidden, and was presented using small-world people and objects to re-enact story scenarios (see Appendix IIIa). In the Sally-Anne task, children were introduced to Sally, whom they were told had a basket and a marble, and Anne whom they were told had a box. Children then watched as Sally put her marble in a covered basket and left the scene. When Sally had disappeared from view, children were shown Anne taking the marble out of the covered basket and placing it inside a covered box. Children were then asked the test question "When Sally gets back, where will she look first for her marble?" and a control memory question "Where is her marble really?" in that order.

The Max-Tom task was similar to the Sally-Anne task, except that Max and Tom replaced Sally and Anne respectively, Max had some drawers and Tom a covered pot instead of Sally and Anne's basket and box, and Max hid a crayon instead of a marble (see Appendix IIIb). In the Max-Tom task, children were asked where Max would look first for his crayon, and where his crayon was really. Incorrect response to a control question meant the response to the test question in that task was automatically recorded as incorrect. The total number of correct responses to the two test questions was recorded as the score to be used in later data analysis, with a possible range of 0-2.

Scalar Implicature (SI) Task. An adaptation of the Some-All SI Task of Papafragou and Musolino (2003) was used to assess CU. Children were introduced to five hand puppets called: Ellie the Elephant, Harry the hedgehog, Suzy the Squirrel, Barry the Badger and Felicity the Fox, whom they were told "sometimes say some silly things". Children were told they were going to play a game with the puppets to "help them say things better". In an initial training phase with Ellie the Elephant, children were presented with four warm-up trials to encourage them to pragmatically assess the puppet's language.

The first warm up scenario presented a truthful and pragmatically felicitous (appropriate) statement – Ellie was shown a toy lion and asked "What is this Ellie?" to which the puppet gave a truthful, and pragmatically felicitous reply- "It is a lion." Children were asked if Ellie "answered well" and if they failed to reply or negatively judged the puppet's performance, the experimenter would comment that "Ellie did answer well because it is a lion". The second warm up statement presented children with a truthful but pragmatically infelicitous (inappropriate) statement – Ellie was shown a toy pig and asked "What is this Ellie?" to which the puppet gave a truthful, but pragmatically infelicitous reply, "It is an animal with four legs." Children were again asked if Ellie "answered well" and if they failed to reply or positively judged the puppet's performance, the experimenter would comment, "Ellie did not answer well because she could have said it was a pig." If children negatively evaluated the puppets responses to the infelicitous statement, they were asked, "How could she have said it better?" Two further warm up trials were given, presenting children with a further felicitous and an infelicitous statement, in that order. The details of these are given in Table 1. The felicity of the warm-up trial statements varied, whilst their logical value remained a constant: true, to indicate to the children that they needed to focus on the conversational appropriateness of statements rather than their logical value. The complete set of warm-ups is shown in Table 3.1.

Table 3.1: *Warm up trials for the SI items in Experiment 1*

Puppet statements on warm up scenarios	Scenario
Felicitous statements	
It's a lion	pointing to a toy lion
It's a duck	pointing to a toy duck
Infelicitous statements	
It's an animal with four legs	pointing to a toy pig
It's a yellow thing you eat	pointing to a toy banana

Following the warm-up statements, children were presented with the test scenarios. In each of the test scenarios a puppet watched whilst a toy character performed a task. For example, the experimenter would say, "This is Teddy. He thinks he is very good at putting the hoops on the pole. Shall we see how he does?" Then the experimenter manipulated the toy character to attempt a task. After watching the task attempt, the puppet would be asked how the character did, "Lets ask Harry how Teddy did. Harry, how did Teddy do?" The puppet would then provide a statement containing either the descriptive term "all" or "some", and children would be asked whether the puppet answered well. If children indicated that the puppet had not answered well, they would be asked how the puppet could have "said it better". All children were presented with four different types of test trials. Sets 1-3 were control trials, used to establish response flexibility, and to gauge whether performance on the SI trials in Set 4 was truly reflective of the ability to compute SIs, or arose from more general processing difficulties. Set 4 was the critical set, containing the SI trials. In Set 1 trials, the puppet used the term "all" in a semantically true and pragmatically appropriate way: for example the bear would put all of the available hoops on the pole and the puppet would describe him as having put "all of the hoops on the pole" (see Appendix IVa). Children were expected to provide "Yes" responses when asked whether the puppet described the characters' performances well in this set of trials. In Set 2 trials, the puppet used the term "all" in a semantically false and pragmatically inappropriate way: for example the bear would put only some of the available hoops on the pole and the puppet would describe him as having put "all of the hoops on the pole" (see Appendix IVb). Children were expected to provide "No" responses when asked whether the puppet described the characters' performances well in this set of trials.

In Set 3 trials, the puppet used the term "some" in a semantically true and pragmatically appropriate way: for example, the bear would put only some of the available hoops on the pole and the puppet would describe him as having put "some of the hoops on the pole" (see Appendix IVc). Children were expected to provide "Yes" responses when asked whether the puppet described the characters' performances well in this set of trials. In Set 4 trials, the puppet used the term "some" in a semantically true but pragmatically inappropriate way: for example the bear would put all of the available hoops on the pole and the puppet would describe him as having put "some of the hoops on the pole" (see Appendix IVd). These were the critical trials assessing the ability to compute SIs. Competent children who were computing the SI ("some implies not all") were expected to provide "No" responses when asked whether the puppet described the characters' performances well in this set of trials. Children who failed to compute the "some = not all" SI were expected to see the meaning of some as consistent with the meaning of all, and so provide "Yes" responses when asked whether the puppet described the characters' performances well in this set of trials. A complete list of these trial statements is presented in Table 3.2.

The presentation of the four sets in the SI task was counterbalanced across children in the 24 possible orders. A different puppet was used to describe the character's performance in each set of trials within each sequence, and ascription of puppets to trial sets was counterbalanced across children. Only two different puppet ascriptions were applied, so for half of the children Harry the Hedgehog gave the descriptions in Set 1, Barry the Badger gave the descriptions in Set 2, Felicity Fox gave the descriptions in Set 3 and Suzy Squirrel gave the descriptions in Set 4. For the other half of the children, Suzy Squirrel gave the descriptions in Set 1, Felicity Fox gave the descriptions in Set 2, Barry the Badger gave the descriptions in Set 3 and Harry the Hedgehog gave the descriptions in Set 4. Presentation of the trials within each set was also counterbalanced using two different orders. For half of the children assigned to one version of the puppet ascriptions, the trials were presented as they read downwards in Table 3.2. For the other half of the children in this version of puppet ascriptions, the trials were presented as they read upwards in Table 3.2. Similarly, for half of the children assigned to the other version of the puppet ascriptions, the trials were presented as they read downwards in Table 3.2. For the other half of the children in that version of puppet ascriptions, the trials were presented as they read upwards in Table 3.2.

The total number of correct responses to enquiries regarding whether the puppets answered well was recorded and translated into a score to be used in later data analysis with a possible range of 0-16. For Sets 1 and 3, each "Yes" response contributed 1 point to the final score and each "No" response was scored as 0. For Sets 2 and 4, each "No" response in these sets contributed 1 point to the final score and each "Yes" response was scored as 0.

Table 3.2: *Test trials for the SI task in Experiment 1*

Puppet statements on test trials	Scenario referred to
Set/Control 1	
He put all of the hoops on the pole	-Teddy bear puts all of the hoops on the pole
He lifted all of the fishes	-Tiger lifts all of the fishes up
She put all of the pieces in the puzzle	-The little girl puts all of the pieces of the puzzle together
He caught all of the horses	-The little boys catches all of the horses
Set/Control 2	
He put all of the hoops on the pole	-Teddy bear puts some of the hoops on the pole
He lifted all of the fishes	-Tiger lifts some of the fishes up
She put all of the pieces in the puzzle	-The little girl puts some of the pieces of the puzzle together
He caught all of the horses	-The little boys catches some of the horses
Set/Control 3	
He put some of the hoops on the pole	-Teddy bear puts some of the hoops on the pole
He lifted some of the fishes	-Tiger lifts some of the fishes up
She put some of the pieces in the puzzle	-The little girl puts some of the pieces of the puzzle together
He caught some of the horses	-The little boys catches some of the horses
Set 4/Scalar Implicatures	
He put some of the hoops on the pole	-Teddy bear puts all of the hoops on the pole
He lifted some of the fishes	-Tiger lifts all of the fishes up
She put some of the pieces in the puzzle	-The girl puts all of the puzzle pieces together
He caught some of the horses	-The little boys catches all of the horses

3.3 Results

Score means and standard deviations are presented in Table 3.3.

Table 3.3: *Mean scores and standard deviations for EF, ToM and CU tasks employed in Experiment 1.*

	4-year-olds	5-year-olds	6-year-olds
N	25	27	23
Age in months	54.88 (3.00)	66.44 (3.32)	75.13 (2.16)
D-N inhib EF/16	10.20 (3.55)	11.93 (3.49)	12.70 (3.82)
DCCS shift EF/6	4.08 (2.86)	5.85 (0.77)	5.52 (1.59)
FB ToM/2	1.24 (0.72)	1.52 (0.58)	1.91 (0.29)
Set 4 SIs/4	1.84 (1.93)	2.04 (1.91)	2.96 (1.58)
Set 1 Control/4	3.80 (0.50)	3.93 (0.27)	3.83 (0.49)
Set 2 Control4	2.72 (1.79)	3.07 (1.36)	3.17 (1.44)
Set 3 Control4	3.48 (1.16)	3.67 (0.96)	3.70 (0.56)

Note: D-N inhib=Day-Night task; DCCS=Dimensional Change Card Sort task; FB=False belief tasks; Set 4 SIs= Set 4 Scalar implicature trials employed in the scalar implicature task, Set 1 Control= Set 1 control trials employed in the scalar implicature task; Set 2 Control= Set 2 control trials employed in the scalar implicature task; Set 3 Control= Set 3 control trials employed in the scalar implicature task.

Relationship between EF and SI computation

Children's inhibitory EF was classed as strong if their scores on the Day-Night task were 12 or greater out of 16 and weak if these were 11 or less. The weak and strong division criteria were selected on the basis that they enabled as equal a split as possible in the number of children assigned to the two EF ability groups. A 2 (inhibitory EF score: weak vs. strong) x 3 (age group: 4-year-olds vs. 5-year-olds vs. 6-year-olds) factorial analysis of variance was conducted on SI scores. N was roughly equal for weak versus strong inhibitory ability respectively across the 4-year-olds, 15:10, 5-year-olds, 12:15 and 6-year-olds, 8:15. This analysis did not reveal a significant effect of inhibitory EF ability on SI score, $F(1,69) = 1.145$, $p>0.20$ or an age interaction, $F(2,69) = 0.995$, $p>0.30$. However, there was a significant main effect of age group, $F(2,69) = 3.392$, $p<0.05$. Table 3.4 reports the means and standard deviations of SI scores for the weak and strong inhibitors across the age groups and Table 3.5 reports the ANOVA summary data. One-tailed t-tests revealed that the SI scores of the 6-year-olds were significantly greater than the SI scores of the 4-year-olds, $t(45.418) = 2.200$, $p<0.05$ and the 5-year-olds, $t(47.967) = 1.862$, $p<0.05$.

Table 3.4: *SI score means and standard deviations for the weak inhibitors (inhibitory EF score: <12) and strong inhibitors (inhibitory EF score ≥12) across the 4-year-old, 5-year-old and. 6-year-old age groups.*

Age group	Inhibitory group	Mean SI score	Standard deviation
4	Weak	2.13	1.96
	Strong	1.40	1.90
5	Weak	1.83	1.85
	Strong	2.20	2.01
6	Weak	3.63	1.06
	Strong	2.60	1.72

Table 3.5: *Summary table for the 2 (inhibitory EF score: <12/weak vs. ≥12/strong) x 3(age group: 4-year-olds vs. 5-year-olds vs. 6-year-olds) analysis of variance of SI scores.*

Source	SS	df	MS	F	p	Partial Eta²
Within	229.675	69	3.329			
Inhibitory EF	3.810	1	3.810	1.145	0.288	0.016
Age group	22.583	2	11.292	3.392	**0.039**	0.090
Inhibitory EF x age group	6.622	2	3.311	0.995	0.375	0.028
(corrected model)	26.512	5	5.302	1.593	0.174	0.103
(corrected total)	256.187	74				

Children's shifting EF was classed as strong if their scores on the DCCS were 6 out of 6 and weak if these were less than 6. Again, the weak and strong division criteria were selected on the basis that they enabled as equal a split as possible in the number of children assigned to the two EF ability groups. A 2 (shifting EF score: weak vs. strong) x 3 (age group: 4-year-olds vs. 5-year-olds vs. 6-year-olds) factorial analysis of variance could not be conducted on SI scores because there were not enough children with weak shifting EF in the 5- and 6-year-old age groups: the 5- and 6-year-olds appeared to be scoring near ceiling. However, N was roughly comparable across the weak versus strong shifting groups, 8:17 respectively for the 4-year-olds. Table 3.6 reports the means and standard deviations of SI scores for the weak and strong shifters in the 4-year-old age group. A t-test did not reveal a significant effect of shifting ability on SI score, $t(23) = 1.051$, $p>0.30$.

Table 3.6: *SI score means and standard deviations for the weak shifters (shifting EF score: <6) and strong shifters (shifting EF score =6) across the 4-year-old age group.*

Age group	Shifting group	Mean SI score	Standard deviation
4	Weak	1.25	1.83
	Strong	2.12	1.97

Relationship between ToM and SI computation

Children's ToM was classed as strong if their scores on the false-belief tasks were 2 out of 2 and weak if these were less than 2. The weak and strong division criteria were selected on the basis that they enabled as equal a split as possible in the number of children assigned to the two ToM ability groups. The 6-year-olds were performing at ceiling on the false-belief tasks so their data was excluded from analysis. A 2 (ToM score: weak vs. strong) x 2 (age group: 4-year-olds vs. 5-year-olds) factorial analysis of variance was conducted on SI scores. This did not reveal a significant effect of ToM on SI score, $F(1,48) = 0.681$, $p>0.40$, an age interaction effect, $F(1,48) = 2.376$, $p>0.10$ or a main effect of age, $F(1,48) = 0.029$, $p>0.80$. N was roughly equal for weak versus strong ToM ability respectively across the 4-year-old age group, 15:10 and the 5-year-old age group, 12:15. Table 3.7 reports the means and standard deviations of SI scores for children with weak and strong ToM in the 4-year-old and 5-year-old age groups.

Table 3.7: *SI score means and standard deviations for children with weak ToM (ToM score: <2) and strong ToM (ToM score =2) across the 4-year-old and 5-year-old age groups.*

Age group	ToM group	Mean SI score	Standard deviation
4	Weak	1.33	1.84
	Strong	2.60	1.90
5	Weak	2.25	2.01
	Strong	1.87	1.89

Relationship between EF and ToM

A 2 (inhibitory EF score: weak vs. strong) x 3 (age group: 4-year-olds vs. 5-year-olds vs. 6-year-olds) factorial analysis of variance was conducted on ToM scores. This analysis did not reveal a significant effect of inhibitory EF ability on ToM score, $F(1,69) = 0.261$, $p>0.60$ or an age interaction, $F(2,69) = 0.709$, $p>0.40$. However, there was a significant main effect of age group, $F(2,69) = 6.965$, $p<0.01$. Table 3.8 reports the means and standard deviations of ToM scores for the weak and strong inhibitors across the age groups and Table 3.9 reports the ANOVA summary data. One-tailed t-tests revealed that the ToM scores of the 6-year-olds (M=1.91, SD=0.29) were significantly greater than the ToM scores of the 4-year-olds (M=1.24 SD=0.72): $t(31.952) = 4.296$, $p<0.001$ and the 5-year-olds (M=1.52, SD=0.58): $t(39.349) = 3.113$, $p<0.01$.

Table 3.8: *ToM score means and standard deviations for the weak inhibitors (inhibitory EF score: <12) and strong inhibitors (inhibitory EF score ≥12) across the 4-year-old, 5-year-old and. 6-year-old age groups.*

Age group	Inhibitory group	Mean ToM score	Standard deviation
4	Weak	1.13	0.74
	Strong	1.40	0.70
5	Weak	1.58	0.52
	Strong	1.47	0.64
6	Weak	1.88	0.35
	Strong	1.93	0.26

Table 3.9: *Summary table for the 2 (inhibitory EF score: <12/weak vs. ≥12/strong) x 3(age group: 4-year-olds vs. 5-year-olds vs. 6-year-olds) analysis of variance of ToM scores.*

Source	SS	df	MS	F	p	Partial Eta²
Within	22.592	69	0.327			
Inhibitory EF	0.085	1	0.085	0.261	0.611	0.004
Age group	4.561	2	2.281	6.965	**0.002**	0.168
Inhibitory EF x age group	0.464	2	0.232	0.709	0.496	0.020
(corrected model)	5.995	5	1.199	3.662	0.005	0.210
(corrected total)	28.587	74				

As with SI scores, a 2 (shifting EF score: weak vs. strong) x 3 (age group: 4-year-olds vs. 5-year-olds vs. 6-year-olds) factorial analysis of variance could not be conducted on ToM scores because there were not enough children with weak shifting EF in the 5- and 6-year-old age groups. However, since N was roughly comparable across the weak versus strong shifting groups for the 4-year-olds, a t-test was performed to look at the difference in ToM scores for weak versus strong shifters in the youngest age-group. This did not reveal a significant effect of shifting EF ability on ToM score, $t(23) = 0.537$, $p>0.50$. Table 3.10 reports the means and standard deviations of ToM scores for children with weak and strong shifting in the 4-year-old age group.

Table 3.10: *ToM score means and standard deviations for the weak shifters (shifting EF score: <6) and strong shifters (shifting EF score =6) across the 4-year-old age group.*

Age group	Shifting group	Mean ToM score	Standard deviation
4	Weak	1.13	0.64
	Strong	1.29	0.77

SIs

A 4 (SI task trial categories: 1= Set 1 control trials vs. 2= Set 2 control trials vs. 3= Set 3 control trials vs. 4= SI trials) x 3 (age group: 4-year-olds vs. 5-year-olds vs. 6-year-olds) factorial analysis of variance, with SI task scores as the dependent measure revealed a significant effect of task trial category on SI task score, $F(3,288) = 21.968$, $p<0.001$. There was no main or interaction effect of age group. Table 3.11 reports the means and standard deviations of SI task scores for Trial Sets 1-4 across the age groups and Table 3.12 reports the ANOVA summary data. One-tailed t-tests revealed that the scores in the SI trials were significantly lower than the scores in the Set 1 Control trials, $t(74) = 7.364$, $p<0.001$, Set 2 Control trials, $t(74) = 4.182$, $p<0.001$ and the Set 3 Control trials, $t(74) = 6.020$, $p<0.001$. Two-tailed t-tests revealed that only the 6-year-olds were performing above chance on the SI trials, $t(22) = 2.902$, $p<0.01$.

Across the age groups, 26 children scored poorly (i.e. achieved a score of 2 or below) in at least one control condition in the SI task (Trial Sets 1-3). Of these children, 22 performed poorly in the SI trials (Trial Set 4), so only 4 children performed well in the SI condition whilst performing poorly in the control conditions. Binomial analyses indicated that the proportion of children scoring poorly on a control condition who also scored poorly on the SI trials, differed significantly from chance, $p<0.001$. In fact of these 22 children, 1 scored poorly in Set 1, 18 in Set 2 and 5 in Set 3. Of the overall number of 2 children who performed poorly in Set 1, 1 child performed poorly and 1 child performed well (i.e. achieved a score of 3 or more) in the SI trials, a proportion not differing from chance expectations, $p>0.40$. Of the overall number of 6 children who performed poorly in Set 3, 5 children performed poorly and 1 child performed well in the SI trials. This proportion did not differ from chance expectations either, $p>0.09$. However, of the overall number of 20 children who performed poorly in Set 2, 18 children performed poorly and 2 children performed well in the SI trials, a proportion which did differ significantly from chance, p<0.001.

Table 3.11: *SI task score means and standard deviations for the SI task trial sets (Sets 1-3: Controls and Set 4: SI trials) across the 4-year-old, 5-year-old and. 6-year-old age groups.*

Age group	SI task trial set	Mean SI task score	Standard deviation
4	1	3.80	0.50
	2	2.72	1.79
	3	3.48	1.16
	4	1.84	1.93
5	1	3.93	0.27
	2	3.07	1.36
	3	3.67	0.96
	4	2.04	1.91
6	1	3.83	0.49
	2	3.17	1.44
	3	3.70	0.56
	4	2.96	1.58

Table 3.12: *Summary table for the 4 (task trial categories: 1=Control 1 vs. 2=Control 2 vs. 3=Control 3 vs. 4=SI trials) x 3(age group: 4-year-olds vs. 5-year-olds vs. 6-year-olds) analysis of variance of SI task scores.*

Source	SS	df	MS	F	p	Partial Eta²
Within	485.741	288	1.687			
Task trial category	111.155	3	37.052	21.968	**0.001**	0.186
Age group	9.835	2	4.917	2.916	0.056	0.020
Task trial category x age group	10.770	6	1.795	1.064	0.384	0.022
(corrected model)	135.895	11	12.354	7.325	0.001	0.219
(corrected total)	621.637	299				

The effect of SI task set presentation order upon performance in the SI trials, was also assessed. Table 3.13 reports the means and standard deviations of SI scores when SI trials were presented before and after Control trial sets 1-3. A t-test revealed that performance in the SI trials altered according to whether these trials were presented before or after Set 1 trials, $t(73) = 2.113$, $p < 0.05$. Higher scores were achieved if SI trials were presented after Set 1 than if they were presented before Set 1. Performance in the SI trials was also influenced by whether they were presented before or after Set 2 trials, $t(73) = 2.112$, $p<0.05$. Higher scores were achieved if SI trials were presented after Set 2 than if they were presented before Set 2. However performance in the SI trials did not appear to be affected by whether Set 3 was presented before or after the SI trials, $t(73) = 0.572$, $p>0.50$.

Table 3.13: *SI score means and standard deviations when the SI trials were presented before and after Control trial sets 1-3.*

SI trial presentation order	Control trial set	Mean SI score	Standard deviation
Before	1	1.68	1.93
	2	1.82	1.86
	3	2.13	1.85
After	1	2.60	1.75
	2	2.70	1.77
	3	2.38	1.89

Excluding poor controls

Some earlier analyses of SI performance were then repeated excluding children performing poorly in 1 or more of the control trials. Because of the low numbers of children in some age groups, age effects were no longer examined.

As in the main sample, children's inhibitory EF was classed as strong if their scores on the Day-Night task were 12 or greater out of 24 and weak if these were 11 or less on the basis that this enabled as equal a split as possible in the number of children assigned to the two EF ability groups. N was roughly equal for weak versus strong inhibitory ability, 24:25. Concordant with the earlier analysis, a 2 (inhibitory

EF score: weak vs. strong) analysis of variance, with SI scores as the dependent measure did not reveal a significant effect of inhibitory EF ability on SI score, $F(1,47) = 0.133$, $p>0.70$. Table 3.14 reports the means and standard deviations of SI scores for children with weak and strong inhibitory EF.

Table 3.14: *SI score means and standard deviations for the weak inhibitors (inhibitory EF score: <12) and strong inhibitors (inhibitory EF score ≥12) when children performing poorly in control trials were excluded.*

Inhibitory group	Mean SI score	Standard deviation
Weak	3.13	1.51
Strong	2.96	1.65

Also as in the main sample, children's shifting EF was classed as strong if their scores on the DCCS were 6 out of 6 and weak if these were less than 6 on the basis that this enabled as equal a split as possible in the number of children assigned to the two EF ability groups. However, a 2 (shifting EF score: weak vs. strong) analysis of variance could not be conducted on SI scores because there were not enough children with weak shifting EF: they appeared to be scoring near ceiling. N was 5:44 across the weak versus strong shifting groups.

Children's ToM was classed as strong if their scores on the false belief tasks were 2 out of 2 and weak if these were less than 2 with the weak and strong division criteria chosen to enable as equal a split as possible in the number of children assigned to the two ToM ability groups. N was roughly comparable for weak versus strong ToM, 14:35. In line with earlier findings, a 2 (ToM score: weak vs. strong) analysis of variance, with SI scores as the dependent measure, did not demonstrate a significant effect of ToM on SI scores, $F(1,47) = 0.265$, $p>0.60$. Table 3.15 reports the means and standard deviations of SI scores for children with weak and strong ToM.

Table 3.15: *SI score means and standard deviations for children with weak ToM (ToM score: <2) and children with strong ToM (ToM score = 2) when children performing poorly in control trials were excluded.*

ToM group	Mean SI score	Standard deviation
Weak	2.86	1.75
Strong	3.11	1.51

A 4 (SI task trial categories: 1= Set 1 control trials vs. 2= Set 2 control trials vs. 3= Set 3 control trials vs. 4= SI trials) analysis of variance, with SI task scores as the dependent measure again revealed a significant effect of task trial category on SI task score, $\underline{F}(3,192) = 12.408$, $\underline{p}<0.001$. Table 3.16 reports the means and standard deviations of SI task scores for Trial Sets 1-4 and Table 3.17 reports the ANOVA summary data. As before, one-tailed t-tests demonstrated that the scores in the SI trials were significantly lower than the scores in the Set 1 Control trials, $\underline{t}(48) = 3.897$, $\underline{p}<0.001$, Set 2 Control trials, $\underline{t}(48) = 3.648$, $\underline{p}<0.001$ and the Set 3 Control trials, $\underline{t}(48) = 3.634$, $\underline{p}<0.001$.

Table 3.16: *SI task score means and standard deviations for the SI task trial sets (Sets 1-3: Controls and Set 4: SI trials) when children performing poorly in control trials were excluded.*

SI task trial set	Mean SI task score	Standard deviation
1	3.92	0.28
2	3.84	0.37
3	3.88	0.33
4	3.04	1.57

Table 3.17: *Summary table for the 4 (task trial categories: 1=Control 1 vs. 2=Control 2 vs. 3=Control 3 vs. 4=SI trials) analysis of variance of SI task scores when poor controls were excluded.*

Source	SS	df	MS	F	p	Partial Eta²
Within	133.551	192	0.696			
Task trial category	25.893	3	8.631	12.408	**0.001**	0.162
(corrected model)	25.893	11	8.631	12.408	0.001	0.162
(corrected total)	159.444	195				

EF

Correlations partialling out age revealed that there was no relationship between performance on the inhibitory and shifting EF measures in the whole sample of 75 children, $r=0.133$, $p>0.10$. Performance on the inhibitory EF task did not appear to improve significantly with age, $F(2,72) = 3.047$, $p>0.05$. However, there were significant age differences in performance on the shifting EF task, $F(2,72) = 6.080$, $p<0.01$. Table 3.18 reports the means and standard deviations of inhibitory and shifting EF scores across the 4-year-old, 5-year-old and. 6-year-old age groups. Table 3.19 reports the ANOVA summary data for the age-shifting EF analysis. One-tailed t-tests revealed that the 5- and the 6-year-olds had significantly higher scores for the shifting EF task than the 4-year-olds, $t(27.223) = 3.002$, $p<0.01$ and $t(38.188) = 2.182$, $p<0.05$, respectively (see Table 3.3 for means and standard deviations).

Table 3.18: *Inhibitory and shifting EF task score means and standard deviations across the 4-year-old, 5-year-old and. 6-year-old age groups.*

EF component	Age group	Mean EF task score	Standard deviation
Inhibitory EF	4	10.20	3.55
	5	11.93	3.49
	6	12.70	3.82
Shifting EF	4	4.08	2.86
	5	5.85	0.77
	6	5.52	1.59

Table 3.19: *Summary table for the 3 (age group: 4-year-olds vs. 5-year-olds vs. 6-year-olds) analysis of variance of shifting EF task scores.*

Source	SS	df	MS	F	p	Partial Eta²
Within	266.987	72	3.708			
Age group	45.093	2	22.547	6.080	**0.004**	0.144
(corrected model)	45.093	2	22.547	6.080	0.004	0.144
(corrected total)	312.080	74				

ToM

Correlations partialling out age revealed that there was no significant relationship between performance on the two ToM tasks either, r=0.167, p>0.07. Performance on the first ToM task did not appear to improve significantly with age, $F(2,72) = 0.843$, p>0.40. However, age did appear to have a significantly improve performance on the second ToM task, $F(2,72) = 10.143$, p<0.001. Table 3.20 reports the means and standard deviations of scores on the first and second ToM tasks across the 4-year-old, 5-year-old and. 6-year-old age groups. Table 3.21 reports the ANOVA summary data for the analysis of scores on the second ToM task. One-tailed t-tests revealed that the 6-year-olds had significantly higher scores for the second ToM task than the 4-year-olds, $t(32.658) = 5.104$, p<0.001 and the 5-year-olds, $t(36.214) = 3.137$, p<0.01.

Only 8 of the 75 children receiving the false belief tasks failed the first false belief task presented: 4 of whom were 4-years-old, 3 of whom were 5-years-old and 1 of whom was 6-years-old, whereas 26 failed the second task presented: 15 of whom were 4-years-old, 10 of whom were 5-years-old and 1 of whom was 6-years-old. A t-test indicated that the difference in the scores for attained for the first and second false belief tasks presented, (M=0.89, SD=0.31) and (M=0.65, SD=0.48) respectively, was significant; $t(74) = 4.030$, p<0.001. Whereas 5 of the children who failed the second task presented, also failed the first task, 21 children passed the first false belief task but failed under further testing. Only 3 children showed the reverse pattern – failing on the first task presented but passing the second.

Table 3.20: *Score means and standard deviations for the ToM tasks presented first and second across the 4-year-old, 5-year-old and. 6-year-old age groups.*

ToM task presentation	Age group	Mean ToM score	Standard deviation
First	4	0.84	0.37
	5	0.89	0.32
	6	0.96	0.21
Second	4	0.40	0.50
	5	0.63	0.49
	6	0.96	0.21

Table 3.21: *Summary table for the 3 (age group: 4-year-olds vs. 5-year-olds vs. 6-year-olds) analysis of variance of scores on the second ToM task.*

Source	SS	df	MS	F	p	Partial Eta²
Within	13.253	72	0.184			
Age group	3.734	2	1.867	10.143	**0.001**	0.220
(corrected model)	3.734	2	1.867	10.143	0.001	0.220
(corrected total)	16.987	74				

3.4 Discussion

This study was concerned with examining the relationship between EF and appreciation of the "some implies not-all" SI. It also served to investigate the validity of the SI task employed by Papafragou and Musolino (2003) Experiment 2.

In contrast to expectations and the masked competence account of the emergence of CU, no relationship was revealed between EF assessed in terms of either inhibitory or shifting ability, and the demonstration of CU assed in terms of the computation of SIs. Although Nilsen and Graham (2009) reported a relationship between EF and CU manifest in an association between inhibitory EF and communicative perspective-taking ability, so far no evidence has been produced to support a relationship between EF and the ability to compute SIs.

Furthermore, support was not provided for expectations and standard accounts of CU which conceptualise CU in terms of the appreciation of the communicative intentions behind messages. Analyses did not indicate that ToM was significantly related to CU. Although previous studies have revealed relationships between ToM and other aspects of CU, such as the ability to detect violations of Gricean maxims (Surian, Baron-Cohen, & van der Lely, 1996) and the ability to appreciate irony (Martin, & McDonald, 2003), evidence has not been provided to support a relationship between ToM and the specific ability to compute SIs to date.

Whilst the lack of association between EF ability and performance on the SI and false belief measures might at first seem problematic for a masked competence account of the development of CU, its significance would seem to be undermined by weaknesses observed with the measures used to assess EF. For example, the older children: 5- and 6-year-olds, were performing at ceiling on the DCCS, precluding analysis of associations between their DCCS performance and responses on the SI task. Further, on the Day-Night task, some children responded very quickly with the required label when presented with a card, but the responses of others were often delayed. It could have been that reaction times during these trials reflected the ease with which children were inhibiting the standard labelling association, so that these times were indexing inhibitory ability. One might thus expect that if response times had been measured in the Day-Night task, they would have correlated positively with scores in the SI trials. Indeed, it might have been that reaction times would have proved similarly informative in the DCCS task. Diamond and Kirkham (2005) report that when reaction times are examined, even adults reveal a cognitive cost when switching in the DCCS task, despite performing at near ceiling with regards to sorting accuracy.

Moreover, in the Day-Night task, children were initially told the rules of the game, and then presented with trials requiring them to act in accordance with the instructions. However initial trials were only considered training trials (i.e. performance on them was did not contribute to the task score) until the point at which two consecutive trials were passed. At this point children were considered to have demonstrated appreciation of the task rules, so these two consecutive trials and all trials presented thereafter were counted as test trials. However there appeared to be quite a large variation in the number of training trials children received before passing two consecutive trials.

Moreover, it was noticed that some children who only required a few (e.g. three) training trials before proceeding to the test trials, often achieved moderate scores on the Day-Night task, whereas children who required a large number of training trials (e.g. 10) often achieved high scores. It is conceivable that because the training trials required demonstration of the inhibition rule, they actually served to hone inhibitory ability. Thus it might have been that children who received more training trials, were initially failing to pass two consecutive trials, not due to a failure to understand task rules, but due to a lack of inhibitory power. The additional training trials received could conceivably have provided more of a chance to practice and strengthen this skill before ability started to make a contribution to task scores. If this had been the case, the training trials would have been serving to diminish differences in inhibitory ability between children, reducing the likelihood that a relationship would be found between these differences and variation in the ability to compute SIs.

Therefore, the use of EF tasks which reduce the need for EF when initially establishing appreciation of task requirements, might increase sensitivity to individual differences in EF ability. For example, in the Simon task measure of inhibitory ability, participants can be asked to provide an initial demonstration of rule appreciation that does not require inhibition. The use of this task, and others which do not require EF during training trials, would enhance the likelihood that a relationship would be found between EF and the ability to compute SIs.

Although the sensitivity of the EF tasks used may have been limited by the inclusion of EF dependent training trials and their failure to incorporate a reaction time performance measure, the differentiation revealed between the two EF measures is supported in the adult literature (Miyake et al., 2000). Nonetheless, the absence of significant relationships between performance on either EF task and false belief understanding that have previously been reported in the literature (e.g. Carlson & Moses, 2001; Hala, Hug, & Henderson, 2003; Frye, Zelazo, & Palfai, 1995; Perner, Lang & Kloo, 2002) further suggests that the EF measures should be improved. Although it is worth noting that the lack of a significant association between performance on the EF tasks and scores in the false-belief ToM tasks, might also be partially attributable to the unsuitability of the ToM tasks employed for the age of the children tested in Experiment 1.

The literature indicates that the first order false-belief tasks employed in this study are typically passed by children as young as 3 to 4 years (Leslie, 1994; Surian & Leslie, 1995), whereas in this study they were given to children of up to 6 years. Indeed the 6-year-olds in this study were performing at ceiling on the false-belief tasks, precluding their ToM data from some analyses. Furthermore, even the score variation of the 4- and 5-year-olds who were not scoring at ceiling was questionable, with analyses indicating that they were biased to pass the first task presented, over the second. Such a pattern would not be expected if passing only one of the two tasks was a genuine sign of a lack of ToM. This leaves open the suggestion that variation in ToM scores, which was calculated from the summation of scores across the two tasks and mainly reflected the performance of children passing either both or only one of the two tasks, revealed fatigue effects rather than genuine ToM difficulties. That performance on the two false belief tasks did not correlate significantly supports this argument. Studies which have reported relationships between performance on the Day-Night task and first-order false belief understanding, have tended to focus on younger children between 3 and 4 years of age (Carlson & Moses, 2001; Hala, Hug, & Henderson, 2003). Although Perner, Lang and Kloo (2002) found a relationship between the DCCS and first-order false belief performance of children aged between 3 and 6 years of age, over two thirds of their participants were under 5-years-old and their oldest participant was 6 years 2 months.

The use of second-order false belief tasks, which are more appropriate for use with the older children tested in this study (standard tasks are typically passed by 6-year-olds; Baron-Cohen, Miller, 2009; O'Riordan, Stone, Jones & Plaisted, 1999; Perner & Wimmer, 1985; Sullivan, Zaitchik, & Tager-Flusberg, 1994), might have enabled the demonstration of ToM-EF and ToM-SI associations.

In accordance with hypotheses, performance on the SI trials of the CU task appeared to be significantly worse than performance in all of the control conditions. The fact that children performed better on the two control conditions requiring positive evaluations of the puppet's descriptions (Sets 1 and 3), than in the SI trials (Set 4), suggests that good performance on the SI trials was not simply reflecting the existence of a negative response bias. If children had been exhibiting a negative response bias, it would have negatively affected their performance on these two control conditions. Additional enhanced performance (again relative to that in the SI trials), in the control condition requiring negative evaluation of the puppet's descriptions (Set 2), indicates further that performance in the SI trials did not simply reflect the ability to evaluate negatively. If this had been the case, one would have expected similar performance across these two sets. The fact that significantly more children who scored poorly in Set 2 control trials, performed poorly in the SI trials than performed well in the SI trials, but the same relationship did not hold for children who scored poorly in Control Sets 1 or 3, does however suggest that the demands of responding negatively to the puppet's descriptions, did indeed evoke an additional processing cost.

Significantly better performance in all of the control conditions relative to the SI trials, suggests that performance in the latter condition reflected more than the general cost of processing associated with the linguistic content and context of the descriptions to be evaluated and responses to be produced. Although these general costs were also demanded in the control conditions, performance was significantly better in the control conditions than in the SI trials. Indeed, this differentiation between performance in the SI and control conditions, remained even when all children scoring poorly in the control trials (i.e. only correctly rejecting or accepting descriptions 50% or less of the time), were removed from analyses, supporting the robustness of the finding. The difference between performance in the SI trials and performance in the control conditions suggests that the SI trials were also tapping a separate ability not evoked in the control trials. Given the nature of the SI trials, this is very likely to be the ability to compute SIs. This data thus serves to support the validity of the SI trials used in Papafragou and Musolino (2003) Experiment 2. As discussed in Section 3.1, the control conditions used in this study were superior to those used by Papafragou and Musolino (2003) Experiment 2, because the linguistic content used in the SI trials was incorporated. This reduced the possibility that factors specific to the linguistic terms used, but not reflecting the actual computation of SIs, could account for the rejection of descriptions during the SI trials.

Further, the considerable demands on response flexibility, invoked by the employment of three control conditions (two requiring responses conflicting, and one requiring responses consistent with those necessary for the SI trials) extended the degree of response alternation evoked in the processing of descriptions relevant to the some-all implicatures of Papafragou and Musolino (2003) Experiment 2. Thus closer examination of the effect of response demands on performance in the SI trials was enabled. There were no significant relationships between poor performance in Control Sets 1 and 3, which required positive evaluation of the puppets' descriptions, and performance in the SI trials, which required negative evaluation of puppet descriptions. However, poor performance in Control Set 2, which required negative evaluation of the puppet's descriptions, was significantly related to poor performance in the SI trials, suggesting that performance in the SI trials might have been partially reflective of the processing demands evoked by responding negatively. The fact remains nonetheless that performance in the SI trials was significantly different from performance in Control Set 2, indicating that performance in the SI trials was not just reflecting response demands. This supports the suggestion that the SI trials were indeed, to some extent, assessing the ability to compute SIs.

The idea that negative responding may make additional demands on processing resources is consistent with Fritzley and Lee (2003)'s report that, initially (around 2 years of age), children demonstrate a bias to respond positively to questions. Fritzley and Lee argue that this response tendency disappears later on (around 4-5 years of age), provided the questions presented are comprehensible. Incomprehensible queries are thought to evoke a negative response bias from 4- to 5-year-olds. If negative responses are more resource-demanding, it makes sense that younger children with under-developed processing systems should demonstrate a preference for a less taxing positive response.

The suggestion that performance in the SI trials might have been partially reflective of the processing demands evoked by responding negatively, is lent further support by the finding that scores for the SI trials appeared to be significantly better when the trials were presented after rather than before the Control Set 2 trials. This is concordant with the idea that the practice of responding negatively in the Control Set 2 trials was honing the ability to evaluate negatively, making it easier to then provide a similar response in the SI trials presented afterwards. However, performance in the SI trials also appeared to be significantly better when they were presented after rather than before the Control Set 1 trials, which demanded positive description evaluations. This suggests that presentation order effects were not just attributable to response training effects. The fact that performance in the SI trials did not seem to be affected by whether these trials were presented before or after Control Set 3, which was similar to Control Set 1 in providing appropriate scenario descriptions, indicates that the facilitatory effect of Control Set 1 was not due to more general task training effects either. However, by presenting children with "all" descriptions in appropriate scenarios, Control Set 1 might have served to remind children of the concept of "all" promoting their grasp of the comparative aspect of the concept of "some".

The level of SIs drawn in this study, was compared to the proportion drawn in Papafragou and Musolino (2003) Experiment 2. Their sample of 5- to 6-year-olds

(M= 67 months), correctly rejected 52.5 % descriptions in the SI trials, whilst the 5- to 6-year-olds tested in this study (mean age 70.44 months) correctly rejected 61.5 % of descriptions in these trials. However, when the older 6-year-olds in the current study were excluded, to reduce the mean age of our sample to a level comparable to that of Papafragou and Musolino's sample (new mean age of our sample being 67.16 Months), the sample of 5- to 6-year-olds from this study were found to have rejected 56.45 % of descriptions in the SI trials. This level is more in line with the findings of Papfragou and Musolino's study.

With the exception of performance on the inhibitory EF task, all the abilities measured in this study improved significantly with age. This is in line with expectations. Concerns were raised previously in this section concerning the sensitivity of the inhibitory EF task which might also account for the lack of age effects found for this task.

3.5 Conclusions

Experiment 1 did not provide support for the masked competence account of CU as neither inhibitory nor shifting EF abilities were found to demonstrate relations to SI appreciation. Furthermore, it did not reveal a significant association between ToM and SI appreciation, in contrast to the conception of SIs in terms of the recognition of communicative intentions. However, comparison of performance on the SI control tasks incorporated in this study, provided support for the validity of the SI task employed by Papafragou and Musolino (2003). Experiment 2 was undertaken to investigate whether the use of different EF measures would allow demonstration of a relationship between EF and CU and whether the incorporation of more advanced ToM tasks would reveal that ToM plays a mediatory role in this relationship.

PAGE MISSING IN ORIGINAL

Chapter 4: Experiment 2 – Using new measures to investigate the relationships between executive functioning, scalar implicatures and theory of mind.

4.1 Introduction

Experiment 1 investigated whether EF abilities are related to measures of CU in young children aged between 4 and 6 years, and whether ToM ability plays a mediating role in this relationship. It also examined the validity of the SI task employed by Papafragou and Musolino (2003). Children performed significantly better in the control trials than in the SI trials of the SI task, supporting the proposal that the SI trials were genuinely reflective of the ability to compute SIs, rather than the more general processing costs demanded in the control conditions. However, the masked competence account of the emergence of CU was not supported, as EF abilities were not found to be related to computation of SIs. Furthermore, ToM ability was not found to be significantly associated with either appreciation of SIs or EF. Possible reasons for these lack of associations were discussed, including the suggestion that EF tasks which placed training trial demands on the EF abilities assessed in proceeding test trials, worked to diminish initial individual differences in EF. The absence of significant relationships must also be seen in terms of performance of the children on the ToM and EF tasks that was often at a ceiling level, particularly for the 5- and 6-year-olds.

The second experiment was conducted using alternative measures of EF and additional measures of ToM to further investigate the relationship of EF to CU and ToM. CU was assessed using the SI trials employed in Experiment 1. The control conditions of the conversational task were not used in Experiment 2 because the validity of the SI trials had already been established in the previous experiment. Furthermore, since only the 6-year-olds were found to demonstrate above chance performance levels in the SI trials Experiment 1, indicating that SI appreciation emerges between 5 and 6 years, the age range of the participants in Experiment 2 was restricted to these ages.

In place of the Day-Night inhibitory task used in Experiment 1, Experiment 2 employed the Frog-Simon task to assess inhibitory ability. Unlike the Day-Night task, the Frog-Simon task did not require a pre-test trial training phase and so avoided the potential confound of inhibitory training affecting the day-night task in Experiment 1. In place of the DCCS shifting task, the Dibbets Updating task was used. Updating ability reflects another of Miyake et al's proposed components of EF and is an EF competence that was not investigated in Experiment 1. This task was expected to be more age appropriate than the DCCS task, on which the older children had performed at ceiling in Experiment 1. Although the updating task did employ an updating training phase, this served as one of the two measures of updating ability.

To assess ToM, Experiment 2 employed the first-order ToM tasks used in Experiment 1. These tasks are known as first-order ToM tasks because they assess the relatively basic ability to conceptualise that people merely represent (i.e. do not have direct access to) the physical world. However, as noted in Section 3.4, such tasks as usually passed by children aged 3-4 years. The relevance of variation in the performance of older children on such tasks is thus questionable. To strengthen conclusions drawn from analyses of performance on ToM measures, second-order ToM tasks were also employed in Experiment 2. Second-order ToM tasks are more difficult than first-order tasks, because they require the recognition that others appreciate that people merely possess representations of the physical world. Second-order tasks are typically passed by 6-year-olds (Baron-Cohen, O'Riordan, Stone, Jones, & Plaisted 1999; Miller, 2009; Perner and Wimmer 1985; Sullivan, Zaitchik, &

Tager-Flusberg, 1994) and so variation in performance on this task is likely to reflect genuine differences in second order ToM appreciation rather than fatigue effects. However, first-order tasks were still included in this study, to gauge individual differences in the ToM of 5-year-olds who might be expected to perform at baseline on the second-order ToM tasks.

In Experiment 2, it was again hypothesised that both EF and appreciation of false beliefs would be positively related to awareness of SIs. A relationship between appreciation of false beliefs and EF abilities was also expected. It was further anticipated that the relationship between awareness of SIs and EF abilities would be mediated by the appreciation of false beliefs. Additionally, all abilities were expected to improve with age.

4.2 Method

Participants. These were 62 children aged between 5 and 6 years, split into two age groups: a younger group (M=68.76 months, range =62-76 months) and an older group (M=79.79 months, range =77-83 months). The children were predominantly white and middle class, and were recruited from a primary school in Sheffield. All children spoke English as their first language, and none were known to have any specific language impairment. Two other children (both aged 77 months) were excluded for failing to participate in all the tasks presented (one was unable to perform the response action required in the inhibitory task due to a prosthetic arm, and the other declined to finish the updating task).

Procedure. The children were tested individually, in a quiet area of their school, and all took part in seven tasks presented in a fixed order. Testing was conducted over the course of one session (lasting approximately 20-25 minutes). The session commenced with the SI task followed by two first-order false-belief tasks, two second-order false-belief tasks, the Frog Simon Inhibitory task, and an Updating task.

SI Task. This was the SI task employed in Experiment 1 that was administered without the control conditions used in Experiment 1, as the previous experiment had already established that performance in the SI trials was not simply reflective of a response bias or general processing demands. Children were introduced to two hand puppets - Ellie the Elephant and Harry the Hedgehog, and took part in warm-up statement training trials led by Ellie the Elephant, as in Experiment 1. Following the warm-up statements, children were presented with the four scalar implicature test scenarios comprising Set 4 in Experiment 1. The implicature descriptions were delivered by Harry the Hedgehog. Presentation of these trials was counterbalanced using two different orders. For half of the children the trials were presented as they read downwards in Set 4 of Table 3.2, Experiment 1. For the other half of the children, the trials were presented as they read upwards in Set 4 of Table 3.2, Experiment 1. The total number of correct responses to enquiries regarding whether the puppet had described the situation well, was recorded and translated into the SI score with a possible range of 0-4. As in Experiment 1, children were awarded 1 point for each "No" response they made, but were not awarded points for "Yes" responses. Children who correctly evaluated the puppet's statements negatively were asked how the puppet could have "said it better". Correct justifications (indicating the puppet's incorrect use of the term "all" and the better suitability of the term "some") were recorded and translated into the SI justification score with a possible range of 0-4. Although a conservative measure of the appreciation of SIs, appropriate justifications were used to support the validity of the SI score.

First-order false belief tasks. These were identical to the tasks used in Experiment 1. Each task score had a possible range of 0-1.

Second-order false belief tasks. Two tasks were used to assess second-order theory of mind ability. One was based on the memory aid version of the second-order task of Perner and Wimmer (1985), as used by Baron-Cohen (1989). Children were presented with four toy characters: a grandmother, a grandfather, a small boy (Johnny) and a baby in a pram along with a small wooden house containing a bed, and a wooden chair (see Appendix Va). The house was placed in the centre of an A3 board, which had two rectangular strips of light green paper stuck to the surface, in front of and behind, where the house was positioned (representing the front and back gardens). A strip of dark green paper was attached to the surface of the board to the left of the front of the house (representing the park), and a strip of blue paper attached to the right of the front of the house (representing the sea). The children were introduced to the four toy characters and then were told the story below:

One day Granny said, "I'm going to take the baby for a walk in the park. Do you want to come with us Johnny?" It was a hot day so Johnny said, "I'm too hot, I don't want to go for a walk!" So Granny went off to the park with baby, while Johnny went to play in the back garden, and Granddad sat at the front of the house. A little later Granddad saw Granny coming back from the park. "Where are you going?" he asked. Granny replied "The park was shut, so I'm going to take baby to the sea instead." Granddad said "Ok, I'm going to have a little sleep." Next, Granny and baby walked by the back of the garden. "Hello Granny, I'm up here!" waved Johnny from the tree. Granny told Johnny that she and baby were going to the seaside.

Children were then asked two control questions to check that they were following the story:

- *Does Granddad know that Granny talked to Johnny?*

- *Where are Granny and baby?*

If these questions were passed, the story continued as below. If they were failed, the relevant part of story was repeated and children were asked these questions again. No child failed the control questions a second time.

A little later, Johnny was bored, and decided to go and find Granny and baby. He ran back through the house and called out "Granddad, I'm going off to play with Granny and baby."

Children were then asked two test questions:

- *Where does Granddad think Johnny will go?*

- *Why does Granddad think Johnny will go there?*

Following this, two more control questions were presented:

- *Where are Granny and baby?* (reality control)

- *Where did Granny go with baby first of all?* (memory control)

The other second-order task was based on that of Sullivan, Zaitchik and Tager-Flusberg (1994). Children were presented with five pictures (see Appendix Vb), using PowerPoint slides. These were shown in the sequence 1-5 on a laptop computer. The pictures were accompanied by the narration below:

Picture 1 - *Granddad has given Mary and Simon some chocolate to share. "Go and put it away now children", says Granddad, "you can have some when mum says so".*

Picture 2 – *The children run into the kitchen and put the chocolate in the fridge. Then they go out to play.*

Picture 3 – *A little later, Simon comes in for a glass of water. He goes to the fridge and he sees the chocolate. He wants to keep the chocolate all for himself, so he takes the chocolate out of the fridge and puts it in his bag.*

At this point children were asked two control questions to ensure that they had been following the story:

– *Where does Mary think the chocolate is?*

- *Where has Simon put the chocolate really?*

The test was discontinued if either control question was failed. No child failed these control questions.

Picture 4 – *Oh look! Mary is playing by the window; she can see everything that Simon is doing! She sees him put the chocolate in his bag! Simon is so busy hiding the chocolate he doesn't see Mary watching him! Later Mum calls Simon and Mary in for tea. She says they can have some of the chocolate. So Simon and Mary come running into the kitchen.*

Children were then presented with Picture 5 and asked two test questions:

- *Where does Simon think Mary will look for the chocolate?*

- *Why does Simon think that?*

Children were asked two more control questions:

– *Where is the chocolate really? (reality control)*

- *Where was the chocolate first of all?* (memory control)

The tasks were presented in a fixed order (the Perner task followed by the Sullivan task). As with the first-order false belief tasks, incorrect responses to the control questions presented after the test questions in a second-order task meant that responses to the test questions in that task were automatically recorded as incorrect. The total number of correct responses to the two test questions in each task was recorded separately as the score for that task, with each task score having a possible range of 0-2.

Frog Simon Inhibitory task.. This task was adapted from Simon and Small (1969). In the original task for adults, participants were asked to respond to high and low pitch tones using their left and right hands. Half of Simon and Small's sample were asked to press a key on their right when they heard a high-pitched tone but to press a key on their left when they heard a low-pitch tone. The instructions were reversed for the other half of the sample. Headphones were then used to deliver high and low pitch tones to the left and right ears of participants. Simon and Small found that responses were significantly faster when the ear to which a tone was played corresponded to the hand required to respond to that tone. The irrelevant presentation-location of the target item appeared to interfere with processing the response required, suggesting that responding in accordance with presentation-location forms a prepotent response to be inhibited. The Simon task would thus seem to reflect a measure of inhibitory ability.

The task was adapted for use with children by replacing high and low pitch tones with red and blue frogs, which were presented singly, on either the left- or right-hand-side of a laptop computer screen using a series of PowerPoint slides. Half of the slides presented a frog on the left-hand-side of the screen, and of these slides, half presented a blue frog and half presented a red frog. The other half of the slides presented a frog on the right-hand-side of the screen and of these slides, half presented a blue frog and half presented a red frog. Children were asked to put a red bracelet on their right wrist and a blue bracelet on their left wrist and were instructed to tap the table with their red (right) hand when they saw a red frog, and to tap the table with their blue (left) hand when they saw a blue frog (see Appendix VI). The colour of the frog was the only relevant piece of information from each slide. However, it was expected that the irrelevant left- or right-hand-side screen location of the frogs would bias children to respond with their respective hands, interfering with their ability to respond in accordance with the frog colour when the screen location and colour response conflicted. Response accuracy was used to assess inhibitory ability in this task and the total number of correct responses to the 20 stimulus slides presented was recorded as the score, with a possible range of 2-20._This was because children had to respond correctly to the first two task slides, in which the frog location did not conflict with the colour response-hand instruction in order to show that they had listened to the task instructions. No child performed incorrectly on the first two slides.

Updating task. This was based on the "Switch task for children" developed by Dibbets and Jolles (2006). Although the task was originally designed as a measure of the ability to switch between mental representations, it actually requires children to detect when these changes are required: in accordance with cueing stimuli. The task thus also involves the ability to update mental representations.

Training phase –Sun: Children viewed a series of PowerPoint slides on a laptop accompanied by the experimenter's narration. They first viewed a slide showing a big cat asleep on a sofa, and a little kitten wide awake at her side. They were told:

"Mummy cat has fallen asleep on the sofa, but her little kitten is feeling a bit bored and so he wanders off and goes outside."

At this point they were shown a slide of the kitten against a yellow background with a big sun in the middle of the sky.

" But it's very hot outside so he starts to feel tired and takes a nap under a bush."

They were then shown a slide with a purple bush on one side of the screen and a red bush on the other, again against a yellow background with a big sun in the middle of the sky. A lawnmower separated the two bushes. Children were asked:

"Which bush do you think kitten's fallen asleep under? Just take a guess".

The experimenter used the mouse to click on the chosen bush. If a child chose the purple bush, it would disappear to reveal the kitten and the experimenter would say *"There he is, under the purple bush".* If the red bush was chosen, it would disappear to reveal a big red cross inside a black square. The experimenter would say *"He isn't under the red bush. So which bush do you think he's under?"* The experimenter then told the child that she needed to check that they knew which bush the kitten was under and the process was repeated until the bush with the kitten had been identified on seven consecutive trials. The experimenter clicked on the lawnmower to move onto the next trial. In the sun slides, the kitten was always revealed when the purple bush was chosen. The positioning of the coloured bushes was counterbalanced across slides between the left and right side of the screen. When the criterion had been reached, the experimenter clicked on the sun, which ended that phase.

Training phase –Rain: After the Sun training phase, the Rain training phase was presented. The experimenter opened up a new slideshow in which the purple and red bush were shown against a grey background. There was now a rain cloud in the sky where the sun had been in the previous slideshow. The lawnmower remained positioned between the two bushes. The experimenter said, *"Well now it's raining! Now which bush do you think kitten is under?"* The mouse was used to click on the chosen bush, but this time the kitten was revealed under the red bush. If the child chose the purple bush, it disappeared to reveal the red cross. The experimenter again told the child that she needed to check that they knew which bush the kitten was under and the process was repeated until the bush with the kitten had been identified on seven consecutive trials.

In the rain slides, the kitten was always revealed when the red bush was chosen. The positioning of the coloured bushes was counterbalanced across slides between the left and right side of the screen. When the criterion had been reached, the experimenter clicked on the rain cloud, which ended that phase.

Training phase –Updating: After the Rain training phase, the Updating training phase was presented. The experimenter opened up a new slideshow in which the child was first shown a sun cue slide: sun in the centre of a yellow slide, followed by four slides from the Sun training phase. A sun was presented in the middle of the sky in these slides and the bushes were shown against a yellow background. Children were asked to identify which bush the kitten was under in each slide. Following the sun slides, children were presented with a rain cue slide: rain cloud in the centre of a grey slide, followed by four slides from the Rain training phase. A rain cloud was presented in the middle of the sky in these slides and the bushes were shown against a grey background. Again the children were asked to identify which bush kitten was under. Children were then presented with an updating cue slide: rain cloud and sun in the centre of a grey-yellow chequered slide, followed by four slides, two of which were from the Rain training phase and two of which were from the Sun training phase. These trials were presented in the order sun, rain, sun, rain. Following the first updating block, "pure" blocks of sun or rain slides (four slides in each block) were presented interspersed with updating blocks, in the order: rain, updating, sun, updating, sun, updating, rain, updating, sun, updating, rain, updating, rain, updating, sun, updating.

After the first updating block (Block 3) had been presented, passing criteria came into effect. These were identical to those applied in the previous two training phases. Children were required to identify the bush revealing the kitten on seven consecutive trials. When criterion had been reached, the experimenter ended that phase in the appropriate manner. The level of training reached during this phase was initially split into three categories for the purpose of identifying children suitable to move onto the switch test trials. Although all children demonstrated the ability to respond appropriately to location cues in the sun and rain training phases, some children completely lost this ability upon viewing the updating blocks. These children became unable even to respond appropriately during the pure blocks presented following the updating blocks in the updating training phase, and so could not pass even four consecutive trials. Such children were assigned to a "Poor Level of Updating Training" category. The "Intermediate Level of Updating Training" category comprised children who had a stable basic appreciation that the kitten could be found under the purple bush in the sun slides and under the red bush in the rain slides, and so could respond appropriately during the pure blocks presented in the updating training phase. These children could thus pass four consecutive trials. However, they were not good enough at attending to the sun/rain cues to demonstrate competence in the updating blocks and thus could not pass seven trials consecutively. Children who did well in both the pure and updating blocks and successfully identified the location of the kitten on seven consecutive trials were assigned to a "Good Level of Updating Training" category.

Updating Test Trials: Children who were assigned to the Good Level of Updating Training or Intermediate Level of Updating Training categories in the training phase, moved onto the Updating test trials. For those children assigned to the

Poor Level of Updating Training category, the updating task ceased with the last training trial. The test trials comprised eight blocks of ten trials. Blocks were presented in the order – sun, updating, rain, updating, sun, updating, rain, updating. Each block was preceded by an appropriate cue slide. The total number of correct responses to the 80 test slides presented was recorded as the "Updating Test" score, with a possible range of 0-80. Correct responses required updating of slide representations, to denote the presence of the sun or of the rain cloud indicators of kitten location, and so the number of correct responses given was considered to reflect updating ability.

However, the switch training phase also required children to constantly update their mental representations of the slides shown on screen. Thus the level of training reached (Updating Training level), could also be considered to reflect updating ability. Although children were initially split into three different groups, to identify children suitable to move onto the test trials of the Dibbets task, analyses of updating ability reflected in the updating training phase recognised only one level distinction. All children assigned to the Poor or Intermediate Level of Updating Training categories were considered together and put in a "weak updating training" category, which was assigned a score of 0. All children assigned to the Good Level of Updating Training category were put in a "strong updating training" category, which was assigned a score of 1. Table 4.1 provides a summary of the slides shown and corresponding kitten locations for the different trials presented during the training and test phases of the Dibbets updating measure.

Table 4.1: *Training and test trials employed in the Dibbets Task updating measure in Experiment 2*

Task Phase	Phase Trials	Trial Slides	Kitten Location
Training	Sun	All with sun cue	Always the purple bush
	Rain	All with rain cue	Always the red bush
	Updating	Blocks of sun trials comprising 4 sun cue slides, and rain trials comprising 4 rain cue slides, separated by blocks of updating trials comprising 2 sun and 2 rain cue slides presented in an interspersed order i.e. sun, rain, sun, rain.	Always the purple bush in sun trial blocks, always the red bush in rain trial blocks, half the time the red bush and half the time the purple bush in updating trial blocks.
Test	Updating	Blocks of sun trials comprising 10 sun cue slides, and rain trials comprising 10 rain cue slides, separated by blocks of updating trials comprising 5 sun and 5 rain cue slides presented in an interspersed order i.e. sun, rain, sun, rain, rain, sun, rain, sun, sun, rain.	Always the purple bush in sun trial blocks, always the red bush in rain trial blocks, half the time the red bush and half the time the purple bush in updating trial blocks.

4.3 Results

The means and standard deviations of the task scores achieved by the 62 participants are presented in Table 4.2. Since only 49 children satisfied the criterion for passing on to the test phase in the Dibbets updating EF task (23 children from the younger age group, 26 from the older age group), the means and standard deviations of the Dibbets updating EF task test scores were derived from this sub-sample of 49 children.

Table 4.2: *Mean scores and standard deviations for EF, ToM and CU tasks employed in Experiment 2*

Age group	Younger	Older
N	33	29
Age (months)	68.76 (4.66)	79.79 (1.93)
SI scores/4	2.24 (1.75)	2.76 (1.79)
SI justifications/4	1.55 (1.82)	1.90 (1.86)
First first-order FB/1	0.94 (0.24)	0.97 (0.19)
Second first-order FB/1	0.82 (0.39)	0.90 (0.31)
Perner second-order FB/2	0.52 (0.83)	0.48 (0.79)
Sullivan second-order FB/2	1.33 (0.82)	1.48 (0.74)
Frog Simon inhibition EF/20	18.52 (1.50)	18.66 (2.18)
Dibbets updating EF test scores/80	69.87 (9.54)	75.08(5.69)
Dibbets updating EF training scores EF/1	0.45 (0.51)	0.59 (0.50)

Note: SI=scalar implicature task; FB=False belief task; EF=Executive functioning task.

Relation between EF and SIs

Children's updating EF training was classed as strong if their scores for the training phase of the Dibbets task were 1 and weak if these were 0. The weak and strong division criteria were selected on the basis that they enabled as equal a split as possible in the number of children assigned to the two EF ability groups. A 2 (updating EF training score: weak vs. strong) x 2 (age group: younger vs. older) factorial analysis of variance was conducted on SI scores. N was comparable for weak versus strong updating groups across the age groups, 18:15, and 12:17 for the younger and older age groups respectively. The effect of updating EF training score on SI score was significant, $F(1,58) = 4.967$, $p<0.05$. Children with strong updating training had greater SI scores (M=3.00, SD=1.61) than children with weak updating training (M=1.93, SD=1.80). There was no significant main or interaction effect involving age group. Table 4.3 reports the means and standard deviations of SI scores for the weak and strong updating training groups across age groups and Table 4.4 reports the ANOVA summary data.

Table 4.3: *SI score means and standard deviations for the weak updaters (updating EF training score: <1) and strong updaters (updating EF training score =1) across the older and younger age groups.*

Age group	Updating group	Mean SI score	Standard deviation
Younger	Weak	1.50	1.62
	Strong	3.13	1.51
Older	Weak	2.58	1.93
	Strong	2.88	1.73

Table 4.4: *Summary table for the 2 (updating EF training score: 0/weak vs. 1/strong) x 2(age group: younger vs. older) analysis of variance of SI scores.*

Source	SS	df	MS	F	p	Partial Eta²
Within	164.915	58	2.843			
Updating EF	14.124	1	14.124	4.967	**0.030**	0.079
Age group	2.621	1	2.621	0.922	0.341	0.016
Updating EF x Age group	6.734	1	6.734	2.368	0.129	0.039
(corrected model)	26.569	3	8.856	3.115	0.033	0.139
(corrected total)	191.484	61				

A 2 (updating EF training score: 0/weak vs. 1/strong) x 2 (age group: younger vs. older) factorial analysis of variance was also conducted on SI justifications. The effect of updating EF training score on SI justifications was also significant, $F(1,58) = 8.838$, $p<0.01$. Children with strong updating training produced more SI justifications (M=2.38, SD=1.77) than children with weak updating training (M=1.00, SD=1.64). Again, there was no significant main or interaction effect for age group. Table 4.5 reports the means and standard deviations of SI justifications for the weak and strong updating training groups across age groups and Table 4.6 reports the ANOVA summary data.

Table 4.5: *SI justification means and standard deviations for the weak updaters (updating EF training score: <1) and strong updaters (updating EF training score =1) across the older and younger age groups.*

Age group	Updating group	Mean SI justifications	Standard deviation
Younger	Weak	0.67	1.41
	Strong	2.60	1.72
Older	Weak	1.50	1.88
	Strong	2.18	1.85

Table 4.6: *Summary table for the 2 (updating EF training score: 0/weak vs. 1/strong) x 2(age group: younger vs. older) analysis of variance of SI justifications.*

Source	SS	df	MS	F	p	Partial Eta²
Within	169.071	58	2.915			
Updating EF	25.763	1	25.763	8.838	**0.004**	0.132
Age group	0.635	1	0.635	0.218	0.642	0.004
Updating EF x Age group	5.975	1	5.975	2.050	0.158	0.034
(corrected model)	35.704	3	11.901	4.083	0.011	0.174
(corrected total)	204.774	61				

Children's updating EF test scores were classed as strong if their scores for the test phase of the Dibbets task were 76 or greater and weak if these were less than 76. Again, the weak and strong division criteria were selected on the basis that they enabled as equal a split as possible in the number of children assigned to the two EF ability groups. A 2 (updating EF test score: weak vs. strong) x 2 (age group: younger vs. older) factorial analysis of variance was conducted on SI scores, and was necessarily restricted to the sub-sample of children taking part in the updating EF test trials. For the weak versus strong updating groups respectively, N was 17:6 for the younger age group and 9:17 for the older age group. This analysis did not reveal a significant effect of updating EF test score, $F(1,45) = 1.317$, $p>0.20$, and there was no significant main or interaction effect involving age group. Table 4.7 reports the means and standard deviations of SI scores for the weak and strong updating test score groups across age groups.

Table 4.7: *SI score means and standard deviations for the weak updaters (updating EF test score: <76) and strong updaters (updating EF test score ≥76) across the older and younger age groups.*

Age group	Updating group	Mean SI score	Standard deviation
Younger	Weak	2.29	1.99
	Strong	1.67	1.51
Older	Weak	3.22	1.39
	Strong	2.53	1.94

Similarly, a 2 (updating EF test score: <76/weak vs. ≥76/strong) x 2 (age group: younger vs. older) factorial analysis of variance with SI justifications as the dependent measure, again restricted to the sub-sample of children taking part in the updating EF test trials, did not reveal a significant effect of updating EF test score, $\underline{F}(1,45) = 0.036$, $\underline{p}>0.80$. As before, there was no significant main or interaction effect for age group. Table 4.8 reports the means and standard deviations of SI justifications for the weak and strong updating test score groups across age groups.

Table 4.8: *SI justification means and standard deviations for the weak updaters (updating EF test score: <76) and strong updaters (updating EF test score ≥76) across the older and younger age groups.*

Age group	Updating group	Mean SI justification	Standard deviation
Younger	Weak	1.82	1.94
	Strong	1.83	1.72
Older	Weak	1.67	1.80
	Strong	1.88	1.90

Children's inhibitory EF was classed as strong if their scores for the Frog Simon task were 19 or greater and weak if these were less than 19, with division criteria chosen on the basis that they enabled as equal a split as possible in the number of children assigned to the two EF ability groups. A 2 (inhibitory EF score: weak vs. strong) x 2 (age group: younger vs. older) factorial analysis of variance with SI scores as the dependent measure was carried out. For the weak versus strong inhibitory groups respectively, N was 12:21 for the younger age group and 8:21 for the older age group. There were no significant main effects of either inhibitory EF, $F(1,58) = 0.094$, $p>0.70$, or age group, $F(1,58) = 3.644$, $p>0.06$. However, there was a significant EF ability-age group interaction effect, $F(1,58) = 4.881$, $p<0.05$ (see Figure 4.1). Table 4.9 reports the means and standard deviations of SI scores for the weak and strong inhibitory groups across age groups and Table 4.10 reports the ANOVA summary data. T-tests revealed that weak inhibitory ability was associated with significantly greater SI scores in the older age group than strong inhibitory ability, $t(26.971) = 2.379$, $p<0.05$. There was no significant difference in the SI scores associated with weak and strong inhibitory ability in the younger age group, $t(31) = 1.453$, $p>0.10$.

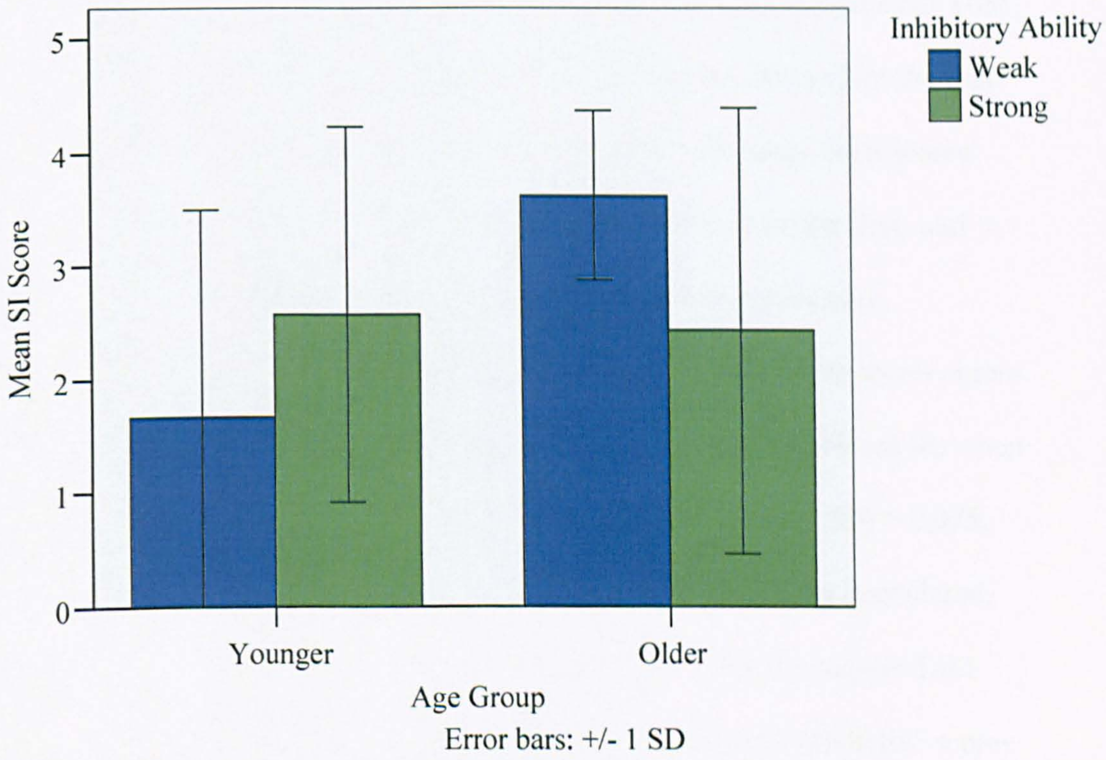
Table 4.9: *SI score means and standard deviations for the weak inhibitors (inhibitory EF score: <19) and strong inhibitors (inhibitory EF score ≥19) across the older and younger age groups.*

Age group	Inhibitory group	Mean SI score	Standard deviation
Younger	Weak	1.67	1.83
	Strong	2.57	1.66
Older	Weak	3.63	0.74
	Strong	2.43	1.96

Table 4.10: *Summary table for the 2 (inhibitory EF score: <19/weak vs. ≥19/strong) x 2(age group: younger vs. older) analysis of variance of SI scores.*

Source	SS	df	MS	F	p	Partial Eta²
Within	172.827	58	2.980			
Inhibitory EF	0.280	1	0.280	0.094	0.760	0.002
Age Group	10.857	1	10.857	3.644	0.061	0.059
Inhibitory EF x Age Group	14.544	1	14.544	4.881	**0.031**	0.078
(corrected model)	18.656	3	6.219	2.087	0.112	0.097
(corrected total)	191.484	61				

Figure 4.1: Mean SI Scores for Strong and Weak Inhibitors across Age Groups



A 2 (inhibitory EF score: <19/weak vs. ≥19/strong) x 2 (age group: younger vs. older) factorial analysis of variance with SI justifications as the dependent measure did not reveal a significant effect of inhibitory EF, $F(1,58) = 0.327$, $p>0.50$, or a significant main or interaction effect involving age group. Table 4.11 reports the means and standard deviations of SI justifications for the weak and strong inhibitory groups across age groups.

Table 4.11: *SI justification means and standard deviations for the weak inhibitors (inhibitory EF score: <19) and strong inhibitors (inhibitory EF score ≥19) across the older and younger age groups.*

Age group	Inhibitory group	Mean SI justifications	Standard deviation
Younger	Weak	1.08	1.78
	Strong	1.81	1.83
Older	Weak	2.00	1.93
	Strong	1.86	1.88

Role of ToM in the relationship between EF and CU

Regressions were conducted on SI scores, using composite measures of ToM as regressors. These composite measures comprised total scores across just the first-order ToM tasks, total scores across just the second-order ToM tasks, total scores across the first- and second-order ToM tasks and total scores across the first- and second-order ToM tasks not including points for appropriate second-order justifications (this last measure will be referred to as the total predictive scores across the first- and second-order ToM tasks). No ToM model proved a significant fit: when total scores across just the first-order ToM tasks were considered, $F(1,60) = 0.075$, $p>0.70$, when total scores across just the second-order ToM tasks were considered, $F(1,60)= 0.815$, $p>0.30$, when total scores across the first- and second-order ToM tasks were considered, $F(1,60)= 1.233$, $p>0.20$ and when only total predictive scores across the first- and second-order ToM tasks were considered, $F(1,60)= 2.195$, $p>0.10$.

Regressions using composite ToM measures were then repeated on SI justifications. As before, no ToM model proved a significant fit: for total scores across just the first-order ToM tasks, $F(1,60) = 0.006$, $p > 0.90$, for total scores across just the second-order ToM tasks, $F(1,60) = 0.368$, $p > 0.50$, for total scores across the first- and second-order ToM tasks, $F(1,60) = 0.713$, $p > 0.40$, for only total predictive scores across the first- and second-order ToM tasks, $F(1,60) = 1.635$, $p > 0.20$.

Regressions were conducted on SI scores next, using EF measures as regressors. Performance on the inhibitory EF task did not prove a significant fit: $F(1,60)=0.327$, $p>0.50$, nor did the background variable of age: $F(1,60)=1.278$, $p>0.20$. The best fitting model for the whole sample of SI score data thus employed only the updating EF training score. This model was a poor fit ($R^2_{adj} = 7.7\%$), but the overall relationship was significant $F(1,60) = 6.080$, $p< 0.05$. Table 4.12 reports the summary data for this regression.

When the sample of SI scores was restricted to those provided by children who had taken part in the updating EF test trials, models employing composite ToM measures again did not provide a significant fit: for total scores on the first-order ToM tasks, $F(1,47) = 0.002$, $p>0.90$; for total scores on the second-order ToM tasks, $F(1,47)= 2.248$, $p>0.10$, for total scores across the first- and second-order ToM tasks, $F(1,47)= 1.605$, $p>0.20$; and for only total predictive scores across the first- and second-order ToM tasks, $F(1,47)= 1.740$, $p>0.10$. Performance on the inhibitory EF task did not prove a significant fit either, $F(1,47)=0.268$, $p>0.60$, and neither did updating EF test scores, $F(1,47) = 0.428$, $p>0.50$, nor the background variable of age, $F(1,47)=1.034$, $p>0.30$. Thus, the best fitting EF model again consisted of a single measure, updating training EF score $F(1,47) = 9.092$, $p<0.01$. However, the amount of variance accounted for by this factor doubled compared to the variance accounted for when analyses had included the whole sample of SI data ($R^2_{adj} = 14.4\%$ for the restricted sample compared to 7.7% for the whole sample). Table 4.13 reports the summary data for this regression.

Table 4.12: *Summary table of regression analysis for variables predicting SI scores using data for the whole sample of 62 children*

	B	SE B	β	t	p
Step 1					
Constant	1.933	0.311		6.221	0.000
Updating EF training scores	1.067	0.433	0.303	2.466	**0.017**

Note: R^2 = 0.092 for step 1, p <.0.05

Table 4.13: *Summary table of regression analysis for variables predicting SI scores using data for the sub-sample of 49 children taking part in updating test trials.*

	B	SE B	β	t	p
Step 1					
Constant	1.471	0.410		3.588	0.001
Updating EF training scores	1.529	0.507	0.403	3.015	**0.004**

Note: R^2 = 0.162 for step 1, p <.0.01

Regressions were also conducted in the main sample on SI justifications using EF measures as regressors. Again, performance on the inhibitory EF task did not prove a significant fit, $F(1,60)=2.559$, $p>0.10$, and neither did the background variable of age, $F(1,60)=0.631$, $p>0.40$. Thus, as before, the best fitting EF model for the whole sample of SI justification data employed only the updating EF training score. This model did not account for an impressive amount of variance (R^2_{adj} = 12.9%), but the overall relationship was significant, $F(1,60) = 10.008$, $p<0.01$. Table 4.14 reports the summary data for this regression. When the sample of SI justifications was restricted to those provided by children who had taken part in the updating test trials models employing composite ToM measures again did not provide a significant fit: when total scores across just the first-order ToM tasks were considered $F(1,47) = 0.177$, $p>0.60$, when total scores across just the second-order ToM tasks were considered $F(1,47)= 1.131$, $p>0.20$, when total scores across the first- and second-order ToM tasks were considered $F(1,47)= 1.147$, $p>0.20$ and when only total predictive scores across the first- and second-order ToM tasks were considered $F(1,47)=2.134$, $p>0.10$. This time however, although updating EF test scores did not prove a significant fit, $F(1,47)=0.134$, $p>0.70$, and age did not account for a significant amount of variance, $F(1,47)=0.001$, $p>0.90$, performance on the inhibitory EF task did prove a significant fit, $F(1,47)=4.404$, $p<0.05$. Table 4.15 reports the summary data for this regression. Nevertheless, the best fitting EF model continued to consist of a single measure, updating EF training score $F(1,47) = 10.363$, $p<0.01$. Table 4.16 reports the summary data for a regression including both inhibitory and updating EF variables and Table 4.17 reports the summary data for the single updating EF measure regression. As in the SI score analyses, the amount of variance in SI justifications accounted for by updating EF training score, increased slightly in the restricted sample compared to the variance accounted for when analyses had included the whole sample (R^2_{adj} = 16.3% in the restricted sample compared to 12.9% in the whole sample).

Table 4.14: *Summary table of regression analysis for variables predicting SI justifications using data for the whole sample of 62 children*

	B	SE B	β	t	p
Step 1					
Constant	1.000	0.312		3.203	0.002
Updating EF training scores	1.375	0.435	0.378	3.164	**0.002**

Note: R^2 = 0.143 for step 1, p <0.01.

Table 4.15: *Summary table of regression analysis for inhibitory EF variable predicting SI justifications using data for the sub-sample of 49 children taking part in updating test trials*

	B	SE B	β	t	p
Step 1					
Constant	-3.106	2.359		-1.317	0.194
Inhibitory EF scores	0.267	0.127	0.293	2.099	**0.041**

Note: R^2 = 0.086 for step 1, $p < 0.05$.

Table 4.16: *Summary table of regression analysis for EF variables predicting SI justifications using data for the sub-sample of 49 children taking part in updating test trials*

	B	SE B	β	t	p
Step 1					
Constant	-2.400	2.222		-1.080	0.286
Inhibitory EF scores	0.178	0.123	0.196	1.448	0.155
Updating EF training scores	1.418	0.512	0.374	2.770	**0.008**

Note: R^2 = 0.216 for step 1, $p < 0.01$

Table 4.17: *Summary table of regression analysis for updating EF variable predicting SI justifications using data for the sub-sample of 49 children taking part in updating test trials*

	B	SE B	β	t	p
Step 1					
Constant	0.765	0.404		1.892	0.065
Updating EF training scores	1.610	0.500	0.425	3.219	**0.002**

Note: R^2 = 0.181 for step 1, $p < 0.01$

Relation between EF and ToM performance

A 2 (inhibitory EF score: <19/weak vs. ≥19/strong) x 2 (age group: younger vs. older) factorial analysis of variance was conducted upon composite predictive scores across the first- and second-order ToM tasks as the dependent measure. As indicated previously, the weak and strong division criteria were selected on the basis that they enabled as equal a split as possible in the number of children assigned to the two EF ability groups. For the weak versus strong inhibitory groups respectively, N was 12:21 for the younger age group and 8:21 for the older age group. There was a significant main effect of inhibitory EF, $\underline{F}(1,58) = 4.441$, $\underline{p}<0.05$. Children with strong inhibitory ability produced greater predictive composite ToM scores (M=3.05, SD=0.73) than children with weak inhibitory ability (M=2.60, SD=0.82). There was no main or interaction effect of age group. Table 4.18 reports the means and standard deviations of composite predictive first- and second-order ToM scores for the weak and strong inhibitors across the age groups and Table 4.19 reports the ANOVA summary data. Nevertheless, given the uneven distribution of children across inhibitory groups, 20:42, the results of this analysis and the analyses immediately below, should be treated cautiously.

Table 4.18: Composite predictive *ToM score means and standard deviations across the first- and second-order ToM tasks for the weak inhibitors (inhibitory EF score: <19) and strong inhibitors (inhibitory EF score ≥19) across the older and younger age groups.*

Age group	Inhibitory group	Mean ToM score	Standard deviation
Younger	Weak	2.13	1.96
	Strong	1.40	1.90
Older	Weak	1.83	1.85
	Strong	2.20	2.01

Table 4.19: *Summary table for the 2 (inhibitory EF score: <19/weak vs. ≥19/strong) x 2(age group: younger vs. older) analysis of variance of composite predictive scores across the first- and second-order ToM tasks*

Source	SS	df	MS	F	p	Partial Eta²
Within	33.839	58	0.583			
Inhibitory EF	2.591	1	2.591	4.441	**0.039**	0.071
Age Group	0.353	1	0.353	0.605	0.440	0.010
Inhibitory EF x Age Group	0.196	1	0.196	0.336	0.564	0.006
(corrected model)	3.580	3	1.193	2.045	0.117	0.096
(corrected total)	37.419	61				

However, a 2 (inhibitory EF score: <19/weak vs. ≥19/strong) x 2 (age group: younger vs. older) factorial analysis of variance with composite first-order ToM scores as the dependent measure did not reveal a significant effect of inhibitory EF, $F(1,58) = 0.397$, $p>0.50$ or a main or interaction effect of age group. Table 4.20 reports the means and standard deviations of composite first-order ToM scores for the weak and strong inhibitors across the age groups. Similarly, a 2 (inhibitory EF score: <19/weak vs. ≥19/strong) x 2 (age group: younger vs. older) factorial analysis of variance with composite second-order ToM scores as the dependent measure did not reveal a significant effect of inhibitory EF, $F(1,58) = 1.828$, $p>0.10$ or a main or interaction effect of age group. Table 4.21 reports the means and standard deviations of composite second-order ToM scores for the weak and strong inhibitors across the age groups. Further, a 2 (inhibitory EF score: <19/weak vs. ≥19/strong) x 2 (age group: younger vs. older) factorial analysis of variance with composite first- and second order ToM scores as the dependent measure did not reveal a significant effect of inhibitory EF, $F(1,58) = 2.676$, $p>0.10$ or a main or interaction effect of age group. Table 4.22 reports the means and standard deviations of composite first- and second-order ToM scores for the weak and strong inhibitors across the age groups.

Table 4.20: Composite *ToM score means and standard deviations across the first-order ToM tasks for the weak inhibitors (inhibitory EF score: <19) and strong inhibitors (inhibitory EF score ≥19) across the older and younger age groups.*

Age group	Inhibitory group	Mean ToM score	Standard deviation
Younger	Weak	1.75	0.45
	Strong	1.76	0.54
Older	Weak	1.75	0.71
	Strong	1.90	0.30

Table 4.21: Composite *ToM score means and standard deviations across the second-order ToM tasks for the weak inhibitors (inhibitory EF score: <19) and strong inhibitors (inhibitory EF score ≥19) across the older and younger age groups.*

Age group	Inhibitory group	Mean ToM score	Standard deviation
Younger	Weak	1.67	1.37
	Strong	1.95	1.24
Older	Weak	1.50	0.76
	Strong	2.05	0.92

Table 4.22: Composite *ToM score means and standard deviations across the first- and second-order ToM tasks for the weak inhibitors (inhibitory EF score: <19) and strong inhibitors (inhibitory EF score ≥19) across the older and younger age groups.*

Age group	Inhibitory group	Mean ToM score	Standard deviation
Younger	Weak	3.42	1.56
	Strong	3.71	1.38
Older	Weak	3.25	0.89
	Strong	4.05	0.87

A 2 (updating EF training score: 0/weak vs. 1/strong) x 2 (age group: younger vs. older) factorial analysis of variance was conducted upon composite first-order ToM scores, with division criteria were selected on the basis that they enabled as equal a split as possible in the number of children assigned to the two EF ability groups. For the weak versus strong updating groups respectively, N was 18:15 for the younger age group and 12:17 for the older age group. Table 4.23 reports the means and standard deviations of composite first-order ToM scores for the weak and strong updaters across age groups. The ANOVA did not reveal a significant effect of updating EF, $F(1,58) = 1.440$, $p>0.20$ or a main or interaction effect of age group. Neither did a 2 (updating EF training score: 0/weak vs. 1/strong) x 2 (age group: younger vs. older) factorial analysis of variance with composite second-order ToM scores as the dependent measure. There was no significant effect of updating EF,

$F(1,58) = 0.362$, $p>0.50$ or main or interaction effect of age group. Table 4.24 reports the means and standard deviations of composite second-order ToM scores for the weak and strong updaters across age groups. Similarly, a 2 (updating EF training score: 0/weak vs. 1/strong) x 2 (age group: younger vs. older) factorial analysis of variance with composite first- and second order ToM scores as the dependent measure did not reveal a significant effect of updating EF, $F(1,58) = 0.545$, $p>0.40$ or a main or interaction effect of age group. Table 4.25 reports the means and standard deviations of composite first- and second-order ToM scores for the weak and strong updaters across age groups. Further, a 2 (updating EF training score: 0/weak vs. 1/strong) x 2 (age group: younger vs. older) factorial analysis of variance with composite predictive scores across the first- and second-order ToM tasks as the dependent measure did not reveal a significant effect of updating EF, $F(1,58) = 0.727$, $p>0.30$ or a main or interaction effect of age group. Table 4.26 reports the means and standard deviations of composite predictive ToM scores across first- and second-order ToM tasks for the weak and strong updaters across age groups.

Table 4.23: Composite *ToM score means and standard deviations across the first-order ToM tasks for the weak updaters (updating EF training score: <1) and strong updaters (updating EF training score =1) across the older and younger age groups.*

Age group	Updating group	Mean ToM score	Standard deviation
Younger	Weak	1.78	0.43
	Strong	1.73	0.59
Older	Weak	1.67	0.65
	Strong	2.00	0.00

Table 4.24: Composite *ToM score means and standard deviations across the second-order ToM tasks for the weak updaters (updating EF training score: <1) and strong updaters (updating EF training score =1) across the older and younger age groups.*

Age group	Updating group	Mean ToM score	Standard deviation
Younger	Weak	1.61	1.34
	Strong	2.13	1.19
Older	Weak	2.00	1.13
	Strong	1.82	0.73

Table 4.25: Composite *ToM score means and standard deviations across the first- and second-order ToM tasks for the weak updaters (updating EF training score: <1) and strong updaters (updating EF training score =1) across the older and younger age groups.*

Age group	Updating group	Mean ToM score	Standard deviation
Younger	Weak	3.39	1.34
	Strong	3.87	1.55
Older	Weak	3.83	1.19
	Strong	3.82	0.73

Table 4.26: Composite predictive *ToM score means and standard deviations across the first- and second-order ToM tasks for the weak updaters (updating EF training score: <1) and strong updaters (updating EF training score =1) across the older and younger age groups.*

Age group	Updating group	Mean ToM score	Standard deviation
Younger	Weak	2.72	0.75
	Strong	2.87	0.99
Older	Weak	2.92	1.00
	Strong	3.12	0.33

A 2 (updating EF test score: <76/weak vs. ≥76/strong) x 2 (age group: younger vs. older) factorial analysis of variance was conducted upon composite first-order ToM scores. Again, the weak and strong division criteria were selected on the basis that they enabled as equal a split as possible in the number of children assigned to the two EF ability groups. For the weak versus strong updating groups respectively, N was 17:6 for the younger age group and 9:17 for the older age group. The analysis did not reveal a significant effect of updating EF, $F(1,45) = 0.158$, $p>0.60$ or a main or interaction effect of age group. Table 4.27 reports the means and standard deviations of composite first-order ToM scores for the weak and strong updaters across age groups. Likewise, a 2 (updating EF test score: <76/weak vs. ≥76/strong) x 2 (age group: younger vs. older) factorial analysis of variance with composite second-order ToM scores as the dependent measure did not reveal a significant effect of updating EF, $F(1,45) = 0.786$, $p>0.30$ or a main or interaction effect of age group. Table 4.28 reports the means and standard deviations of composite second-order ToM scores for the weak and strong updaters across age groups.

A 2 (updating EF test score: <76/weak vs. ≥76/strong) x 2 (age group: younger vs. older) factorial analysis of variance with composite first- and second-order ToM scores as the dependent measure did not reveal a significant effect of updating EF, $F(1,45) = 0.838$, $p>0.30$ or a main or interaction effect involving age group. Table 4.29 reports the means and standard deviations of composite ToM scores across first- and second-order ToM tasks for the weak and strong updaters across age groups. Similarly a 2 (updating EF test score: <76/weak vs. ≥76/strong) x 2 (age group: younger vs. older) factorial analysis of variance with composite predictive scores across the first- and second-order ToM tasks as the dependent measure did not reveal a significant effect of updating EF, $F(1,45) = 1.023$, $p>0.30$ or a main or interaction effect of age group. Table 4.30 reports the means and standard deviations of composite predictive ToM scores across first- and second-order ToM tasks for the weak and strong updaters across age groups.

Table 4.27: Composite *ToM score means and standard deviations across the first-order ToM tasks for the weak updaters (updating EF test score: <76) and strong updaters (updating EF test score ≥76) across the older and younger age groups.*

Age group	Updating group	Mean ToM score	Standard deviation
Younger	Weak	1.71	0.59
	Strong	2.00	0.00
Older	Weak	2.00	0.00
	Strong	1.82	0.53

Table 4.28: Composite *ToM score means and standard deviations across the second-order ToM tasks for the weak updaters (updating EF test score: <76) and strong updaters (updating EF test score ≥76) across the older and younger age groups.*

Age group	Updating group	Mean ToM score	Standard deviation
Younger	Weak	1.65	1.32
	Strong	2.17	1.47
Older	Weak	1.89	1.17
	Strong	2.00	0.71

Table 4.29: Composite *ToM score means and standard deviations across the first- and second-order ToM tasks for the weak updaters (updating EF test score: <76) and strong updaters (updating EF test score ≥76) across the older and younger age groups.*

Age group	Updating group	Mean ToM score	Standard deviation
Younger	Weak	3.35	1.62
	Strong	4.17	1.47
Older	Weak	3.89	1.17
	Strong	3.82	0.88

Table 4.30: Composite predictive *ToM score means and standard deviations across the first- and second-order ToM tasks for the weak updaters (updating EF test score: <76) and strong updaters (updating EF test score ≥76) across the older and younger age groups.*

Age group	Updating group	Mean ToM score	Standard deviation
Younger	Weak	2.59	1.00
	Strong	3.17	0.75
Older	Weak	3.11	0.60
	Strong	3.06	0.75

SI performance

One-tailed t-tests were performed to compare SI scores and justifications with chance performance. The SI scores of the older age group were found to be significantly greater than a chance level of 50%, $t(28) = 2.287$, $p<0.05$ (see Table 4.2 for group mean and standard deviation). However, the SI justifications of the older age group were found to be at chance levels, $t(28) = 0.300$, $p>0.07$. The SI scores and justifications of the younger age group were both found to be at chance levels, $t(32) = 0.796$, $p>0.40$ and $t(32) = 1.433$, $p>0.10$ respectively.

EF performance

A 2 (age group: younger vs. older) analysis of variance with updating EF test scores as the dependent measure revealed that the positive effect of age was significant, $F(1,47) = 5.530$, $p<0.05$ (see Table 4.2 for means and standard deviations across the age groups). Table 4.31 reports the summary data for this analysis. An analysis of variance could not be conducted on updating training score as this dependent measure had only 2 levels. However a correlational analysis revealed that updating training score was not significantly associated with age in months, $r=0.057$, $p>0.30$.

Table 4.31: *Summary table for the 2(age group: younger vs. older) analysis of variance of updating EF test scores.*

Source	SS	df	MS	F	p	Partial Eta²
Between Groups	330.933	1	330.933	5.530	**0.023**	0.105
Within Groups	2812.455	47	59.839			
Total	261643.000	49				

Updating EF test scores were not found to correlate significantly with inhibition EF scores when the effects of age were partialled out, $r=0.072$, $p>0.30$.

Similarly Updating EF training scores were not found to correlate significantly with inhibition EF scores when the effects of age were partialled out, $r=0.146$, $p>0.10$.

ToM performance

Correlations partialling out age revealed that scores for the two first-order ToM tasks were significantly correlated, $r = 0.331$, $p<0.01$. However, scores for the two second-order tasks were not significantly correlated, $r =-0.065$, $p>0.30$. A t-test revealed a significant difference between mean scores for the two first-order tasks $t(61)=2.185$, $p<0.05$. Scores for the task presented first were greater (M=0.95, SD=0.22) than scores for the task presented second (M=0.85, SD=0.36). A significant difference was also found between the mean scores for the second-order tasks $t(61)=6.156$, $p<0.001$. Scores for the task presented second were greater (M=1.40, SD=0.78) than scores for the task presented first (M=0.50, SD=0.81).

4.4 Discussion

This study was concerned with examining the relationship between EF, appreciation of the "some-all" SI and ToM, using different measures of EF to those employed in Experiment 1.

In line with the hypotheses and in support of the masked competence account of the emergence of CU, a relationship was found between EF, assessed in terms of updating ability, and the demonstration of CU assed in terms of the computation of SIs. The fact that a relationship was revealed not only between EF and SI scores, but also between EF and SI justifications (a more conservative measure of SI appreciation), strengthens the validity of this finding. The relationship found between updating EF and CU was based upon updating training rather than updating test scores but this was likely due to the comparatively reduced proportion of variance in the test scores. Although inhibition did not demonstrate a main effect on SI performance, it was found to interact with age. Significantly better SI scores were associated with weak inhibitory performance in the older age group, than were associated with strong inhibitory performance in this age group, whilst inhibitory ability did not significantly affect SI scores in the younger age group. However, there were over twice as many strong inhibitors as weak inhibitors overall and the proportion of weak to strong inhibitors was even smaller in the older age group, with only just over one third of children in this age group being classified as weak inhibitors. Although the difference in SI scores for the weak and strong inhibitors in the younger age group was not significant, there was a trend for higher SI scores to be associated with strong inhibitors in this age group. Since the ratio of weak to strong inhibitors was less skewed in the younger age group than in the older age group, this relationship is more likely to reflect the true nature of the relationship between inhibitory ability and SI appreciation. One might thus expect that a more even proportion of weak to strong inhibitors would demonstrate a significant positive relation between inhibitory ability and SI appreciation.

Although recent research has indicated that EF is related to another aspect of CU relating to the ability to take account of common ground information (Nilsen & Graham, 2009), the relationship found between updating EF and SIs is the first time that a relationship has been demonstrated between EF and the specific ability to compute SIs. However, in contrast to expectations, regressions conducted to investigate the role of ToM in the relationship between EF and CU, indicated that ToM did not account for a significant amount of variance in SI scores. Only updating EF ability was found to predict SI scores. The same relationship dynamic, i.e., updating training but not ToM, updating test scores or inhibitory performance, successfully predicting SI appreciation, existed when analyses considered both SI score and also SI justifications. Since SI justifications are a more conservative measure of SI appreciation, the replication of this relationship with this measure strengthens the validity of this finding. Thus the masked competence account has been partially supported. EF has been found to play a role in the demonstration of CU as assessed through SI appreciation, but it is not clear that the effect of EF on CU is due to its impact on ToM.

The lack of support for the ToM-CU relationship contrasts with standard accounts of CU which conceptualise CU in terms of the appreciation of communicative intentions. However, it concords with the findings from Experiment 1. Although previous studies have demonstrated relationships between ToM and other aspects of CU such as the ability to detect violations of Gricean maxims and the ability to appreciate irony, no evidence has so far been provided evidence for a relationship between ToM and the ability to compute SIs.

Nevertheless, the absence of a significant link between ToM and CU, might be attributable to problems with the ToM measures used. The validity of the ToM scores produced in Experiment 1, was questioned in Section 3.4. The comparatively high number of children failing on the second but not the first ToM task was interpreted as a sign that some of the variation in overall ToM scores might have been reflecting fatigue rather than individual differences in ToM. This was especially likely given that the first-order ToM tasks used in Experiment 1 are typically passed by children who are of the same age as the participants taking part in that study. To gain a more accurate assessment of ToM appreciation, Experiment 2 thus also incorporated more difficult second-order ToM tasks. These are typically passed by 6- to 7-year-olds.

Since the likelihood of the 5- to 6-year-olds passing these tasks was expected to be smaller compared performance on first-order measures, failure on these tasks was expected to be more reflective of ToM deficiency than fatigue. The first-order ToM tasks used in Experiment 1 were also presented in Experiment 2. These were employed to distinguish children failing the second-order tasks who had a degree of ToM appreciation but hadn't quite achieved second-order competency, from those failing due to a more comprehensive ToM deficit.

In line with the findings from Experiment 1, more children in Experiment 2 failed the first-order task presented second than failed the first-order task presented first, 9 and 3 respectively. Indeed, analyses indicated that scores for the first-order task presented first were significantly greater than scores for the task presented second. Although 2 children failed both tasks, indicating a genuine ToM deficit, 7 children passed the first task but went on to fail the second, whilst only child demonstrated the opposite pattern. This result indicates that some of the variation in the composite first-order ToM task scores may have been attributable to fatigue rather than genuine lack of ToM. However, the strong relation found between scores from the two first-order tasks supports the suggestion that the two tasks were both tapping the same ability, and thus fatigue had a limited effect.

In contrast to the sequential performance patterns on the first-order ToM tasks, more children actually failed the second-order ToM task presented first than failed the task presented second, 43 compared to 11. In accord with this, analyses indicated that scores for the second-order ToM task presented second were significantly greater than scores for the task presented first. This suggests that fatigue was unlikely to account for much of the composite second-order score variation. The performance pattern on the second-order tasks makes sense because the second task presented was specifically designed by Sullivan, Zaitchik, and Tager-Flusberg (1994) to be a simplified version of the first task given (developed by Perner and Wimmer, 1985). Indeed Sullivan et al. found that the age at which children usually passed their simplified task was considerably younger than the age at which children were found to pass the Perner and Wimmer task (5.5 and 7 years respectively).

Unlike performance in the two first-order ToM tasks, performance on the two second-order tasks did not appear to be related. This suggests that a considerable degree of score variation for at least one of the second-order tasks was not reflective of individual differences in ToM. Indeed performance on the more difficult Perner and Wimmer (1985) task might have actually been more reflective of processing demands than ToM appreciation. However, analyses indicate that whilst EF performance was related to performance on the Sullivan et al. (1994) task, it was not related to performance on the Perner and Wimmer task, which is the reverse pattern to that which would be expected if the Perner and Wimmer task was more reflective of processing demands. Whatever the reason for the variation discrepancy in the two second-order tasks, the lack of relation between them does suggest that a factor other than ToM appreciation might have accounted for some of the composite second-order task score variation.

Analysis of performance on both the first- and second-order ToM tasks supports the suggestion that a considerable proportion of variance in ToM scores might have been reflecting a factor other than ToM ability. Thus, the absence of a relationship between ToM and CU, could have been attributable to problems with the specific ToM measures used. If more suitable ToM measures had been employed it is expected that a relationship would have been uncovered between ToM and CU and that this would have accounted for the relationship between EF and CU.

However, as hypothesised, a positive relationship was found between EF assessed via the new Frog Simon inhibitory EF task, and ToM appreciation, although the uneven spread of children across the weak and strong inhibitory groups entails that this relationship is drawn cautiously. The uncovering of this relationship is in accordance with previous reports of an association between inhibitory EF ability and ToM in the literature (Carlson & Moses, 2001; Hala, Hug, & Henderson, 2003). The demonstration of a significant relationship between inhibitory EF ability assessed via the new Frog Simon inhibitory EF task and ToM also supports the suggestion that inhibition-requiring training trials present in the inhibitory EF task employed in Experiment 1 but not required for the Frog Simon Task, served to mask individual differences in inhibitory ability displayed in test trials. Finding a relationship between EF and ToM indicates that at least some of the variance in ToM scores was valid, reflecting differences in ToM appreciation. Nevertheless, updating EF ability was not significantly associated with ToM.

The discovery that the amount of variance in SI score accounted for by updating EF ability, doubled when analysis focused on a restricted sample identified by their superior updating EF performance, is noteworthy. This finding suggests that the better children are at updating, the greater the role that updating EF ability is assigned in computing SIs. Not only does this demonstrate the significance of updating EF ability in the expression of CU, but also highlights that other factors are playing a role, and are more responsible for the level of CU demonstrated when updating EF ability is less advanced.

In line with the findings from Experiment 1, in which performance on the two EF tasks employed (measuring inhibitory and shifting ability) appeared to be differentiated once the effects of age had been partialled out, performance on the two

EF tasks employed in Experiment 2 (reflecting inhibitory and updating ability) was found to be unrelated. Although Miyake et al (2000) found that inhibitory, shifting and updating ability all correlated moderately with each other, their statistical analyses also indicated that these abilities were clearly distinguishable. Friedman et al. (2006) replicated these findings. Moreover, Miyake et al. found that these abilities actually contributed differentially to complex EF tasks, further promoting the idea of EF as a compound construct. The differentiation between the two EF abilities measured in this study supports the conceptual differentiation of these skills and so lends strength to the possibility that the two EF factors might be affecting CU in different ways.

The level of implicatures drawn in this study, as indicated by the SI score, was compared to the proportion drawn in Papafragou and Musolino (2003) Experiment 2. The 5- and 6-year-olds tested in this study (M= 73.92 months) correctly rejected 62.1 % of descriptions in the SI trials. This can be contrasted with Papafragou and Musolino's sample of 5- and 6-year-olds (M= 67 months), who correctly rejected 52.5 % descriptions in these trials. However, as in Experiment 1, when the oldest 6-year-olds in the current study were excluded, to reduce the mean age of our sample to a level comparable to that of Papafragou and Musolino's sample (new M for the present sample = 67.04 months), the sample of 5- and 6-year-olds from this study were found to have rejected 57.69 % of descriptions in the SI trials. This level is more in line with the findings of Papfragou and Musolino's study. Furthermore, consistent with the findings from Experiment 1, only the older 6-year-old age-group was performing at above chance-levels on the SI task in Experiment 2.

In contrast to both expectations, and the findings from the previous study in which all the abilities measured improved significantly with age, only updating ability appeared to get significantly better with age in Experiment 2. However, Experiment 2 focused on a smaller age range (5-6 years) than that investigated in Experiment 1 (4-6 years), so there was less variance in age available to exert an effect.

4.5 Conclusions

Experiment 2 provided partial support for the masked competence account of CU. It provided evidence of relations between updating EF and SI appreciation, and of an intermediary relation between inhibitory EF and ToM. However, it did not reveal an association between ToM and SI appreciation and ToM was not found to play a mediatory role in the relationship between EF and CU. The ability to compute SIs is only one of many skills involved in CU. Experiment 3 was therefore undertaken to investigate whether EF would be found to contribute to a different measure of CU: The Conversational Violation Task (CVT) and whether ToM would be found to play a mediatory role in this relationship.

Chapter 5: Experiment 3 – Investigating the relationships between executive functioning, conversational understanding and theory of mind using a broad measure of conversational understanding: The Conversational Violations Task

5.1 Introduction

Experiment 2 investigated whether inhibitory and updating components of EF are related to a specific aspect of CU in 5- and 6-year-olds: the appreciation of SIs. Children with strong updating EF performed significantly better in the SI task than children with weak EF, supporting the proposal that EF enhances CU. These results concur with the masked competence account of the emergence of CU. However, the ability to appreciate SIs is only one of many skills that comprise CU.

The third experiment employed a more comprehensive measure of CU to see if EF is related to CU more broadly. CU was assessed using a Conversational Violations Test (CVT) employed previously by Surian, Baron-Cohen and van der Lely (1996) and Siegal, Iozzi and Surian (2009). The CVT presents children with questions, each of which is paired with two possible answers:

Question: "Have you seen my dog?"

Answer 1: "Yes, he's in the garden."

Answer 2: "Yes, he's in the clouds."

Children have to choose which answer is the most appropriate response to the question. In order to determine the most appropriate response, they need to be aware of conversational maxims relating to the informativeness (Maxims of First and Second Quantity), truthfulness (Maxim of Quality), relevance (Maxim of Relation) and courteousness (Maxim of Politeness) of communication. One of the two answers to each question violates a maxim, and the task on the CVT as presented here was to select the alternative response each time. In the above example, Answer 2 violates the maxim relating to the validity of communication – it could not be true that the dog is in the clouds, so children should select Answer 1 as the most appropriate response.

The CVT has the advantage that maxim violations are counterbalanced across puppets over the course of the trials and the consequences of the violating puppet's responses are not made explicit. It is therefore not susceptible to the criticism directed at the study of Eskritt, Whalen, and Lee (2008) mentioned in Section 1.5.1: that task success might simply reflect learning over the course of trials. In Eskritt et. al.'s study, such learning would not require children to draw conclusions about intended meanings such as deception, purposeful vagueness or deliberate irrelevance. Instead, they could come to believe that asking for help from a particular puppet would not aid them in locating the sticker.

Because the CVT presents children with violations of a number of different maxims, this task requires broad knowledge of Grice's conversational maxims. Thus the CVT is a more comprehensive measure of CU than the SI task that only requires appreciation of the First Maxim of Quantity. This maxim dictates that people should try to be as informative as possible in communication, leading to the use of the stronger term "all" in conversation when it is applicable, rather than the logically compatible, but weaker term "some". Knowledge of the First Maxim of Quantity should lead children in the SI task to reject a puppet's statement that a character had performed "some" of a task when in fact the character had completed all of their task. However the SI task does not require appreciation of Grice's other conversation maxims. The puppet's statement should not be rejected on the grounds that it provided more information than was necessary, so the task does not require appreciation of the Second Maxim of Quantity. Neither should the statement be discredited on the grounds that it was irrelevant, as the puppet's statement was given in response to an enquiry about the character's performance. Appreciation of the Maxim of Relation is thus not required either. Additionally, since it is actually true that the character in the SI task does complete some of their task, the puppet's statement should also not be rejected on the grounds that it provided false information. Appreciation of the Maxim of Quality is therefore also not required. Further, the puppet's statement should not be rejected on the grounds that it is impolite, so the task does not require appreciation of the Maxim of Politeness.

Moreover, the CVT can be seen as more appropriate for a younger age group than the SI task which was established as suitable for 5- and 6-year-olds in Experiment 1. The age of participants selected for the current study: 4- and 5-years, was based on the findings of Eskritt, Whalen, and Lee (2008) who reported that 3-year-olds failed to demonstrate appreciation of two of three maxims assessed: the maxims of Quality and Politeness. However, they reported that 4-year-olds revealed appreciation of the maxims of Quality, Politeness and Relation.

The other tasks employed in Experiment 3 were chosen to be suitable for this younger age group. The Frog Simon Inhibitory and Dibbets Updating tasks, used to assess EF in Experiment 2, were also employed in Experiment 3. However, in Experiment 2, almost a third of the younger age group had failed to achieve a sufficient level of training in the updating task, to enable them to pass on to the test phase of this task. Given that the mean age of the younger group in Experiment 2 (68 months), corresponded to the older age group in the current study, it was anticipated that an even greater proportion of the younger age group in Experiment 3 would fail to achieve a sufficient level of updating training to enable them to move on to the test phase of this task. Therefore, in Experiment 3, updating scores were based only on performance in the updating training phase of the task. The children in Experiment 3 were not presented with the updating test trials employed in Experiment 2. The DCCS shifting task was also employed in Experiment 3, as Experiment 1 had established that the task was suitable for this younger age group. All three of Miyake et al.'s (2000) proposed components of EF: inhibition, updating and shifting, were therefore assessed in Experiment 3. ToM was assessed using the first- and second-order false belief tasks employed in Experiment 2.

In Experiment 3, it was hypothesised that all abilities would improve significantly with age, and that appreciation of false beliefs and EF abilities would be significantly related to performance on the CVT. Demonstration of a relationship between appreciation of false beliefs and EF abilities was also envisaged. It was further expected that the relationship between CVT performance and EF abilities would be mediated by appreciation of false beliefs.

5.2 Method

Participants These were 60 children aged between 4 and 5 years, split equally into two age groups: a younger group (M=56.5 months, range =48-59 months) and an older group (M= 67.07 months, range =60-71 months). These children were predominantly white and were recruited from primary schools located in middle class areas of Sheffield and Surrey. All children spoke English as their first language, and none were known to have any specific language impairment. Three other children (aged 59, 62 and 64 months) were excluded. Of the children excluded, one child was too distracted to complete all of the tasks presented, one child failed to complete initial training in the updating task and one child was unable to attend the second testing session due to illness.

Procedure. Children were tested individually, in a quiet area of their school, and all took part in eight tasks. Testing was conducted over the course of two sessions, lasting between approximately 10 and 20 minutes, and the order of task presentation was fixed. The first session commenced with the two first-order false-belief tasks, followed by the two second-order false-belief tasks, the DCCS task, the Frog Simon Inhibitory task, and the Dibbets Updating task. The second session was used to present the CVT.

First-order false belief tasks. These were identical to the first-order false belief tasks used in Experiments 1 and 2. Each task score had a possible range of 0-1.

Second-order false belief tasks. These were identical to the Second-order false belief tasks used in Experiment 2. As in the previous study, the total number of correct responses to the two test questions in each task was recorded separately for each task, with each task score having a possible range of 0-2. No children failed the first set of control questions.

DCCS. This was identical to the DCCS task used in Experiment 1. The post-switch score for this task had a possible range of 0-6 and no child failed to meet the pre-switch score criterion for task inclusion.

Frog Simon Inhibitory task. This was identical to the task used in Experiment 2. The score for this task had a possible range of 2-20 and no child failed to respond correctly to the first two slides.

Updating task. This was identical to the Updating task used in Experiment 2, except that children were not presented with Test trials in Experiment 3. Thus only the Training score was recorded. This had a possible range of 0-1.

CVT. The CVT was presented on a laptop and was an English translation of the task employed by Siegal, Iozzi, and Surian (2009). Children were introduced to three puppets: Mark, Lucy and Jane. Mark then proceeded to put 25 questions to Lucy and Jane. Both Lucy and Jane provided answers to each question. However, one of them (counterbalanced across questions) would provide an inappropriate response, which violated one of Grice's five conversational maxims. The other puppet would provide an acceptable response which didn't violate any of the maxims. For example:

Mark: "What food do you like?"

Lucy: "I like the sea."

Jane: "I like ice-cream."

In this trial, Lucy's response is inappropriate, as "the sea" is not a type of food and so her answer violates Grice's maxim of Relevance. However, Jane's response is appropriate as "ice-cream" is a type of food. Prior to the start of the puppet trials, children were told "We're going to see some puppets on the computer. One of them is going to ask some questions and the others are going to give some answers. Each time, one of the puppets is going to give a bit of a silly or a rude answer and one is going to answer well and I want to see if you can point to the puppet who answers well." Following each trial, children were asked "Can you point to the puppet that answered well?"

All four of Grice's conversational maxims, together with the Maxim of Politeness ("Be Polite), were represented in the CVT (see Appendix VIII for the CVT Script providing the full list of CVT items). Questions violating the different maxims were interspersed with each other (see Appendix IX for the sequence of violations). Presentation order of the questions was counterbalanced across participants, with half of the participants receiving the sequence of questions as it is read downwards in Appendix IX, and half receiving the order as it is read upwards.

5.3 Results

Score means and standard deviations are presented in Table 5.1.

Table 5.1: *Mean scores and standard deviations for EF, ToM and CU tasks employed in Experiment 3.*

Age group (in years)	4	5
No. of children	30	30
Age (months)	56.50 (2.39)	67.07 (3.80)
First first-order FB/1	0.87 (0.35)	1.00 (0.00)
Second first-order FB/1	0.70 (0.47)	0.93 (0.25)
Perner second-order FB/2	0.10 (0.40)	0.43 (0.77)
Sullivan second-order FB/2	0.93 (0.79)	1.40 (0.77)
DCCS/6	4.57 (2.57)	5.53 (1.55)
Frog Simon/20	17.87 (1.63)	18.73 (1.26)
Dibbets updating training/1	0.37 (0.49)	0.70 (0.47)
CVT/25	15.57 (3.10)	18.30 (3.44)
CVT Maxims/5:		
First Quantity	2.60 (0.93)	3.10 (1.27)
Second Quantity	2.83 (0.87)	2.70 (0.99)
Quality	3.10 (1.30)	4.07 (1.23)
Relation	3.27 (0.87)	3.93 (1.23)
Politeness	3.77 (1.10)	4.50 (0.78)

Note: FB=False belief task; DCCS=Dimensional Change Card Sort task; CVT=Conversational Violations Task.

Relation between EF and CVT performance

Table 5.2 displays the results of ANOVAs investigating the effects of EF, age and maxims on CVT performance and Table 5.3 reports the means and standard deviations of CVT scores for the weak and strong EF groups across maxims and age groups. Children's shifting EF was classed as strong if their scores for the DCCS task were 6 and weak if these were less than 6. Their inhibitory EF was classed as strong if scores for the Frog Simon task were 19 or greater and weak if these were less than 19. Updating EF was classed as strong if scores for the training phase of the Dibbets task

Table 5.2: *Results of 2 (age group: 4-year-olds vs. 5-year-olds) x 2 (EF level: shifting EF scores<6=weak vs. shifting EF scores=6=strong; inhibitory EF scores <19 =weak vs. inhibitory EF scores ≥19 =strong; updating EF scores <1 =weak vs. updating EF scores =1=strong) x 5 (type of maxim: First Quantity vs. Second Quantity vs. Quality vs. Relation vs. Politeness) ANOVAs on scores for the CVT measure. N.B. Shifting analyses did not include the age group variable.*

	df	F	p	partial Eta^2
Age group				
Inhibitory EF	1,280	13.982	**0.001**	0.048
Updating EF	1,280	12.070	**0.001**	0.041
EF level				
Shifting EF	1,140	4.455	**0.037**	0.031
Inhibitory EF	1,280	3.190	0.075	0.011
Updating EF	1,280	3.231	0.073	0.011
Type of maxim				
Shifting EF	4,140	2.544	**0.042**	0.068
Inhibitory EF	4,280	14.575	**0.001**	0.172
Updating EF	4,280	15.558	**0.001**	0.182
Age group x EF level				
Inhibitory EF	1,280	1.036	0.310	0.004
Updating EF	1,280	0.585	0.445	0.002
Age group x type of maxim				
Inhibitory EF	4,280	1.631	0.167	0.023
Updating EF	4,280	1.766	0.136	0.025
EF level x type of maxim				
Shifting EF	4,140	1.542	0.193	0.042
Inhibitory EF	4,280	0.503	0.734	0.007
Updating EF	4,280	0.634	0.639	0.009
Age group x EF level x type of maxim				
Inhibitory EF	4,280	0.769	0.546	0.011
Updating EF	4,280	0.178	0.950	0.003

Table 5.3: *CVT score means and standard deviations for weak and strong EF groups across the 4-year-old and 5-year-old age groups. N.B. shifting EF was only considered for the 4-year-olds.*

Age group	EF group	Maxim	Mean CVT score	Standard deviation
4	Shifting weak	First Quantity	2.63	0.74
		Second Quantity	2.88	0.99
		Quality	2.75	1.75
		Relation	3.00	0.76
		Politeness	2.88	0.99
	Shifting strong	First Quantity	2.59	1.01
		Second Quantity	2.82	0.85
		Quality	3.23	1.11
		Relation	3.36	0.90
		Politeness	4.09	0.97
4	Inhibition weak	First Quantity	2.47	1.01
		Second Quantity	2.76	0.90
		Quality	3.24	1.03
		Relation	3.12	0.70
		Politeness	3.76	1.09
	Inhibition strong	First Quantity	2.77	0.83
		Second Quantity	2.92	0.86
		Quality	2.92	1.61
		Relation	3.46	1.05
		Politeness	3.77	1.17
5	Inhibition weak	First Quantity	2.80	1.32
		Second Quantity	2.90	0.99
		Quality	3.70	1.57
		Relation	3.60	1.58
		Politeness	4.10	0.99
	Inhibition strong	First Quantity	3.25	1.25
		Second Quantity	2.60	1.00
		Quality	4.25	1.02
		Relation	4.10	1.02
		Politeness	4.70	0.57
4	Updating weak	First Quantity	2.63	0.96
		Second Quantity	2.79	0.92
		Quality	2.89	1.24
		Relation	3.21	0.98
		Politeness	3.79	1.18
	Updating strong	First Quantity	2.55	0.93
		Second Quantity	2.91	0.83
		Quality	3.45	1.37
		Relation	3.36	0.67
		Politeness	3.73	1.01
5	Updating weak	First Quantity	2.78	1.20
		Second Quantity	2.56	1.01
		Quality	3.67	1.00
		Relation	3.56	1.01
		Politeness	4.56	0.53
	Updating strong	First Quantity	3.24	1.30
		Second Quantity	2.76	1.00
		Quality	4.24	1.30
		Relation	4.10	1.30
		Politeness	4.48	0.87

were 1 and weak if these were 0. The weak and strong division criteria were selected on the basis that they enabled as equal a split as possible in the number of children assigned to the two EF ability groups for each task. Age group had two levels: 4-year-olds vs. 5-year-olds. Type of maxim had five levels: First Quantity vs. Second Quantity vs. Quality vs. Relation vs. Politeness. For weak versus strong shifting groups respectively, N was 8:22 for the 4-year-olds and 3:27 for the 5-year-olds. Since the 5-year-olds were performing at ceiling on the DCCS task, only the shifting data of the 4-year-olds was analysed. Thus, analyses looking at the effects of shifting and maxims on CVT performance did not also investigate age effects. For weak versus strong inhibitory groups respectively, N was 17:13 for the 4-year-olds and 10:20 for the 5-year-olds. For weak versus strong updating groups respectively, N was 19:11 for the 4-year-olds and 9:21 for the 5-year-olds.

There were significant main effects for shifting EF, age group and type of maxim. Children with strong shifting EF attained significantly greater CVT maxim scores (M=3.22, SD=1.09) than children with weak shifting EF (M=2.83, SD=1.06), $F(1,140)= 4.455$, $p<0.05$. Although, the uneven distribution of N across the weak and strong shifting EF groups indicates that the findings of the shifting analysis should be treated cautiously. Furthermore, 5-year-olds had significantly higher CVT maxim scores (M=3.66, SD=1.28) than 4-year olds (M=3.11, SD=1.09), $F(1,280)= 12.070$ and 13.982, p's<0.001. Scheffe tests following the significant main effect for maxims, $F(4,280)= 14.575$ and 15.558, p's <0.001, $F(4,140)=2.544$, $p<0.05$, found that the data for the inhibitory and updating analyses revealed performance on CVT items reflecting the Maxims of First Quantity (M=2.85, SD=1.13) and Second Quantity (M=2.77, SD=0.93) was significantly lower than performance on items reflecting the Maxims of Quality (M=3.58, SD=1.34), Relation (M=3.60, SD=1.11) and Politeness

(M=4.13, SD=1.02), p's <0.01. Consistent with this, Scheffe tests on data for the shifting analysis revealed that performance on CVT items reflecting the Maxims of First Quantity (M=2.60, SD=0.93) and Second Quantity (M= 2.83, SD=0.87) was significantly lower than performance on items reflecting the Maxims of Politeness (M=3.77, SD=1.10), p's <0.05.

Role of ToM in the relationship between EF and CU

Regressions were conducted on CVT scores, using composite measures of ToM comprising either total scores on the first-order ToM tasks, total scores on the second-order ToM tasks, total scores on both the first- and second-order ToM tasks or total scores both on the first- and second-order ToM tasks not including points for appropriate second-order justifications (i.e. total predictive scores across the first- and second-order ToM tasks). Total scores across just the first-order ToM tasks did not prove a significant fit, $F(1,58) = 2.094$, $p>0.10$. However the remaining ToM composite measures accounted for a significant amount of variance. When total scores across just the second-order ToM tasks were considered, $F(1,58)= 5.468$, $p<0.05$, when total scores across the first- and second-order ToM tasks were considered, $F(1,58)= 5.786$, $p<0.05$ and when only total predictive scores across the first- and second-order ToM tasks were considered, $F(1,58)= 5.362$, $p<0.05$.

Nevertheless, when ToM was entered into regressions on CVT performance with age and DCCS scores, ToM scores did not make a significant contribution. Both age and DCCS scores predicted a significant amount of variance (see Table 5.4) accounting for a quarter of the variance in CVT scores, total $R^2_{adj} = 25.6\%$. However no composite ToM measure accounted independently for a significant amount of variance (see Table 5.5).

Table 5.4: *Summary table of regression analysis looking at the contributions of age and DCCS scores to CVT performance when age was entered first.*

	B	SE B	β	t	p
Step 1					
Constant	2.511	4.240		0.592	0.556
Age in months	0.233	0.068	0.410	3.419	**0.001**
Step 2					
Constant	3.608	3.992		0.904	0.370
Age in months	0.168	0.068	0.295	2.488	**0.016**
DCCS scores	0.580	0.194	0.355	2.994	**0.004**

Table 5.5: *Step 3 results for regression analyses looking at the contributions of age, various ToM composite measures and DCCS performance to CVT scores.*

	B	SE B	β	t	p
1st order ToM					
Constant	3.661	4.032		0.908	0.368
Age	0.162	0.073	0.284	2.222	**0.030**
ToM composite	0.187	0.798	0.029	0.235	0.815
DCCS	0.579	0.196	0.355	2.963	**0.004**
2nd order ToM					
Constant	5.337	4.212		1.267	0.210
Age	0.128	0.075	0.225	1.718	0.091
ToM composite	0.484	0.391	0.155	1.239	0.221
DCCS	0.590	0.193	0.361	3.054	**0.003**
1st + 2nd order ToM					
Constant	4.945	4.169		1.186	0.241
Age	0.128	0.077	0.225	1.673	0.100
ToM composite	0.347	0.317	0.141	1.093	0.279
DCCS	0.585	0.194	0.358	3.023	**0.004**
1st + 2nd order ToM Predict					
Constant	4.526	4.048		1.118	0.268
Age	0.129	0.075	0.227	1.735	0.088
ToM composite	0.539	0.446	0.151	1.207	0.232
DCCS	0.589	0.193	0.360	3.047	**0.004**

Further regressions were conducted on the individual maxims reflected in the CVT, using age, ToM composite measures and DCCS scores as predictors. None of these factors explained a significant amount of variance in scores for CVT items reflecting the Maxims of First and Second Quantity. However, age alone accounted for a significant amount of variation in scores for CVT items reflecting the Maxim of Relation explaining 18.3% of the variance, DCCS performance alone provided the best model for variation in scores for CVT items reflecting the Maxim of Politeness explaining 36.8% of the variance and both age and DCCS performance independently accounted for a significant amount of variation in scores for CVT items reflecting the Maxim of Quality together explaining 21.4% of the variance. ToM did not contribute significantly to any model of maxim score variance. Table 5.6 presents the final steps in regressions focused upon age and DCCS predictors of maxim scores.

Table 5.6: *Step 2 results for regression analyses looking at the contributions of age, and DCCS performance to scores for CVT items reflecting individual maxims.*

	B	SE B	β	t	p
Step 2 –First Quantity					
Constant	1.410	1.492		0.945	0.349
Age	0.019	0.025	0.105	0.763	0.449
DCCS	0.049	0.072	0.094	0.681	0.499
Step 2 – Second Quantity					
Constant	3.489	1.234		2.828	0.006
Age	-0.011	0.021	-0.073	-0.523	0.603
DCCS	-0.009	0.060	-0.021	-0.153	0.879
Step 2 –Quality					
Constant	-1.505	1.563		-0.963	0.340
Age	0.067	0.026	0.310	2.547	**0.014**
DCCS	0.182	0.076	0.293	2.403	**0.020**
Step 2 –Relation					
Constant	-0.953	1.313		-0.726	0.471
Age	0.066	0.022	0.369	2.973	**0.004**
DCCS	0.093	0.064	0.180	1.451	0.152
Step 2 –Politeness					
Constant	1.167	1.050		1.112	0.271
Age	0.026	0.018	0.160	1.480	0.144
DCCS	0.265	0.051	0.564	5.206	**0.001**

Relationship between EF and ToM performance

Tables 5.7 and 5.8 display the results of 2 (age group) X 2 (shifting, inhibitory and updating EF level: weak vs. strong) ANOVAs with ToM scores as the dependent measures. Children's shifting EF was classed as strong if their scores for the DCCS task were 6 and weak if these were less than 6. Their inhibitory EF was classed as strong if scores for the Frog Simon task were 19 or greater and weak if these were less than 19. Updating EF was classed as strong if scores for the training phase of the

Dibbets task were 1 and weak if these were 0. As before, weak and strong division criteria were selected on the basis that they enabled as equal a split as possible in the number of children assigned to the two EF ability groups for each task. Age group had two levels: 4-year-olds vs. 5-year-olds. For weak versus strong shifting groups respectively, N was 8:22 for the 4-year-olds and 3:27 for the 5-year-olds. Since the 5-year-olds were performing at ceiling on the DCCS task, only the shifting data of the 4-year-olds was analysed. Thus, analyses looking at the effects of shifting on ToM scores did not also investigate age effects. For weak versus strong inhibitory groups respectively, N was 17:13 for the 4-year-olds and 10:20 for the 5-year-olds. For weak versus strong updating groups respectively, N was 19:11 for the 4-year-olds and 9:21 for the 5-year-olds.

For both inhibitory and updating EF, there were consistent significant age group main effects, with 5-year-olds having higher ToM scores than 4-year-olds, F's ≥ 6.089, p's $<.03$. However, there were no significant main or interaction effects involving EF for any of the analyses. Nevertheless, as indicated previously, the disproportionality of N across the shifting EF groups entails that the results of analyses using shifting EF should be treated cautiously.

Age and EF performance

Table 5.10 displays the results of ANOVAs investigating the effect of age group on EF. A main effect of age group was uncovered, with 5-year-olds again demonstrating greater inhibitory and updating EF than 4-year-olds (see Table 5.1 for age group means and standard deviations), $F(1,58)= 5.298$ and 12.100, p's<0.05. Neither shifting EF nor updating EF was found to correlate significantly with inhibition EF scores when age was partialled out, $r=0.015$, $p>0.40$ and $r=0.004$,

$p>0.40$ respectively. Similarly, partial correlations covarying out the effects of age revealed that shifting EF was not related to updating EF, $r=0.038$, $p>0.30$.

Table 5.7: *Results of 2 (age group) x 2 (EF level) ANOVAs on the first order composite and second order composite ToM measures for the shifting, inhibition and updating EF measures. N.B. Shifting analyses did not include the age group variable.*

		df	F	p	partial Eta²
Age group					
First order ToM:	Inhibitory EF	1,56	6.746	**0.012**	0.108
	Updating EF	1,56	6.264	**0.015**	0.101
Second order ToM:	Inhibitory EF	1,56	6.089	**0.017**	0.098
	Updating EF	1,56	6.864	**0.011**	0.109
EF level					
First order ToM:	Shifting EF	1,28	0.102	0.752	0.004
	Inhibitory EF	1,56	0.000	0.999	0.000
	Updating EF	1,56	0.011	0.917	0.000
Second order ToM:	Shifting EF	1,28	0.012	0.915	0.000
	Inhibitory EF	1,56	1.100	0.299	0.019
	Updating EF	1,56	0.010	0.923	0.000
Age group x EF Level					
First order ToM:	Inhibitory EF	1,56	0.128	0.721	0.002
	Updating EF	1,56	0.115	0.736	0.002
Second order ToM:	Inhibitory EF	1,56	1.581	0.214	0.027
	Updating EF	1,56	1.587	0.213	0.028

Table 5.8: *Results of 2 (age group) x 2 (EF level) ANOVAs on the first and second order ToM composite and the first and second order predictive ToM composite measures for the shifting, inhibition and updating EF measures. N.B. Shifting analyses did not include the age group variable.*

		df	F	p	partial Eta²
Age group					
1st & 2nd ToM:	Inhibitory EF	1,56	9.013	**0.004**	0.139
	Updating EF	1,56	9.534	**0.003**	0.145
1st & 2nd Predict ToM:	Inhibitory EF	1,56	7.568	**0.008**	0.119
	Updating EF	1,56	7.221	**0.009**	0.114
EF level					
1st & 2nd ToM:	Shifting EF	1,28	0.052	0.822	0.002
	Inhibitory EF	1,56	0.705	0.405	0.012
	Updating EF	1,56	0.014	0.905	0.000
1st & 2nd Predict ToM:	Shifting EF	1,28	0.021	0.886	0.001
	Inhibitory EF	1,56	0.210	0.648	0.004
	Updating EF	1,56	0.122	0.728	0.002
Age group x EF Level					
1st & 2nd ToM:	Inhibitory EF	1,56	1.318	0.256	0.023
	Updating EF	1,56	1.310	0.257	0.023
1st & 2nd Predict ToM:	Inhibitory EF	1,56	0.904	0.346	0.016
	Updating EF	1,56	1.170	0.284	0.020

Table 5.9: *Composite ToM score means and standard deviations for weak and strong EF groups across the 4-year-old and 5-year-old age groups. N.B. shifting EF was only considered for the 4-year-olds.*

Age group	ToM composite	EF group	Mean ToM score	Standard deviation
4	1st order	Shifting weak	1.50	0.76
		Shifting strong	1.59	0.67
		Inhibition weak	1.59	0.62
		Inhibition strong	1.54	0.78
		Updating weak	1.58	0.69
		Updating strong	1.55	0.69
	2nd order	Shifting weak	1.00	1.31
		Shifting strong	1.05	0.90
		Inhibition weak	1.06	0.90
		Inhibition strong	1.00	1.16
		Updating weak	1.16	1.12
		Updating strong	0.82	0.75
	1st+2nd order	Shifting weak	2.50	1.77
		Shifting strong	2.64	1.33
		Inhibition weak	2.65	1.22
		Inhibition strong	2.54	1.71
		Updating weak	2.74	1.56
		Updating strong	2.36	1.21
	1st+2nd order predictive	Shifting weak	2.25	1.28
		Shifting strong	2.32	1.09
		Inhibition weak	2.35	1.00
		Inhibition strong	2.23	1.30
		Updating weak	2.37	1.17
		Updating strong	2.18	1.08
5	1st order	Inhibition weak	1.90	0.32
		Inhibition strong	1.95	0.22
		Updating weak	1.89	0.33
		Updating strong	1.95	0.22
	2nd order	Inhibition weak	1.40	0.70
		Inhibition strong	2.05	1.23
		Updating weak	1.56	1.24
		Updating strong	1.95	1.07
	1st+2nd order	Inhibition weak	3.30	0.68
		Inhibition strong	4.00	1.34
		Updating weak	3.44	1.42
		Updating strong	3.90	1.09
	1st+2nd order predictive	Inhibition weak	2.80	0.42
		Inhibition strong	3.15	0.75
		Updating weak	2.78	0.83
		Updating strong	3.14	0.57

Table 5.10: *Results of 2 (age group) ANOVAs on the shifting, inhibition and updating EF measures.*

	df	F	p	partial Eta^2
Shifting EF	1,58	3.117	0.083	0.051
Inhibitory EF	1,58	5.298	**0.025**	0.084
Updating EF	1,58	12.100	**0.001**	0.173

ToM performance

Scores for the two first-order ToM tasks were found to be strongly positively correlated, $r = 0.391$, $p<0.01$. However, scores for the two second-order tasks did not appear to be related, $r = -0.210$, $p>0.05$. A t-test revealed that scores for the first, first-order ToM task presented were significantly greater than scores for the second, first-order ToM task presented (see Table 5.1 for age group means and standard deviations), $t(59)=2.427$, $p<0.05$. A further t-test indicated that scores for the first, second-order ToM task presented were significantly lower than scores for the second, second-order ToM task presented (see Table 5.1 for age group means and standard deviations), $t(61)=7.619$, $p<0.001$.

CVT performance

The coefficient for internal consistency of the CVT task demonstrated reasonably good reliability, Cronbach's alpha = 0.613. One-sample t-tests revealed that children scored significantly above chance for each group of CVT items reflecting an individual maxim or rule: The Maxim of First Quantity ($M=2.85$, $SD=1.13$), $t(59)=2.394$, $p<0.05$; Maxim of Second Quantity ($M=2.77$, $SD=0.93$), $t(59)=2.228$, $p<0.05$; Maxim of Quality ($M=3.58$, $SD=1.34$), $t(59)= 6.243$, $p<0.001$; Maxim of Relation ($M=3.60$, $SD=1.11$), $t(59)=7.692$, $p<0.001$; Maxim of Politeness ($M=4.13$, $SD=1.02$), $t(59)=12.449$, $p<0.001$.

5.4 Discussion

This study was concerned with extending the results of Experiment 2 by examining the relationship between all three of the proposed components of EF and a broad measure of CU: the CVT.

In line with expectations, and in support of the masked competence account of the emergence of CU, a relationship was found between EF, assessed in terms of DCCS performance, and the demonstration of CU assessed using the CVT. Although the skewed nature of DCCS performance indicates that the relationship between shifting EF and CVT performance should be considered tentatively.

Although previous research has revealed relationships between EF and other specific aspects of CU such as the ability to take account of common ground information (Nilsen & Graham, 2009) and to compute SIs (Chapter 4), this is the first time that a relationship has been demonstrated between EF and a general measure of CU. However, in contrast to expectations and standard accounts of CU, although in line with the findings from Experiments 1 and 2, analyses did not reveal convincing evidence of a significant association between CU and ToM measures. Although ToM initially appeared to predict CVT performance, this effect disappeared once age and EF ability were considered. The lack of relation between ToM and performance on the CVT is at odds with the findings of Surian, Baron-Cohen and van der Lely (1996) and Tedoldi, Surian and Siegal (2005), both of whom found a relation between ToM and CVT performance. However, the relationship found by Surian, Baron-Cohen and van der Lely (1996) was demonstrated by children with autism. Because this population is known to demonstrate difficulties with ToM, the children were given first-order false-belief tasks typically passed by 3- to 4-year-olds, despite having a mean mental age of 5 years 7 months. Indeed, in support of the use of these tasks, only 3 of the 8 children with autism (37.5 %) passed these tasks. The relationship found by Tedoldi, Surian and Siegal (2005) was demonstrated by a sample of deaf children with a mean age of approximately 9 years. However these children also appeared to demonstrate a ToM deficit, with only 24 % scoring correctly on the first-order false belief tasks (in comparison to the 79 % of hearing children with a mean age of approximately 7 years scoring correctly).

In both the aforementioned studies, the relationships found between ToM and CU were borne by associations between first-order ToM and CVT performance. However, in the current study, the children tested were performing at ceiling on the first-order tasks, 93% scored correctly on the first task presented, 81.5% on the second task. This would be expected of typically developing children of the age sampled (4 and 5 years), and provided the rationale for including second-order false belief tasks in the study. The age of participants selected for the study had been based more on their expected performance on the CVT, than the suitability of ToM tests. Although there were no ceiling effects in the second order tasks (the mean score across both tasks was 1.43 out of 4), the lack of association between second-order ToM and CVT performance does not contrast with the Surian, Baron-Cohen and van der Lely (1996) or Tedoldi, Surian and Siegal (2005) studies, in which second-order ToM ability was not investigated. Such findings might simply reflect the fact that first-order ToM plays a more central role in success on the CVT than second-order ToM.

The fact that the relationship between first-order ToM and CVT performance appears restricted to populations demonstrating ToM deficits, suggests that although ToM plays a role in CVT performance, it has a greater significance for those whose ToM is less well developed. It could well be that once ToM comes online, other factors which don't come into play without ToM appreciation, begin to play more of a role.

Furthermore, also in contrast to expectations, and Experiment 2, analyses failed to reveal an association between ToM and EF measures. This is also at odds with the findings of Carlson and Moses (2001) and Hala, Hug, and Henderson (2003) who demonstrated a relationship between ToM and inhibitory ability and Frye, Zelazo and Palfai (1995) and Perner, Lang and Kloo (2002) who reported an association between ToM and DCCS performance. The age of the children tested in the first three of these studies was younger than the age tested in the current study. However, the suggestion that the older age of children in the current study might have reduced performance variation, and so the likelihood of finding evidence for the ToM-EF relationship, appears to be weakened by Perner, Lang and Kloo (2002).

Perner, Lang, and Kloo (2002) found evidence for a relationship between ToM and performance on the DCCS in a sample of 3-, 4- and 5-year-olds. Perner et al. demonstrated this association using only first-order false belief tasks. However, they reported that only 66% of their participants scored correctly on the traditional false belief tasks they employed, and an even smaller proportion (51.7%) on a more contemporary task used. This contrasts greatly with the 93% of participants who scored correctly on the first traditional first-order task presented and the 81.5% who passed on the second traditional first-order task presented in the current study. Such a discrepancy in pass-rates can probably be attributed to the inclusion of 3-year-olds in the Perner, Lang and Kloo study who would be less likely to pass first-order ToM tasks, and so might be expected to have brought overall performance measures down below the ceiling functioning demonstrated in the current study. Therefore, the initial suggestion that the older age of children in the current study might have reduced performance variation, and so the likelihood of finding evidence for the ToM-EF relationship, would appear plausible with more sensitive measures.

Thus, as in the previous studies, the masked competence account has been partially supported. EF has again been found to play a role in the demonstration of CU, and this time the effect has been demonstrated using the CVT. However, because the current study did not reveal relationships between ToM and CU or between EF and ToM, it is still not clear that the effect of EF on CU is due to its impact on ToM.

As suggested in Section 4.4, the failure to find significant relationships with ToM might be due to the inappropriateness of the ToM measures used for the age range tested. The validity of the ToM scores produced in Experiments 1 and 2 was questioned in Sections 3.4 and 4.4. In line with the findings from the previous two studies, the results of the current study revealed that significantly more children failed the second, first-order ToM task they were presented with, than failed the first, first-order task. Furthermore, consistent with the previous pattern of results, whilst 3 children failed both tasks in the current study, indicating a genuine ToM deficit, 8 children passed the first task but went on to fail the second, whilst only one child demonstrated the opposite pattern. In view of the fact that first-order ToM tasks are typically passed by children of the age of the participants in this study, it seems highly likely that some of the variation in composite first-order ToM scores might have been reflecting fatigue rather than individual differences in ToM. Although, as in the previous two studies, the strong correlation found between performance on the two first-order tasks in the current study, suggests that the effects of fatigue were limited.

Second-order ToM tasks were included to differentiate between children with, and children without, a more advanced level of ToM, and so increase ToM score variation. Such tasks are typically passed by 6- to 7-year-olds. Just as in Experiment

2, in Experiment 3 significantly more children failed the first, second-order task presented, than failed the second, second-order task. This suggests that fatigue was unlikely to account for much of the composite second-order score variation. However, also in line with the findings from Experiment 2, performance on the two second-order tasks did not appear to be related. This suggests that a considerable degree of score variation for at least one of the second-order tasks was not reflective of individual differences in ToM.

As argued in Section 4.4, analysis of performance on both the first- and second-order ToM tasks indicates that a considerable degree of variance in ToM scores might have been reflecting a factor other than ToM ability. Therefore the failure to find relationships with ToM in the current study, could have been attributable to the inappropriateness of the specific ToM measures used. Unfortunately, there is a paucity of evidence to suggest a ToM task suitable for the precise age range targeted in this study. However, a scale of ToM tasks put forward by Wellman and Liu (2004) suggests that children find tasks which assess appreciation of the difference between real and apparent emotion, more difficult than traditional first-order ToM tasks. Furthermore, Wellman and Liu indicated that the mean age at which children pass tasks measuring awareness of the differentiation between real and apparent emotion was just over 5 years. This is younger than the age at which children are reported to pass second-order ToM tasks. Wellman and Liu's report indicates that, if tasks assessing appreciation of the difference between real and apparent emotion had been employed in place of the second-order ToM tasks in this study, a greater degree of variation in ToM score might have been produced. It is expected that greater variation in ToM scores would have resulted in significant relationships between ToM and EF and between ToM and CU.

It is also worth noting that although updating ability appeared to be related to the appreciation of SIs in Experiment 2, this did not seem to be associated with the more general CVT measure of CU in Experiment 3. However, the SI task assesses appreciation of only one component of the CVT: the First Maxim of Quantity. The CVT requires appreciation of four more conversational maxims, in addition to the First Maxim of Quantity. It could be that updating ability play a rather narrow role in CU, associated specifically with the First Maxim of Quantity, whilst shifting ability plays a more general role associated with appreciation of all of the conversational maxims. Shifting ability was not measured in Experiment 2, as the task was deemed too easy for the age range tested, but had a more age-appropriate measure been employed, it is expected that a relationship would have been revealed between shifting ability and the appreciation of SIs. Such a scenario fits in well with the general finding that the Maxim of First Quantity is grasped after the rest of the conversational maxims (Siegal, Iozzi, & Surian, 2009; Surian, Baron-Cohen, & van der Lely, 1996). One would expect appreciation of a maxim to emerge later on if it placed additional demands on EF. Despite the absence of significant EF and maxim interaction effects, the significant links between DCCS scores and responses on the Quality and Politeness items are consistent with the notion that distinct EF components can exert a differential influence on the CVT maxim sub-scales.

In line with the findings of Experiments 1 and 2, in which the EF components assessed appeared to be unrelated: inhibitory and shifting ability in Experiment 1, inhibitory and updating ability in Experiment 2, the three EF tasks components examined in Experiment 3: inhibitory, updating and shifting ability were also not found to be related. This is in concordance with the suggestion of Miyake et al. (2000) that inhibition, shifting and updating ability are differentiated components of EF, and leaves open the possibility that different EF factors might be affecting CU in different ways, as has been suggested in the preceding paragraph.

The CVT was found to demonstrate good internal reliability, indicating that that all of the items in the task were tapping the same basic ability: CU. However, the level of performance in individual trials appeared to depend on the specific maxim being tapped. The children in this study demonstrated a greater understanding of the maxims of Quality, Relation and Politeness than the maxims of First or Second Quantity. This is in concordance with the findings of Siegal, Iozzi, and Surian (2009) who reported the same performance pattern in bilingual children aged between 3 years 9 months and 6 years, and consistent patterns in monolingual Italians and Slovenians aged between 3 years 6 months and 6 years. The implication that children have more difficulty appreciating the Maxim of First Quantity than some of the other conversational maxims is also in concordance with the findings of Surian, Baron-Cohen and van der Lely (1996). These authors reported that children with SLI (age M= 11 years 10 months) and typically developing children (age M= 6 years 7 months) demonstrated only chance performance on CVT items tapping appreciation of the Maxim of First Quantity. However, the same children demonstrated above chance performance on CVT items tapping appreciation of the rest of the conversational maxims.

It has been suggested (by an anonymous reviewer of the Siegal, Iozzi and Surian 2009 paper) that poor performance on the CVT items representing the Maxim of First Quantity, can be attributed to the inappropriateness of the exact statements used in the CVT to represent this Maxim. The Maxim of First Quantity calls for sufficient information to be provided in conversation, in an effort to avoid ambiguity and so enable precise meanings to be identified. This maxim was represented in the

CVT by pairs of statements, one of which was intended to be ambiguous, and one of which was considered detailed enough to communicate an exact meaning. However, it has been proposed that some of the responses which had been intended to be ambiguous, could in fact be construed as appropriate (and thus adequately informative) alternatives to additional possible responses. To illustrate this, consider the following exchange (Appendix VIII, Item A1):

MARK : What would you like to buy in this pet-shop?

LUCY : An animal.

JANE : A cat.

It had been intended that children should interpret Lucy's response of "An animal" as ambiguous and so inappropriate. Unlike Jane, she doesn't inform Mark which animal she wants, and so doesn't enable Mark to identify what exactly she would like to buy. However, a child might not initially construe the situation as requiring information alluding to which animal is to be bought. It could be they consider the first priority to be to establish the type of thing to be bought, in this case an animal, instead of pet food for example, which is often bought from a pet shop.

The exchange below (Appendix VIII, Item A2), can be reinterpreted in a similar fashion:

MARK : How would you like your tea?

LUCY: with milk.

JANE : In a cup

It had been intended that children should interpret Jane's response of "In a cup" as ambiguous and so inappropriate. Unlike Lucy, she doesn't tell Mark what she wants in her drink, and so doesn't enable Mark to identify what he should put in the tea. However, a child might consider it an initial priority to establish that the tea is actually to be made in a vessel they don't have too much experience with at this age: a cup, rather than a tumbler, which 4- and 5-year-olds more standardly drink from.

In support of the suggestion that the Maxim of First Quantity was inadequately represented in the CVT, Cronbach's Alpha rose from 0.613 to 0.650 when the five items used to represent this maxim were removed. This indicates that the CVT is a slightly more reliable measure of CU without the First Quantity items. However, since DCCS performance had not been found to predict scores for CVT items reflecting the Maxim of First Quantity, these items did not affect the relationship established between EF and CU in Experiment 3.

Of the 11 children who scored less than perfectly on the DCCS (i.e. scored less than 6), none scored over 20 on the CVT and most (72.7%) scored 14 or less. This contrasts with the 49 who did well on the DCCS (i.e. scoring 6), of whom only 14.3% scored 14 or less on the CVT, with 75.5% scoring 16 or more. However, whilst no child who performed poorly on the DCCS gained a high CVT score, supporting the proposal that the shifting aspect of EF plays a necessary role in the emergence of CU, nearly a quarter of children who performed well on the DCCS nevertheless performed rather poorly on the CVT (24.5 % scored 15 or less). This indicates that the shifting aspect of EF is necessary but not sufficient for the emergence of CU. Other factors such as breadth of general knowledge might be expected to affect CU. It could be that those children with relatively poor general knowledge are less familiar with different contexts and so are less able to make inferences about the influence of these contexts on the meaning of language used.

Although it has been suggested that the effect of EF on CU is mediated by ToM, it is unlikely that ToM is the factor identified here, as it is not clear to see why children would demonstrate unequal competence across the different maxims/conversational rules. It seems much more likely that this other factor is world knowledge, whose demands seem much more likely to differ across the maxim-based groups of CVT items. For the CVT items reflecting the Second Maxim of Quantity for instance, children need to possess quite specific knowledge: that puppies have four legs, that eggs are boiled in a saucepan. Similarly, for the CVT items reflecting the First Maxim of Quantity, children need to bring to mind comprehensive food knowledge relating to the fact that there are a number of different types of cuisine one could eat, and a similarly broad knowledge of gifts relating to the fact that there are many different types of presents that could be received. However, for CVT items reflecting the Maxim of Quality, children need only to know basic facts relating to more common knowledge such as whether a dog could be in the clouds, or whether people live on the moon. For CVT items reflecting the Maxim of Politeness, children need only to recognise when someone is being rude.

For items reflecting the Maxim of Relevance, children need to know if the sea is a type of food, or if wearing trousers is a type of holiday activity, which would seem to tax general knowledge slightly more than the items reflecting the Maxims of Quality and Politeness, but to a lesser degree than the items reflecting the two Maxims of Quantity. Indeed in support of this proposal, whilst 21 and 29 children achieve a perfect score on the items reflecting the Maxims of Quality and Maxim of Politeness respectively, and only 6 and 3 children achieve this on items reflecting the First and Second Maxim of Informativeness respectively, an intermediate number of children: 16, manage to achieve the top score for items reflecting the Maxim of Relevance.

Siegal, Iozzi and Surian (2009) looked at the effect of bilingualism on CVT performance and investigated the role that inhibitory and shifting aspects of EF played in this relationship. The day-night task employed in Experiment 1 of the current study, and the DCCS used in Experiments 1 and 3 of the current study, were used to assess inhibitory and shifting ability respectively. They did not look directly at the relation between either aspect of EF and CVT performance, but investigated whether either inhibitory or shifting proficiency could account for the bilingual advantage in CVT performance. Siegal et al. did not find a significant relationship between bilingualism and inhibitory ability, but found that Italian-Slovenian bilinguals and Slovenian monolinguals outperformed Italian monolinguals on shifting ability. However, they reported that entering shifting ability as a covariate did not remove the bilingual advantage in CVT performance.

The CVT performance of the children in this study was compared to that of the Italian monolinguals in Siegal et al's Experiment 1. The 4- and 5-year-olds tested in this study (M=61.78 months) achieved means scores of 2.85, 2.77, 3.58, 3.60 and 4.13 for groups of CVT items assessing the appreciation of the maxims of First Quantity, Second Quantity, Quality, Relation and Politeness respectively. This can be contrasted with Siegal et al.'s sample of 4- to 6-year-olds (M=64 months), who achieved mean scores of 3.37, 3.26, 3.84, 3.84 and 4.26 respectively in these trials. The Siegal et al. sample appear to be scoring more highly in all of the maxim-based item groups. However, although Siegal et al.'s sample was older, this was only by 3 months. Such a slight age advantage would not be expected to result in the consistent display of enhanced performance across the different groups of CVT items. However, the performance superiority of the Siegal et al. sample might have been attributable a difference in CVT task demands. In the Siegal et al. study, children were directed to identify the silly or rude responses in each statement pair. However, children in the current study were asked to identify the appropriate statement in each pair. It could well be that the silly or rude responses are actually more salient to the children precisely because of their inappropriateness. This could mean that children in the current study who were asked to identify the appropriate statements, had to exert inhibitory ability to overcome their attention to the inappropriate responses, to enable them to respond correctly in this study. This inhibitory demand would not have been made of the Siegal et al. sample, who in contrast, were asked to identify the inappropriate responses.

In line with expectations, and as in Experiment 1, all the abilities measured in Experiment 3 improved significantly with age. The age range tested in Experiment 3 spanned only two years (4-5 years), as in Experiment 2 (5-6 years). This indicates that the failure to find a general significant association between performance and age in Experiment 2, was not due to the restricted age variance available to exert an effect. It suggests however, that age has more of an influence on the performance of younger children, recruited in both Experiments 1 and 3, but not Experiment 2.

5.5 Conclusions

Experiment 3 provided partial support for the masked competence account of CU. Although ToM was not found to demonstrate associations with either CU or EF, a relationship was discovered between the shifting component of EF and a broad measure of CU: the CVT. Experiment 4 was conducted to examine whether training children on the shifting component of EF would improve their CU. An EF training effect would provide further support for the masked competence account of CU.

Chapter 6: Experiment 4 – Effects of executive functioning training on conversational understanding.

6.1 Introduction

Experiment 3 investigated whether inhibitory, updating and shifting components of EF are related to the CVT as a broad measure of CU in 4- and 5-year-olds. Results indicated that shifting EF assessed using the DCCS was significantly associated with CVT scores. This finding is consistent with the proposal that EF enhances CU and concurs with the masked competence account of the emergence of CU. However, the relationship found between shifting EF and CU could be attributable to a third mediating factor which promotes both shifting EF and CU. Establishing that training EF results in enhanced CU would provide evidence of a more direct link between EF and CU.

A number of studies have attempted to train children's EF abilities. For example, Dowsett and Livesey (2000) trained 3-year-olds on the DCCS and an adaptation of the Change task (Logan & Burkell, 1986). The Change task, like the DCCS, requires shifting from a previously activated response set to an alternative response set. Dowsett and Livesey found that training children over the course of three sessions on these two EF tasks led to significant improvements in performance on a Go/No-Go EF task that requires production of a response in accord with the presence of a "go" stimulus and inhibition of the response in accord with the presence of a "no-go" stimulus. Similar improvements were neither demonstrated by a control no-intervention group nor by a control practice-task group who were not given EF training but practiced on the Go/No-Go task for three sessions.

As part of their investigation into the relationship between EF and ToM in 3- and 4-year-olds, Kloo and Perner (2003) also employed EF training using the DCCS.

In one condition, children were trained on the DCCS. In another condition, children were trained on ToM. In a control condition, children were trained on either number conservation or relative clauses. Kloo and Perner found that DCCS training significantly improved not only DCCS performance, but also ToM performance. The DCCS improvement shown by the DCCS training group was significantly greater than the non-significant DCCS improvement shown by the control group. However, although the ToM training group also improved significantly on the DCCS task, their performance was not significantly greater than that of the control training group. Furthermore, as with the control training, ToM training was not found to significantly improve ToM performance.

More recently, Mack (2007) reported that 3-year-olds given pre-training on the DCCS task performed significantly better on a post-training DCCS than controls who did not receive training. The pre-training comprised drawing attention to the two dimensions of shape and colour evident in the target and sorting cards, and explaining and demonstrating how each sorting card could be sorted in two different ways, one of which was in accord with the shape of the target cards and one of which was in accord with the colour of the target cards. Children then practiced sorting to each rule and were provided with feedback on their performance. Mack found children in the group who received pre-training and children in another group who did not receive training but were asked an irrelevant question between pre-switch and post-switch sorting performed significantly better than control children following the standard DCCS format (as illustrated in Section 3.2).

Tasks other than the DCCS have also been employed in EF training studies. Diamond, Barnett, Thomas, and Munro (2007) compared the influence of EF training with the impact of a balanced literacy program (dBL) on the EF ability of 4- and 5-

year-old U.S. preschoolers. The EF program promoted children's use of self-regulatory private speech, provided EF aids such as pictures to cue self-inhibition and encouraged use of planning. By contrast, behavioural control in the dBL program was predominantly teacher-, rather than self-imposed. EF ability was measured using a dot task and a flanker task, both of which required inhibition of pre-potent responses. In the dot task, children were required to follow rules which frequently required them to inhibit a pre-potent response. For example, they would be asked to press on the opposite side of a page that a flower appeared – a task which demanded inhibition of the pre-potent tendency to signal the same side on which an image appears. In the flanker task, children were asked to respond to a central shape (either a circle or a triangle) which often required inhibition of attention to a conflicting external shape. Diamond et al. reported that EF-trained children performed significantly better than dBL control group children in trials requiring inhibition in both the dot and flanker tasks.

Further evidence of the effectiveness of EF training was provided by Klingberg et al. (2005). These authors investigated whether EF training using visuo-spatial and verbal working memory (WM) span computer tasks could lead to EF improvements in 7- to 12-year-olds with ADHD. EF training comprised presentation of tasks taxing memory for multiple object-locations and phoneme, letter or digit strings. The precise quantity of items in a string was adjusted over trials to match each child's ability. Control training comprised presentation of the same WM span tasks, but the number of objects/phonemes/letters/digits to be remembered was kept at a minimum level of two to three items in each string. Children who underwent EF training were found to perform significantly better than children in the control-trained group, on a span-board visuo-spatial WM span task (requiring memory for object locations presented on a three-dimensional board rather than the computer grid used in training), a digit-recall verbal WM span task, a colour-word Stroop response-inhibition task and the Raven's Coloured Progressive Matrices non-verbal reasoning ability EF task.

However, EF training has not always been found to significantly improve EF ability. Rueda et al. (2005) used EF computer programs to train 4- and 6-year olds to track and anticipate stimuli, to resolve conflict and to enhance their inhibitory ability. Training effects were established by analysis of pre- and post-test performance on a Flanker EF task. Task performance was calculated in terms of conflict scores generated by subtracting reaction times for trials in which the central item was surrounded or flanked by congruent items from reaction times for trials in which the central item was flanked by conflicting items. Rueda et al. found that training led to enhanced functioning on the Flanker task. The children's low conflict scores were similar to an adult performance. Although the training effect was not found to reach significance (post-training conflict scores were not significantly smaller than pre-training scores), an electrophysiological investigation indicated that the training led to an altered distribution of task-related neural activity relative to the controls, resembling the distribution found in adults during the Flanker task.

Similarly, Fisher and Happé (2005) were not able to find a significant effect of EF training on EF ability. Fisher and Happé used EF and ToM training in a study of 6- to 15-year-olds with autistic spectrum disorders. Children with autism in the ToM training condition were encouraged to think of beliefs as "photos in the head", and children with autism in the EF training condition were taught to conceptualise the brain as a machine which used EF "tools" which help it to think. Children in the ToM and EF training groups performed significantly better on post-test and follow-up ToM tests relative to pre-test tasks, and showed greater improvement than children with autism in the control group, who improved in only one of four ToM tasks given at post-test and did not improve in any of the ToM tasks given at follow-up. However, EF-trained children failed to demonstrate enhanced performance on EF tasks given post-test or follow-up. Although the ToM-trained children performed significantly better on the DCCS EF task, the level of improvement was comparable to that of the children in the control group and so could not be attributed to ToM training.

Therefore, neither Rueda et al. (2005) nor Fisher and Happé (2005) reported significant effects of EF training on EF ability. Furthermore, when considering the comparable success of EF training studies, the bias to publish positive results ought also to be borne in mind (Scargle, 2000).

Nonetheless, the DCCS training program used by Kloo and Perner (2003) appeared to be effective, and was based on the DCCS task that was significantly related to CVT scores in Experiment 3. Therefore, this program formed the basis for the EF training in Experiment 4 to establish if shifting EF thought to be tapped by the DCCS task is directly related to performance on the CVT. Children aged 3 and 4 years were recruited for Experiment 4, since the DCCS task is first passed between 3 and 5 years of age (Zelazo, 2006). Thus, many of the 3- and 4-year-olds could be expected to fail the DCCS task and so require DCCS training.

In addition to DCCS training, children in Experiment 4 were given pre- and post DCCS tasks to determine whether the training was successful, and pre- and post-CVT tasks to assess whether the training led to the hypothesised improvement in CVT scores. Control number-conservation pre- and post-test tasks, taken from Kloo and Perner (2003), were also presented to establish whether the facilitatory effect of EF training was specific to EF and CVT competence, or whether it led to a broad improvement in post-test performance. Additionally, a picture vocabulary task (BPVS) was given to establish mental age.

It was initially intended that Experiment 4 would also include CVT training to investigate the possibility that the relationship between EF and CU stemmed from the positive influence of CU on EF. However, a pilot study indicated that a variety of CVT training methods did not raise CVT performance above chance levels. Similarly, the administration of control-task training would have enabled examination of the hypothesis that CVT and EF competence were facilitated by general training effects, rather than improvements in EF per se. However, the DCCS training condition was presented first and the data produced from this consequently rendered a control condition redundant.

In Experiment 4, it was hypothesised that all abilities would improve significantly with age, and that EF training would lead to a significant improvement on both the DCCS as a measure of shifting EF ability and the CVT. It was not expected that EF training would enhance performance on a number conservation control task.

6.2 Method

Participants. These were 29 children, ranging in age from 40- to 57-months (M = 46.93 months, SD = 5.26). The children were selected from a larger sample of 72 children aged between 40 and 57 months (M =49.07 months, SD = 5.21) who had been given an initial DCCS screening task. There were 35 children, ranging between 42 and 57 months-old (M =51.06 months, SD = 4.28), who were excluded from the study for attaining perfect scores of 6 out of 6 in the DCCS task. Exclusion would also have been a consequence of error-less performance on the pre-training CVT, reflected in scores of 12 out of 12. However, no child achieved a perfect pre-training CVT score. A further 8 children were also excluded from the study: 2 children (aged 41 and 48 months) were unable to identify shapes and/or colours appropriately, 1 child (56 months) was scared of the puppets used in the CVT and asked to withdraw, 1 child (44 months) missed her second testing session, 1 child (47 months) had to leave before completing their third testing session, 1 child (41 months) could not sort according to the DCCS pre-switch dimension and 2 children (53 and 55 months) were unable to complete the third testing session because the testing environment was too noisy. In total, 43 children ranging between 41- and 57-months-old (M =50.51 months, SD = 4.71) were excluded. The children included in the study were predominantly white, and were recruited from primary schools located in middle class areas of Sheffield and Surrey. All children spoke English as their first language, and none were known to have any specific language impairment.

Procedure. Children were tested individually, in a quiet area of their school, and were presented with tasks in a fixed order. Those who were not excluded on the basis of the criteria stated above, took part in three testing sessions, each of which lasted between 10 and 15 minutes.

Design. This consisted of three phases in which the children in the study were tested over three sessions.

Session 1 – Screening, pre-training measures and first training period. The initial DCCS screening task was given, which also served as the pre-training DCCS (PreDCCS). Children who scored less than 6 out of 6 on the DCCS went on to receive the pre-training CVT (PreCVT), the pre-training number conservation control task (PreCon) and the first phase of DCCS training.

Session 2 – Vocabulary measure and second training period. The British Picture Vocabulary Scale (BPVS) was administered and children received the second phase of DCCS training.

Session 3 – Post- training measures. Children received post-training measures of DCCS (Post2BoxDCCS and Post3BoxDCCS), CVT (PostNewCVT and PostOldCVT) and control-task (PostCon) performance.

DCCS

Screening/pre-training task (PreDCCS): The procedure was identical to that used in the DCCS task given in Experiment 1, except that children were asked to sort by size and colour. Half of the children were presented with target cards depicting a big red cat and a small blue snail, and sorting cards showing a big blue snail or a small red cat (see Appendix Xa). When sorting by colour, these children were instructed "Here's a red/blue one, where does it go?" When sorting by size, the children were instructed "Here's a small/big one, where does it go?" The other half of the children were presented with an alternative sorting set, comprising target cards depicting a big green snake and a small yellow horse, and sorting cards showing a small green snake or a big yellow horse (see Appendix Xb). When sorting by colour, these children were instructed, "Here's a yellow/green one, where does it go?" When sorting by size, the children were instructed, "Here's a small/big one, where does it go?" As before, children were asked to sort 6 cards in the pre- training and 6 cards in the post-training. Task scores were based on post- training sorting, with a possible range of 0-6.

Training: First session – Children were presented with an initial sorting task which used the same general procedure followed in the pre-training. However, in the training session, participants were asked to sort the cards four times, twice according to number and twice according to colour. Feedback was given on their performance. If a child sorted incorrectly, the experimenter repeated the current sorting rule, pointed to the relevant dimension on the target and sorting cards and showed the child how to sort correctly. As in the pre-training task, children were asked to sort the first six cards according to one dimension, and the remaining six cards according to the alternative dimension. The sorted cards were then removed and shuffled and children asked to sort the first six cards again according to the initial dimension and the remaining cards according to the alternative dimension.

Half of the children received one sorting set comprising target cards of two red strawberries and one green strawberry and sorting cards of one red strawberry or two green strawberries (see Appendix Xc). When sorting by colour, these children were instructed, "Here's a red/green one, where does it go?" When sorting by number, the children were instructed "Here is one, where does it go?/Here are two, where do they go?" The other half of the children received an alternative sorting set. This set comprised target cards consisting of two blue aeroplanes and one yellow aeroplane and sorting cards consisting of one blue aeroplane or two yellow aeroplanes (see Appendix Xd). When sorting by colour, these children were instructed "Here's a blue/yellow one, where does it go?" When sorting by number, the children were instructed "Here is one, where does it go?/Here are two, where do they go?"

When children had sorted the first set of cards four times, the sorting cards were removed and children were presented with six more sorting cards. The target cards were kept to guide sorting. The six new sorting cards were identical to one another and revealed an image suitable for contrastive sorting along the number and colour dimensions. Children who had received the strawberry cards in the prior task were given cards depicting a single red car, and children who had previously sorted cards with aeroplanes, received cards depicting two yellow apples (see Appendix Xe). Children were asked to sort all six cards according to one dimension and provided with feedback, as in the prior task. If the sorting dimension was colour, and children received the cards depicting a single red car, they would be asked "Here's a red one, where does it go?" These cards would be sorted in the tray with the two red strawberry target cards. When all the cards had been sorted, the cards were removed from the sorting tray and children were asked to sort according to the alternative dimension of number. If the children received the cards depicting two yellow apples, they would be asked "Here are two. Where do they go?" As in all card-sorting tasks, order of dimension sorting was counterbalanced across children.

When children had sorted this set of six cards according to both dimensions, the target and sorting cards were removed. They were then given two new target cards differing on the dimensions of number and colour and a final set of six identical sorting cards. As in the previous sorting task, these cards revealed an image suitable for contrastive sorting along the number and colour dimensions depicted on the target cards. There were two sets of target and sorting cards, and those children who had started the session by sorting the strawberry cards were given target cards comprising two yellow flowers and one red flower, and sorting cards revealing one yellow flower (see Appendix Xf). When sorting by colour, these children were instructed, "Here's a yellow one, where does it go?" When sorting by number, the children were instructed, "Here is one, where does it go?" Children who had started the session by sorting the aeroplane cards received target cards depicting two green trees and one blue tree, and sorting cards revealing two blue trees (see Appendix Xg). When sorting by colour, these children were instructed "Here's a blue one, where does it go?" When sorting by number, the children were instructed "Here are two, where do they go?"

Training: Second session – This was identical to the first session, except that children completed the sorting tasks using the sets of sorting cards they had not used in the previous training session.

Post- training tasks: Post training two box task (Post2BoxDCCS) - This followed the same procedure as the pre-training task. However, children were presented with the sorting set they had not been presented with in the pre-training task. Thus the score range for this task was 0-6.

Post-training tasks: Post training three box task (Post3BoxDCCS)– This followed the same basic procedure as the pre-training /post-training two box task. However, children were presented with three target cards depicting a yellow hat, a red pencil and a green teddy and twenty-one sorting cards revealing a green pencil, a yellow teddy and a red hat (see Appendix Xh). Children were shown how to sort three different sorting cards according to either colour or shape, and then asked to sort nine more cards according to this rule. When sorting by colour, these children were instructed "Here's a green/yellow/red one, where does it go?" When sorting by shape, the children were instructed "Here is a hat/pencil/teddy, where does it go?" When these nine cards had been sorted correctly (no child failed to achieve this), children were asked to sort the remaining nine cards according to the alternative rule (i.e. if they had started by sorting the cards according to colour, then they were asked to sort the last nine cards according to shape). The order of rule sorting was counterbalanced across children. Task scores were based on post-switch performance and so had possible range of 0-9.

The use of a three-box task was in accordance with the suggestion of Kloo and Perner (2003) that enhanced performance on a two-box DCCS task following DCCS training might only indicate that children had learnt to switch sorting boxes following notice of a rule change. Thus an improvement in sorting on the two-box task need not necessarily demonstrate that children have acquired the ability to shift between alternative mental representations of a figure on a sorting card. However, correct performance on a three-box task requires more than simply switching sorting to an alternative box. As accurate sorting following a rule change in the three-box task requires selection of the appropriate sorting box from the two possible alternatives, children are required to shift their conceptualisation of the card figure in accordance with the rule-change.

CVT

Pre- training task (PreCVT): This was based on the CVT task used in Experiment 3. However, Item 21 Appendix VIII was removed from the set of twenty-five items used in Experiment 3, and the remaining twenty-four items divided into two sets of twelve items. Each of the two sets contained either two or three items representing the First and Second Maxim of Quantity, the Maxims of Quality, Relation and Politeness (see Appendix XI: Series A and B). The removal of Item 21 (Appendix VIII), enabled an even split between the number of items used in the two pre-training CVT sets. This item had been used to represent the Maxim of Politeness in Experiment 3. However, as the Politeness Maxim had not, in fact, been included in Grice's original list of conversational maxims, a trial representing the Maxim of Politeness was thus deemed the most appropriate item to remove.

Within each series, items reflecting the different conversational maxims and rules were presented in almost the same quasi-random order in which they had been presented, in Experiment 3. Each child was presented with one of the two series of items (counterbalanced across children). As in Experiment 3, children watched the CVT series on a laptop and were introduced to the three puppets: Mark, Lucy, and

Jane, who were involved in the task. Task scores were based on children's accurate identification of the puppet who gave the inappropriate response in each trial. The task had a possible score range of 0-12.

Post- training task (PostNewCVT and PostOldCVT): Children were given a new pre-training series of CVT items that they had not received during the Pre-training phase (PostNewCVT), followed immediately by the series of pre-training items that had been given previously (PostOldCVT). Thus if they had received Series A in the pre-training phase, they received Series B followed immediately by Series A in the post-training phase. The PostNewCVT and the PostOldCVT each had a possible score range of 0-12. Performance on the post-training Old Item CVT was used to provide a basic indication of CU. Post-training NewCVT scores were used to assess application of CU to new CVT items.

Control Task

Pre-training task (PreCon): Half the children were presented with two parallel, equal length rows of seven yellow counters, and half were presented with two parallel, equal length rows of six green paperclips. Children were told, "Here we have some yellow counters/ green paperclips." The experimenter pointed to the row nearest to the child and asked, "Are there more counters/paperclips in this row?" Then she pointed to the other row and asked, "Or are there more counters/paperclips in this row?" Finally, she pointed to both rows alternately, asking "Or are there the same number in both rows?" When the child had responded, the experimenter pushed the counters/clips further apart in one of the rows to lengthen it, and asked the child "Now, are there more counters/paperclips in this row?" pointing to the row nearest to the child, "Or are there more counters/paperclips in this row?" pointing to the other row, "Or are there the same number in both rows?" pointing to both rows alternately.

When the child had responded, the experimenter then pushed the counters/clips together in one of the rows to shorten it. She then asked the child, "Now, are there more counters/paperclips in this row?" pointing to the row nearest to the child, "Or are there more counters/paperclips in this row?" pointing to the other row, "Or are there the same number in both rows?" pointing to both rows alternately.

Post-training task (PostCon): This followed the same procedure as the pre-training task. However, if children were given the yellow counters in the pre-training task, they were given the green paperclips in the post-training task. Children given the green paperclips in the pre-training task were given the yellow counters in the post-training task.

6.3 Results

Score means and standard deviations on the measures are presented in Table 6.1. Correlations between the tasks are shown in Table 6.2.

Table 6.1: *Mean scores and standard deviations of the pre-training and post-training tasks given to the children in Experiment 4 (N=29)**

	Pre-training	Post-training
Two-Box DCCS/6	0.69 (1.47)	5.24 (1.55)
Three-Box DCCS/9	N/A	6.55 (3.03)
Pre-training CVT/CVT Old/12	6.10 (1.59)	6.93 (1.62)
% scores for pre-training CVT/CVT Old		
First Quantity maxim/100	52.90 (33.13)	55.69 (32.91)
Second Quantity maxim/100	48.79 (27.15)	47.07 (28.25)
Quality maxim/100	50.10 (30.96)	61.48 (28.31)
Relation maxim/100	47.66 (28.49)	59.21 (31.75)
Politeness maxim/100	58.62 (30.09)	67.24 (36.05)
CVT New/12	N/A	6.86 (1.25)
% scores for CVT New First Quantity maxim/100	N/A	56.93 (32.34)
Second Quantity maxim/100	N/A	58.03 (23.12)
Quality maxim/100	N/A	60.34 (32.62)
Relation maxim/100	N/A	53.97 (30.17)
Politeness maxim/100	N/A	58.62 (32.92)
Control Conservation Task/4	1.76 (1.38)	2.21 (1.42)

**Post-training control data is missing for one child who did not seem to understand the questions posed. Note: DCCS=Dimensional Change Card Sort task; CVT=Conversational Violations Task;* CVT Old: Post-training CVT using old items; CVT New: Post-training CVT using new items.

Effect of DCCS training on CVT performance

A 3 (CVT condition: Pre vs. PostNew vs. PostOld) X 5 (Maxim) ANOVA did not reveal a significant difference between performance in the pre- and post-training CVTs, $F(2, 420) = 2.012$, $p>0.10$. There was also no significant difference between performance on CVT items representing different Maxims, $F(4, 420) = 1.388$, $p>0.20$, and no significant interaction between CVT condition and Maxim, $F(8, 420) = 0.547$, $p>0.80$. Table 6.2 reports the means and standard deviations of CVT percentage scores for the different maxims across the CVT conditions.

Table 6.2: *CVT percentage score means and standard deviations for the maxims (first quantity, second quantity, quality, relation and politeness) across the pre-training, post-training new items and post-training old items CVT conditions.*

CVT condition	Maxim	Mean % CVT score	Standard Deviation
Pre-training	First Quantity	52.90	33.13
	Second Quantity	48.79	27.15
	Quality	50.10	30.96
	Relation	47.66	28.49
	Politeness	58.62	30.09
Post-training New Items	First Quantity	56.93	32.34
	Second Quantity	58.03	23.12
	Quality	60.34	32.62
	Relation	53.97	30.17
	Politeness	58.62	32.92
Post-training Old Items	First Quantity	55.69	32.91
	Second Quantity	47.07	28.25
	Quality	61.48	28.31
	Relation	59.21	31.75
	Politeness	67.24	36.05

However, one-sample t-tests determining CVT scores of 6 out of 12 as chance levels, indicated that although scores for the pre-training CVT were not significantly different from chance, $t(28) = 0.351$, $p>0.70$, scores for the post-training NewCVT, which used items not seen in the pre-training, were greater than chance, $t(28) = 3.727$, $p<0.01$ (see Table 6.1 for CVT score means and standard deviations). Scores for the post-training OldCVT, which was identical to the pre-training CVT but was presented after training, were also greater than chance, $t(28) =3.087$, $p<0.01$(see Table 6.1 for CVT score means and standard deviations). Furthermore, one sample t-tests on the individual conversational maxims, revealed that children were only scoring above chance levels: i.e. accurately detecting maxim violations over 50% of the time on groups of items presented post-training. More specifically, analyses found above chance performance on the CVT items representing the Maxim of Quality and the Rule of Politeness in the post-training OldCVT, $t(28) = 2.185$, $p<0.05$ and $t(28) = 2.576$, $p<0.05$ respectively (Table 6.1 presents the means and standard deviations for percentages of maxim violations correctly detected).

Effect of DCCS training on DCCS performance

To assess the effect of DCCS training on DCCS performance, a paired samples t-test was performed to compare performance on the two-box DCCS task before and after DCCS training had been received. This analysis indicated that DCCS training led to a significant improvement in performance on the standard two-box DCCS task, $t(28) = 12.769$ $p<0.001$ (see Table 6.1 for pre- and post training mean scores and standard deviations). One-sample t-tests were then performed on DCCS scores achieved in the two-box task, prior to and after DCCS training. Actual scores were compared to chance-level performance, determined as a score of 3 out of 6 in the post-switch phase. Pre-training performance on the two-box task was significantly worse than chance selection between the two post-switch alternatives, $t(28) = 8.485$, $p<0.001$ (see Table 6.1 for pre-training two-box DCCS score means and standard deviations). However, post-training performance on the two-box task was significantly better than chance selection between the two post-switch alternatives, $t(28) = 7.785$, $p<0.001$(see Table 6.1 for post-training two-box DCCS score means and standard deviations). Further, a one-sample t-test revealed that post-training performance on the three-box task was also significantly better than chance selection between the three post-switch alternatives determined as a score of 3 out of 9 in the post-switch phase, $t(28) = 6.311$, $p<0.001$ (see Table 6.1 for post-training three-box DCCS score means and standard deviations).

Effect of DCCS training on Control task performance

Although children achieved higher scores in the post-training control conservation task than in the pre-training control task, they were not performing significantly better at post-training, $t(27) = 1.694$, $p>0.10$. One-sample t-tests revealed that neither pre- nor post-training performance on the control task was significantly better than chance determined as a score of 2 out of 4, $t(28) = 0.942$, $p>0.30$ and $t(27) = 0.797$, $p>0.40$ respectively.

Table 6.3: *Task score intercorrelations (n = 29 except for the post-training control scores: data is missing for one child who did not seem to understand the questions posed).*

	Age (m)	BPVS	PreDCCS	PreCVT	PreCon	Post2Box DCCS	Post3Box DCCS	PostCVT New	PostCVT Old	PostCon
Age(m)	-									
BPVS	**0.332***	-								
PreDCCS	0.182	**0.324***	-							
PreCVT	-0.226	0.293	**0.321***	-						
PreCon	**0.391***	0.291	-0.021	-0.070	-					
Post2BoxDCCS	0.125	0.155	0.191	-0.097	0.028	-				
Post3BoxDCCS	0.090	0.141	0.249	-0.049	0.084	**0.746*****	-			
PostCVTNew	0.168	0.068	0.269	0.170	-0.020	-0.056	-0.008	-		
PostCVTOld	0.259	**0.414***	**0.381***	0.307	0.215	-0.050	0.015	0.030	-	
PostCon	0.118	0.161	0.100	0.278	**0.613*****	-0.104	0.033	-0.023	-0.054	-

N.B. ***=p<0.001 (1-tailed), *= p < 0.05 (1-tailed), BPVS: British Picture Vocabulary Scale, PreDCCS: Pre-training two box Dimensional Change Card Sort Task, PreCVT: Pre-training CVT items, PreCon: Pre-training control task, Post2BoxDCCS: Post-training two-box Dimensional Change Card Sort task, Post3BDCCS: Post-trainng three-box Dimensional Change Card Sort task, PostCVTNew: Post-trainng CVT using new items, PostCVTOld: Post-training CVT using old/pre-training items, PostCon: Post-training control task.

Relationship between DCCS and CVT

BPVS scores were found to correlate significantly with pre-training DCCS performance (r=0.324, p<0.05) and PostOldCVT scores (r=0.414, p<0.05). However, age in months was not found to demonstrate significant correlations with DCCS or CVT performance. Thus BPVS, but not age, was partialled out of one-way ANCOVAs to determine the effects of DCCS ability on CVT performance. Because of the number of analyses conducted to investigate the relationship between DCCS and CVT scores, a conservative significance level of 0.01 was adopted.

DCCS improvement was calculated by subtracting pre-training 2 box DCCS scores from post-training 2 box DCCS scores. Children's DCCS improvement was classed as strong if their scores increased by 6 from pre- to post-training levels and weak if their scores increased by less than 6. The weak and strong division criteria were selected on the basis that they enabled as equal a split as possible in the number of children assigned to the two DCCS improvement groups. Of the 29 children given DCCS training, 13 demonstrated weak DCCS improvement and 16 demonstrated strong DCCS improvement. Table 6.4 displays the results of ANCOVAs investigating the effects of strong vs. weak DCCS improvement and maxim type on post-training CVT scores and pre to post-training CVT improvement, whilst partialling out BPVS scores. No effect of DCCS improvement was revealed. Table 6.5 reports the estimated marginal means and standard errors of post-training CVT percentage scores for the weak and strong shifters across maxims when the effect of vocabulary score was taken into account and Table 6.6 reports the estimated marginal means and standard errors of pre to post-training CVT percentage improvement for the weak and strong shifters across maxims when the effect of vocabulary score was taken into account.

Tables 6.7 and 6.8 display the results of ANCOVAs investigating the effects of post-training DCCS performance and maxims on post-training CVT scores and pre- to post-training CVT improvement, whilst partialling out BPVS scores. Tables 6.9-6.12 report the estimated marginal means and standard errors of post-training CVT percentage scores and pre- to post-training CVT improvement for the weak and strong shifters across maxims when the effects of vocabulary scores were taken into account. Children's 2 box post-training DCCS performance was classed as strong if their scores were 6 and weak if their scores were less than 6. Again, the weak and strong division criteria were selected on the basis that they enabled as equal a split as possible in the number of children assigned to the two DCCS performance groups. Of 29 children, 7 demonstrated weak 2 box performance and 22 demonstrated strong 2 box performance. Children's 3 box post-training DCCS performance was classed as strong if their scores were 9 and weak if their scores were less than 9, with weak and strong division criteria chosen to enable as equal a split as possible in the number of children assigned to the two DCCS performance groups. Of 29 children, 18 demonstrated weak 3 box performance and 11 demonstrated strong 3 box performance. No effect of post-training DCCS performance was revealed either, although the numbers were unevenly distributed between 2 box DCCS groups so only tentative implications can be drawn from these analyses. Further, the correlation between pre-training DCCS and pretraining CVT scores reported in Table 6.3 was not significant once BPVS scores were partialled out, $\underline{r}=0.263$, $\underline{p}>0.08$.

Table 6.4: *Results for the 2 (2 box DCCS post training improvement level: <6 point increase = weak vs. 6 point increase = strong) x 5 (Maxim type) ANCOVAs partialling out standardised BPVS scores on post-training and improvement CVT measures.*

		df	F	p	partial Eta²
DCCS Improvement:	PostNewCVT	1,134	0.159	0.691	0.001
	PostOldCVT	1,134	3.293	0.072	0.024
	PostCombinedCVT	1,134	2.320	0.130	0.017
	ImproveNewCVT	1,134	0.103	0.748	0.001
	ImproveOldCVT	1,134	0.568	0.452	0.004
Maxim:	PostNewCVT	4,134	0.191	0.943	0.006
	PostOldCVT	4,134	1.609	0.176	0.046
	PostCombinedCVT	4,134	1.241	0.297	0.036
	ImproveNewCVT	4,134	0.211	0.932	0.006
	ImproveOldCVT	4,134	0.664	0.618	0.019
DCCS Improvement x Maxim:	PostNewCVT	4,134	0.574	0.682	0.017
	PostOldCVT	4,134	0.223	0.925	0.007
	PostCombinedCVT	4,134	0.233	0.920	0.007
	ImproveNewCVT	4,134	0.225	0.924	0.007
	ImproveOldCVT	4,134	0.872	0.482	0.025

N.B. PostNewCVT= % scores for maxims in post-training new item CVT measure; PostOldCVT= % scores for maxims in post-training old item CVT measure; PostCombinedCVT= a combined % score measure produced by summing raw performance on each of the post-training CVTs and then converting into overall % scores; ImproveNewCVT=a CVT improvement measure calculated by subtracting % scores for maxims in the pre-training measure from % scores for maxims in the post-training new item CVT; ImproveOldCVT= a CVT improvement measure calculated by subtracting % scores for maxims in the pre-training CVT measure from% scores for maxims the post-training old item CVT.

Table 6.5: *Estimated marginal mean CVT percentage scores and standard errors for the maxims (first quantity, second quantity, quality, relation and politeness) produced across the post-training new items, post-training old items and combined post-training CVT conditions by children demonstrating weak and strong improvement in shifting when individual differences in vocabulary score were taken into account.*

CVT measure	Shifting EF improvement	Maxim	Mean % CVT score	Standard Error
PostNewCVT score	Weak	First Quantity	65.4	08.6
		Second Quantity	56.4	08.6
		Quality	60.3	08.6
		Relation	49.9	08.6
		Politeness	61.5	08.6
	Strong	First Quantity	50.1	07.7
		Second Quantity	59.4	07.7
		Quality	60.4	07.7
		Relation	57.2	07.7
		Politeness	56.2	07.7
PostOldCVT	Weak	First Quantity	61.1	08.8
		Second Quantity	52.1	08.8
		Quality	62.5	08.8
		Relation	68.8	08.8
		Politeness	72.7	08.8
	Strong	First Quantity	51.3	07.9
		Second Quantity	43.0	07.9
		Quality	60.7	07.9
		Relation	51.4	07.9
		Politeness	62.8	07.9
PostCombinedCVT	Weak	First Quantity	61.3	06.0
		Second Quantity	53.6	06.0
		Quality	61.3	06.0
		Relation	58.2	06.0
		Politeness	67.1	06.0
	Strong	First Quantity	50.2	05.4
		Second Quantity	50.2	05.4
		Quality	60.2	05.4
		Relation	53.9	05.4
		Politeness	59.5	05.4

Table 6.6: *Estimated marginal mean CVT percentage improvements and standard errors for the maxims (first quantity, second quantity, quality, relation and politeness) calculated using the post-training new items CVT measure and calculated using the post-training old items CVT measure for children demonstrating weak and strong shifting improvement when individual differences in vocabulary score were taken into account.*

CVT measure	Shifting EF Improvement	Maxim	Mean % CVT score	Standard Error
ImproveNewCVT	Weak	First Quantity	06.9	12.4
		Second Quantity	03.1	12.4
		Quality	04.3	12.4
		Relation	08.3	12.4
		Politeness	00.5	12.4
	Strong	First Quantity	01.7	11.2
		Second Quantity	14.2	11.2
		Quality	15.1	11.2
		Relation	04.7	11.2
		Politeness	-00.4	11.2
ImproveOldCVT	Weak	First Quantity	02.5	11.1
		Second Quantity	-01.2	11.1
		Quality	06.5	11.1
		Relation	27.0	11.1
		Politeness	11.7	11.1
	Strong	First Quantity	02.9	10.0
		Second Quantity	-02.2	10.0
		Quality	15.5	10.0
		Relation	-01.2	10.0
		Politeness	06.1	10.0

Table 6.7: *Results for the 2 (post-training DCCS level: 2 box scores: <6 = weak vs. = 6 = strong) x 5 (Maxim type) ANCOVAs partialling out standardised BPVS scores on post-training and improvement CVT measures.*

		df	F	p	partial Eta²
2 box DCCS:	PostNewCVT	1,134	0.532	0.467	0.004
	PostOldCVT	1,134	0.225	0.636	0.002
	PostCombinedCVT	1,134	0.149	0.700	0.001
	ImproveNewCVT	1,134	0.962	0.328	0.007
	ImproveOldCVT	1,134	0.025	0.875	0.000
Maxim:	PostNewCVT	4,134	0.584	0.675	0.017
	PostOldCVT	4,134	0.639	0.636	0.019
	PostCombinedCVT	4,134	0.751	0.559	0.022
	ImproveNewCVT	4,134	0.278	0.892	0.008
	ImproveOldCVT	4,134	0.856	0.492	0.025
2 box DCCS x Maxim:	PostNewCVT	4,134	0.791	0.533	0.023
	PostOldCVT	4,134	1.156	0.333	0.033
	PostCombinedCVT	4,134	1.192	0.317	0.034
	ImproveNewCVT	4,134	0.066	0.992	0.002
	ImproveOldCVT	4,134	1.373	0.247	0.039

Table 6.8: *Results for the 2 (post-training DCCS level: 3 box scores: <9 = weak vs. =9 = strong) x 5 (Maxim type) ANCOVAs partialling out standardised BPVS scores on post-training and improvement CVT measures.*

		df	F	p	partial Eta²
3 box DCCS:	PostNewCVT	1,134	0.733	0.393	0.005
	PostOldCVT	1,134	0.214	0.644	0.002
	PostCombinedCVT	1,134	0.566	0.453	0.004
	ImproveNewCVT	1,134	0.508	0.477	0.004
	ImproveOldCVT	1,134	0.239	0.626	0.002
Maxim:	PostNewCVT	4,134	0.037	0.997	0.001
	PostOldCVT	4,134	1.391	0.240	0.040
	PostCombinedCVT	4,134	1.020	0.399	0.030
	ImproveNewCVT	4,134	0.272	0.895	0.008
	ImproveOldCVT	4,134	0.966	0.428	0.028
3 box DCCS x Maxim:	PostNewCVT	4,134	1.978	0.101	0.056
	PostOldCVT	4,134	0.547	0.701	0.016
	PostCombinedCVT	4,134	2.018	0.096	0.057
	ImproveNewCVT	4,134	3.106	0.018	0.085
	ImproveOldCVT	4,134	1.948	0.106	0.055

Table 6.9: *Estimated marginal mean CVT percentage scores and standard errors for the maxims (first quantity, second quantity, quality, relation and politeness) produced by weak and strong post-training 2 box shifters across the post-training new items, post-training old items and combined post-training CVT conditions when individual differences in vocabulary score were taken into account.*

CVT measure	Shifting EF level	Maxim	Mean % CVT score	Standard Error
PostNewCVT score	Weak	First Quantity	59.6	11.6
		Second Quantity	57.1	11.6
		Quality	66.7	11.6
		Relation	38.0	11.6
		Politeness	50.0	11.6
	Strong	First Quantity	56.1	06.5
		Second Quantity	58.3	06.5
		Quality	58.3	06.5
		Relation	59.0	06.5
		Politeness	61.4	06.5
PostOldCVT	Weak	First Quantity	47.9	11.9
		Second Quantity	64.5	11.9
		Quality	67.1	11.9
		Relation	64.7	11.9
		Politeness	57.5	11.9
	Strong	First Quantity	58.2	06.7
		Second Quantity	41.5	06.7
		Quality	59.7	06.7
		Relation	57.5	06.7
		Politeness	70.3	06.7
PostCombinedCVT	Weak	First Quantity	51.6	08.1
		Second Quantity	60.2	08.1
		Quality	65.9	08.1
		Relation	48.8	08.1
		Politeness	53.8	08.1
	Strong	First Quantity	56.3	04.6
		Second Quantity	49.0	04.6
		Quality	59.0	04.6
		Relation	58.1	04.6
		Politeness	65.8	04.6

Table 6.10: *Estimated marginal mean CVT percentage improvements and standard errors for the maxims (first quantity, second quantity, quality, relation and politeness) calculated using the post-training new items CVT measure and calculated using the post-training old items CVT measure for children demonstrating weak and strong post-training 2 box shifting when individual differences in vocabulary score were taken into account.*

CVT measure	Shifting EF Level	Maxim	Mean % CVT score	Standard Error
ImproveNewCVT	Weak	First Quantity	-07.6	16.9
		Second Quantity	04.2	16.9
		Quality	09.0	16.9
		Relation	-00.5	16.9
		Politeness	-07.6	09.5
	Strong	First Quantity	07.7	09.5
		Second Quantity	10.8	09.5
		Quality	10.6	09.5
		Relation	08.5	09.5
		Politeness	02.4	09.5
ImproveOldCVT	Weak	First Quantity	-19.1	15.0
		Second Quantity	11.8	15.0
		Quality	09.5	15.0
		Relation	25.9	15.0
		Politeness	-00.1	15.0
	Strong	First Quantity	09.7	08.5
		Second Quantity	-06.0	08.5
		Quality	12.1	08.5
		Relation	06.8	08.5
		Politeness	11.4	08.5

Table 6.11: *Estimated marginal mean CVT percentage scores and standard errors for the maxims (first quantity, second quantity, quality, relation and politeness) produced by weak and strong post-training 3 box shifters across the post-training new items, post-training old items and combined post-training CVT conditions when individual differences in vocabulary score were taken into account.*

CVT measure	Shifting EF level	Maxim	Mean % CVT score	Standard Error
PostNewCVT score	Weak	First Quantity	62.1	07.1
		Second Quantity	60.2	07.1
		Quality	67.6	07.1
		Relation	45.3	07.1
		Politeness	61.1	07.1
	Strong	First Quantity	48.4	09.1
		Second Quantity	54.5	09.1
		Quality	48.4	09.1
		Relation	68.2	09.1
		Politeness	54.5	09.1
PostOldCVT	Weak	First Quantity	58.1	07.5
		Second Quantity	46.1	07.5
		Quality	66.5	07.5
		Relation	55.4	07.5
		Politeness	69.3	07.5
	Strong	First Quantity	51.7	09.6
		Second Quantity	48.7	09.6
		Quality	53.2	09.6
		Relation	65.4	09.6
		Politeness	63.9	09.6
PostCombinedCVT	Weak	First Quantity	57.7	05.0
		Second Quantity	52.1	05.0
		Quality	66.6	05.0
		Relation	49.9	05.0
		Politeness	65.2	05.0
	Strong	First Quantity	51.0	06.4
		Second Quantity	51.0	06.4
		Quality	51.0	06.4
		Relation	65.6	06.4
		Politeness	59.2	06.4

Table 6.12: *Estimated marginal mean CVT percentage improvements and standard errors for the maxims (first quantity, second quantity, quality, relation and politeness) calculated using the post-training new items CVT measure and calculated using the post-training old items CVT measure for children demonstrating weak and strong post-training 3 box shifting when individual differences in vocabulary score were taken into account.*

CVT measure	Shifting EF Level	Maxim	Mean % CVT score	Standard Error
ImproveNewCVT	Weak	First Quantity	16.0	10.1
		Second Quantity	16.9	10.1
		Quality	19.6	10.1
		Relation	-07.3	10.1
		Politeness	-05.4	10.1
	Strong	First Quantity	-15.5	12.9
		Second Quantity	-03.3	12.9
		Quality	-05.0	12.9
		Relation	28.6	12.9
		Politeness	08.8	12.9
ImproveOldCVT	Weak	First Quantity	11.9	09.3
		Second Quantity	02.8	09.3
		Quality	18.5	09.3
		Relation	02.8	09.3
		Politeness	02.8	09.3
	Strong	First Quantity	-12.2	11.9
		Second Quantity	-09.2	11.9
		Quality	00.1	11.9
		Relation	25.7	11.9
		Politeness	18.1	11.9

Since the 3- to 4-year-old group in the current study was younger than the 4-year old group who had demonstrated the DCCS-CVT relationship in Experiment 3, the data for the current study was split into separate 3-year-old and 4-year-old age groups. The previous two set of analyses focusing on pre-to-post training DCCS improvement and post-training DCCS scores were re-run on the two separate age groups (see Tables 6.13-6.18). Tables 6.19-6.30 report the estimated marginal means and standard deviations for these analyses. As for the previous analyses, children's DCCS improvement was classed as strong if their scores increased by 6 from pre- to post-training levels and weak if their scores increased by less than 6. As before, weak and strong division criteria were selected on the basis that they enabled as equal a split as possible in the number of children assigned to the two DCCS improvement groups. Of the 18, 3-year-olds given DCCS training, 7 demonstrated weak DCCS improvement and 11 demonstrated strong DCCS improvement. Of the 11, 4-year-olds given DCCS training, 6 demonstrated weak DCCS improvement and 5 demonstrated strong DCCS improvement. Children's 2 box post-training DCCS performance was classed as strong if their scores were 6 and weak if their scores were less than 6. Again, weak and strong division criteria were selected on the basis that they enabled as equal a split as possible in the number of children assigned to the two DCCS performance groups. Of 18, 3-year-olds, 5 demonstrated weak post-training 2 box performance and 13 demonstrated strong post-training 2 box performance. Of 11, 4-year-olds, 2 demonstrated weak post-training 2 box performance and 9 demonstrated strong post-training 2 box performance. Children's 3 box post-training DCCS performance was classed as strong if their scores were 9 and weak if their scores were less than 9, with weak and strong division criteria chosen to enable as equal a split as possible in the number of children assigned to the two DCCS performance groups. Of 18, 3-year-olds, 10 demonstrated weak 3 box performance and 8 demonstrated strong 3 box performance. Of 11, 3-year-olds, 8 demonstrated weak 3 box performance and 3 demonstrated strong 3 box performance.

As in the previous analyses, there was no effect of pre-to-post-training DCCS improvement or post-training DCCS scores on post-training CVT scores or pre- to post-training CVT improvement. However, reflecting the pattern present in the combined age analysis, N was unevenly distributed between post-training 2 box DCCS score level groups formed for both the 3- and 4-year old age groups. Furthermore, N was found to be disproportionately spread between the post-training 3 box DCCS score groups formed for the 4-year old age group. Thus implications should be drawn from these analyses cautiously. There was no correlation between pre-training DCCS and pretraining CVT scores, even before BPVS scores had been partialled out, $r=0.264$, $p>0.140$ for the 3-year-olds and $r=0.447$, $p>0.08$ for the 4-year-olds.

Table 6.13: *Results for the 3-year-olds of 2 (2 box DCCS post training improvement level: <6 point increase = weak vs. 6 point increase = strong) x 5 (Maxim type) ANCOVAs partialling out standardised BPVS scores on post-training and improvement CVT measures.*

		df	F	p	partial Eta²
DCCS Improvement:	PostNewCVT	1,79	0.082	0.776	0.001
	PostOldCVT	1,79	1.151	0.287	0.014
	PostCombinedCVT	1,79	1.283	0.261	0.016
	ImproveNewCVT	1,79	0.525	0.471	0.007
	ImproveOldCVT	1,79	0.013	0.909	0.000
Maxim:	PostNewCVT	4,79	1.199	0.318	0.057
	PostOldCVT	4,79	0.624	0.647	0.031
	PostCombinedCVT	4,79	1.257	0.294	0.060
	ImproveNewCVT	4,79	0.129	0.971	0.007
	ImproveOldCVT	4,79	0.792	0.534	0.039
DCCS Improvement x Maxim:	PostNewCVT	4,79	2.323	0.064	0.105
	PostOldCVT	4,79	0.080	0.988	0.004
	PostCombinedCVT	4,79	0.682	0.606	0.033
	ImproveNewCVT	4,79	1.811	0.135	0.084
	ImproveOldCVT	4,79	0.597	0.666	0.029

Table 6.14: *Results for the 3-year-olds of 2 (post-training DCCS level: 2 box scores: <6 = weak vs. = 6 = strong) x 5 (Maxim type) ANCOVAs partialling out standardised BPVS scores on post-training and improvement CVT measures.*

		df	F	p	partial Eta²
2 box DCCS:	PostNewCVT	1,79	0.110	0.741	0.001
	PostOldCVT	1,79	0.946	0.334	0.012
	PostCombinedCVT	1,79	0.226	0.635	0.003
	ImproveNewCVT	1,79	1.119	0.293	0.014
	ImproveOldCVT	1,79	0.013	0.910	0.000
Maxim:	PostNewCVT	4,79	1.568	0.191	0.074
	PostOldCVT	4,79	0.257	0.905	0.013
	PostCombinedCVT	4,79	0.847	0.499	0.041
	ImproveNewCVT	4,79	0.223	0.925	0.011
	ImproveOldCVT	4,79	0.758	0.556	0.037
2 box DCCS x Maxim:	PostNewCVT	4,79	1.321	0.269	0.063
	PostOldCVT	4,79	0.680	0.608	0.033
	PostCombinedCVT	4,79	0.786	0.538	0.038
	ImproveNewCVT	4,79	0.491	0.742	0.024
	ImproveOldCVT	4,79	0.545	0.703	0.027

Table 6.15: *Results for the 3-year-olds of 2 (post-training DCCS level: 3 box scores: <9 = weak vs. = 9 = strong) x 5 (Maxim type) ANCOVAs partialling out standardised BPVS scores on post-training and improvement CVT measures.*

		df	F	p	partial Eta²
3 box DCCS:	PostNewCVT	1,79	1.390	0.242	0.017
	PostOldCVT	1,79	0.452	0.504	0.006
	PostCombinedCVT	1,79	1.644	0.204	0.020
	ImproveNewCVT	1,79	0.040	0.842	0.001
	ImproveOldCVT	1,79	0.011	0.918	0.000
Maxim:	PostNewCVT	4,79	0.537	0.709	0.026
	PostOldCVT	4,79	0.668	0.616	0.033
	PostCombinedCVT	4,79	0.890	0.474	0.043
	ImproveNewCVT	4,79	0.464	0.762	0.023
	ImproveOldCVT	4,79	0.903	0.466	0.044
3 box DCCS x Maxim:	PostNewCVT	4,79	3.060	0.021	0.134
	PostOldCVT	4,79	0.615	0.653	0.030
	PostCombinedCVT	4,79	2.883	0.028	0.127
	ImproveNewCVT	4,79	2.520	0.048	0.113
	ImproveOldCVT	4,79	0.903	0.466	0.044

Table 6.16: *Results for the 4-year-olds of 2 (2 box DCCS post training improvement level: <6 point increase = weak vs. 6 point increase = strong) x 5 (Maxim type) ANCOVAs partialling out standardised BPVS scores on post-training and improvement CVT measures.*

		df	F	p	partial Eta²
DCCS Improvement:	PostNewCVT	1,44	0.120	0.730	0.003
	PostOldCVT	1,44	0.942	0.337	0.021
	PostCombinedCVT	1,44	0.357	0.553	0.008
	ImproveNewCVT	1,44	0.289	0.593	0.007
	ImproveOldCVT	1,44	1.328	0.255	0.029
Maxim:	PostNewCVT	4,44	0.305	0.873	0.027
	PostOldCVT	4,44	2.074	0.100	0.159
	PostCombinedCVT	4,44	1.270	0.296	0.103
	ImproveNewCVT	4,44	0.572	0.684	0.049
	ImproveOldCVT	4,44	0.853	0.500	0.072
DCCS Improvement x Maxim:	PostNewCVT	4,44	0.418	0.795	0.037
	PostOldCVT	4,44	1.049	0.393	0.087
	PostCombinedCVT	4,44	0.269	0.896	0.024
	ImproveNewCVT	4,44	1.894	0.128	0.147
	ImproveOldCVT	4,44	1.173	0.336	0.096

Table 6.17: *Results for the 4-year-olds of 2 (post-training DCCS level: 2 box scores: <6 = weak vs. = 6 = strong) x 5 (Maxim type) ANCOVAs partialling out standardised BPVS scores on post-training and improvement CVT measures.*

		df	F	p	partial Eta²
2 box DCCS:	PostNewCVT	1,44	0.581	0.450	0.013
	PostOldCVT	1,44	0.043	0.837	0.001
	PostCombinedCVT	1,44	1.093	0.302	0.024
	ImproveNewCVT	1,44	0.045	0.833	0.001
	ImproveOldCVT	1,44	0.062	0.804	0.001
Maxim:	PostNewCVT	4,44	0.505	0.733	0.044
	PostOldCVT	4,44	0.990	0.423	0.083
	PostCombinedCVT	4,44	1.287	0.290	0.105
	ImproveNewCVT	4,44	0.320	0.863	0.028
	ImproveOldCVT	4,44	0.895	0.475	0.075
2 box DCCS x Maxim:	PostNewCVT	4,44	0.332	0.855	0.029
	PostOldCVT	4,44	0.629	0.644	0.054
	PostCombinedCVT	4,44	0.654	0.627	0.056
	ImproveNewCVT	4,44	1.214	0.319	0.099
	ImproveOldCVT	4,44	1.300	0.285	0.106

Table 6.18: *Results for the 4-year-olds of 2 (post-training DCCS level: 3 box scores: <9 = weak vs. = 9 = strong) x 5 (Maxim type) ANCOVAs partialling out standardised BPVS scores on post-training and improvement CVT measures.*

		df	F	p	partial Eta²
3 box DCCS:	PostNewCVT	1,44	0.133	0.718	0.003
	PostOldCVT	1,44	0.037	0.848	0.001
	PostCombinedCVT	1,44	0.371	0.545	0.008
	ImproveNewCVT	1,44	0.000	0.991	0.000
	ImproveOldCVT	1,44	0.033	0.856	0.001
Maxim:	PostNewCVT	4,44	0.407	0.803	0.036
	PostOldCVT	4,44	1.336	0.272	0.108
	PostCombinedCVT	4,44	1.172	0.336	0.096
	ImproveNewCVT	4,44	0.189	0.943	0.017
	ImproveOldCVT	4,44	0.644	0.634	0.055
3 box DCCS x Maxim:	PostNewCVT	4,44	0.252	0.907	0.022
	PostOldCVT	4,44	0.491	0.742	0.043
	PostCombinedCVT	4,44	0.581	0.678	0.050
	ImproveNewCVT	4,44	0.725	0.580	0.062
	ImproveOldCVT	4,44	0.896	0.475	0.075

Table 6.19: *Estimated marginal mean CVT percentage scores and standard errors for the maxims (first quantity, second quantity, quality, relation and politeness) produced across the post-training new items, post-training old items and combined post-training CVT conditions by three-year-olds demonstrating weak and strong improvement in shifting when individual differences in vocabulary score were taken into account.*

CVT measure	Shifting EF improvement	Maxim	Mean % CVT score	Standard Error
PostNewCVT score	Weak	First Quantity	81.0	11.1
		Second Quantity	54.8	11.1
		Quality	57.2	11.1
		Relation	33.3	11.1
		Politeness	64.3	11.1
	Strong	First Quantity	47.0	08.9
		Second Quantity	56.0	08.9
		Quality	68.2	08.9
		Relation	56.0	08.9
		Politeness	54.5	08.9
PostOldCVT	Weak	First Quantity	59.6	12.2
		Second Quantity	52.4	12.2
		Quality	57.2	12.2
		Relation	59.6	12.2
		Politeness	64.4	12.2
	Strong	First Quantity	54.4	09.8
		Second Quantity	42.3	09.8
		Quality	46.9	09.8
		Relation	48.4	09.8
		Politeness	63.6	09.8
PostCombinedCVT	Weak	First Quantity	68.6	08.0
		Second Quantity	54.3	08.0
		Quality	57.2	08.0
		Relation	45.8	08.0
		Politeness	64.3	08.0
	Strong	First Quantity	50.9	06.4
		Second Quantity	47.2	06.4
		Quality	56.3	06.4
		Relation	50.9	06.4
		Politeness	59.1	06.4

Table 6.20: *Estimated marginal mean CVT percentage improvements and standard errors for the maxims (first quantity, second quantity, quality, relation and politeness) calculated using the post-training new items CVT measure and calculated using the post-training old items CVT measure for three-year-olds demonstrating weak and strong shifting improvement when individual differences in vocabulary score were taken into account.*

CVT measure	Shifting EF Improvement	Maxim	Mean % CVT score	Standard Error
ImproveNewCVT	Weak	First Quantity	11.8	15.9
		Second Quantity	02.2	15.9
		Quality	-14.5	15.9
		Relation	-04.8	15.9
		Politeness	07.1	15.9
	Strong	First Quantity	-15.2	12.7
		Second Quantity	06.1	12.7
		Quality	31.8	12.7
		Relation	12.1	12.7
		Politeness	00.0	12.7
ImproveOldCVT	Weak	First Quantity	-10.0	15.7
		Second Quantity	00.0	15.7
		Quality	-14.3	15.7
		Relation	21.3	15.7
		Politeness	07.1	15.7
	Strong	First Quantity	-07.6	12.5
		Second Quantity	-07.5	12.5
		Quality	10.7	12.5
		Relation	04.6	12.5
		Politeness	09.1	12.5

Table 6.21: *Estimated marginal mean CVT percentage scores and standard errors for the maxims (first quantity, second quantity, quality, relation and politeness) produced by three-year-old weak and strong post-training 2 box shifters across the post-training new items, post-training old items and combined post-training CVT conditions when individual differences in vocabulary score were taken into account.*

CVT measure	Shifting EF level	Maxim	Mean % CVT score	Standard Error
PostNewCVT score	Weak	First Quantity	73.4	13.4
		Second Quantity	56.6	13.4
		Quality	70.0	13.4
		Relation	26.6	13.4
		Politeness	50.0	13.4
	Strong	First Quantity	55.2	08.3
		Second Quantity	55.1	08.3
		Quality	61.5	08.3
		Relation	55.1	08.3
		Politeness	61.5	08.3
PostOldCVT	Weak	First Quantity	50.0	14.3
		Second Quantity	63.2	14.3
		Quality	60.0	14.3
		Relation	63.4	14.3
		Politeness	60.0	14.3
	Strong	First Quantity	58.9	08.9
		Second Quantity	39.7	08.9
		Quality	47.4	08.9
		Relation	48.7	08.9
		Politeness	65.4	08.9
PostCombinedCVT	Weak	First Quantity	60.0	09.5
		Second Quantity	60.0	09.5
		Quality	64.0	09.5
		Relation	44.0	09.5
		Politeness	55.0	09.5
	Strong	First Quantity	56.9	05.9
		Second Quantity	46.2	05.9
		Quality	53.8	05.9
		Relation	50.8	05.9
		Politeness	63.5	05.9

Table 6.22: *Estimated marginal mean CVT percentage improvements and standard errors for the maxims (first quantity, second quantity, quality, relation and politeness) calculated using the post-training new items CVT measure and calculated using the post-training old items CVT measure for three-year-olds demonstrating weak and strong post-training 2 box shifting when individual differences in vocabulary score were taken into account.*

CVT measure	Shifting EF Level	Maxim	Mean % CVT score	Standard Error
ImproveNewCVT	Weak	First Quantity	03.4	19.3
		Second Quantity	03.2	19.3
		Quality	03.2	19.3
		Relation	-16.6	19.3
		Politeness	-10.0	19.3
	Strong	First Quantity	-07.8	12.0
		Second Quantity	05.2	12.0
		Quality	17.8	12.0
		Relation	14.1	12.0
		Politeness	07.7	12.0
ImproveOldCVT	Weak	First Quantity	-20.0	18.6
		Second Quantity	10.0	18.6
		Quality	-06.6	18.6
		Relation	19.8	18.6
		Politeness	00.0	18.6
	Strong	First Quantity	-04.2	11.5
		Second Quantity	-10.2	11.5
		Quality	03.8	11.5
		Relation	07.7	11.5
		Politeness	11.5	11.5

Table 6.23: *Estimated marginal mean CVT percentage scores and standard errors for the maxims (first quantity, second quantity, quality, relation and politeness) produced by three-year-old weak and strong post-training 3 box shifters across the post-training new items, post-training old items and combined post-training CVT conditions when individual differences in vocabulary score were taken into account.*

CVT measure	Shifting EF level	Maxim	Mean % CVT score	Standard Error
PostNewCVT score	Weak	First Quantity	70.1	09.1
		Second Quantity	56.6	09.1
		Quality	76.7	09.1
		Relation	33.2	09.1
		Politeness	65.0	09.1
	Strong	First Quantity	47.9	10.2
		Second Quantity	54.1	10.2
		Quality	47.9	10.2
		Relation	64.6	10.2
		Politeness	50.0	10.2
PostOldCVT	Weak	First Quantity	64.8	10.2
		Second Quantity	44.7	10.2
		Quality	53.1	10.2
		Relation	48.1	10.2
		Politeness	69.8	10.2
	Strong	First Quantity	45.9	11.4
		Second Quantity	48.1	11.4
		Quality	48.1	11.4
		Relation	58.6	11.4
		Politeness	56.4	11.4
PostCombinedCVT	Weak	First Quantity	65.9	06.4
		Second Quantity	49.9	06.4
		Quality	63.9	06.4
		Relation	39.9	06.4
		Politeness	67.4	06.4
	Strong	First Quantity	47.6	07.1
		Second Quantity	50.1	07.1
		Quality	47.6	07.1
		Relation	60.1	07.1
		Politeness	53.3	07.1

Table 6.24: *Estimated marginal mean CVT percentage improvements and standard errors for the maxims (first quantity, second quantity, quality, relation and politeness) calculated using the post-training new items CVT measure and calculated using the post-training old items CVT measure for three-year-olds demonstrating weak and strong post-training 3 box shifting when individual differences in vocabulary score were taken into account.*

CVT measure	Shifting EF Level	Maxim	Mean % CVT score	Standard Error
ImproveNewCVT	Weak	First Quantity	06.9	13.2
		Second Quantity	08.5	13.2
		Quality	28.5	13.2
		Relation	-13.2	13.2
		Politeness	-04.8	13.2
	Strong	First Quantity	-19.2	14.7
		Second Quantity	-00.3	14.7
		Quality	-04.7	14.7
		Relation	29.0	14.7
		Politeness	12.2	14.7
ImproveOldCVT	Weak	First Quantity	01.4	13.1
		Second Quantity	-03.2	13.1
		Quality	05.1	13.1
		Relation	01.7	13.1
		Politeness	00.1	13.1
	Strong	First Quantity	-21.0	14.6
		Second Quantity	-06.4	14.6
		Quality	-04.2	14.6
		Relation	22.8	14.6
		Politeness	18.6	14.6

Table 6.25: *Estimated marginal mean CVT percentage scores and standard errors for the maxims (first quantity, second quantity, quality, relation and politeness) produced across the post-training new items, post-training old items and combined post-training CVT conditions by four-year-olds demonstrating weak and strong improvement in shifting. when individual differences in vocabulary score were taken into account.*

CVT measure	Shifting EF improvement	Maxim	Mean % CVT score	Standard Error
PostNewCVT score	Weak	First Quantity	47.7	13.5
		Second Quantity	58.9	13.5
		Quality	64.5	13.5
		Relation	69.9	13.5
		Politeness	58.9	13.5
	Strong	First Quantity	56.2	14.8
		Second Quantity	66.2	14.8
		Quality	42.6	14.8
		Relation	59.4	14.8
		Politeness	59.4	14.8
PostOldCVT	Weak	First Quantity	62.5	12.4
		Second Quantity	51.3	12.4
		Quality	68.1	12.4
		Relation	79.1	12.4
		Politeness	82.0	12.4
	Strong	First Quantity	44.9	13.6
		Second Quantity	45.1	13.6
		Quality	91.7	13.6
		Relation	58.5	13.6
		Politeness	61.7	13.6
PostCombinedCVT	Weak	First Quantity	52.7	09.3
		Second Quantity	52.7	09.3
		Quality	66.1	09.3
		Relation	72.7	09.3
		Politeness	70.2	09.3
	Strong	First Quantity	48.7	10.2
		Second Quantity	56.7	10.2
		Quality	68.7	10.2
		Relation	60.7	10.2
		Politeness	60.7	10.2

Table 6.26: *Estimated marginal mean CVT percentage improvements and standard errors for the maxims (first quantity, second quantity, quality, relation and politeness) calculated using the post-training new items CVT measure and calculated using the post-training old items CVT measure for four-year-olds demonstrating weak and strong shifting improvement when individual differences in vocabulary score were taken into account.*

CVT measure	Shifting EF Improvement	Maxim	Mean % CVT score	Standard Error
ImproveNewCVT	Weak	First Quantity	03.3	18.6
		Second Quantity	06.1	18.6
		Quality	28.3	18.6
		Relation	25.6	18.6
		Politeness	-05.0	18.6
	Strong	First Quantity	36.3	20.4
		Second Quantity	29.7	20.4
		Quality	-24.1	20.4
		Relation	-13.9	20.4
		Politeness	-03.9	20.4
ImproveOldCVT	Weak	First Quantity	18.1	15.1
		Second Quantity	-01.4	15.1
		Quality	31.9	15.1
		Relation	34.7	15.1
		Politeness	18.1	15.1
	Strong	First Quantity	24.9	16.6
		Second Quantity	08.3	16.6
		Quality	24.9	16.6
		Relation	-15.1	16.6
		Politeness	-01.7	16.6

Table 6.27: *Estimated marginal mean CVT percentage scores and standard errors for the maxims (first quantity, second quantity, quality, relation and politeness) produced by four-year-old weak and strong post-training 2 box shifters across the post-training new items, post-training old items and combined post-training CVT conditions when individual differences in vocabulary score were taken into account.*

CVT measure	Shifting EF level	Maxim	Mean % CVT score	Standard Error
PostNewCVT score	Weak	First Quantity	24.6	23.3
		Second Quantity	58.1	23.3
		Quality	58.1	23.3
		Relation	66.1	23.3
		Politeness	49.6	23.3
	Strong	First Quantity	57.5	11.0
		Second Quantity	63.1	11.0
		Quality	53.8	11.0
		Relation	64.9	11.0
		Politeness	61.2	11.0
PostOldCVT	Weak	First Quantity	42.9	21.9
		Second Quantity	67.9	21.9
		Quality	84.9	21.9
		Relation	67.9	21.9
		Politeness	51.4	21.9
	Strong	First Quantity	57.0	10.3
		Second Quantity	44.1	10.3
		Quality	77.5	10.3
		Relation	70.1	10.3
		Politeness	77.5	10.3
PostCombinedCVT	Weak	First Quantity	30.5	15.6
		Second Quantity	60.5	15.6
		Quality	70.5	15.6
		Relation	60.5	15.6
		Politeness	50.5	15.6
	Strong	First Quantity	55.4	07.4
		Second Quantity	53.2	07.4
		Quality	66.5	07.4
		Relation	68.8	07.4
		Politeness	69.3	07.4

Table 6.28: *Estimated marginal mean CVT percentage improvements and standard errors for the maxims (first quantity, second quantity, quality, relation and politeness) calculated using the post-training new items CVT measure and calculated using the post-training old items CVT measure for four-year-olds demonstrating weak and strong post-training 2 box shifting when individual differences in vocabulary score were taken into account.*

CVT measure	Shifting EF Level	Maxim	Mean % CVT score	Standard Error
ImproveNewCVT	Weak	First Quantity	-36.1	33.1
		Second Quantity	05.9	33.1
		Quality	22.4	33.1
		Relation	38.9	33.1
		Politeness	-02.6	33.1
	Strong	First Quantity	30.4	15.6
		Second Quantity	19.2	15.6
		Quality	00.5	15.6
		Relation	00.7	15.6
		Politeness	-05.0	15.6
ImproveOldCVT	Weak	First Quantity	-17.3	26.3
		Second Quantity	15.7	26.3
		Quality	49.2	26.3
		Relation	40.7	26.3
		Politeness	-00.8	26.3
	Strong	First Quantity	29.7	12,4
		Second Quantity	00.2	12.4
		Quality	24.2	12.4
		Relation	05.7	12.4
		Politeness	11.3	12.4

Table 6.29: *Estimated marginal mean CVT percentage scores and standard errors for the maxims (first quantity, second quantity, quality, relation and politeness) produced by four-year-old weak and strong post-training 3 box shifters across the post-training new items, post-training old items and combined post-training CVT conditions when individual differences in vocabulary score were taken into account.*

CVT measure	Shifting EF level	Maxim	Mean % CVT score	Standard Error
PostNewCVT score	Weak	First Quantity	51.6	11.8
		Second Quantity	64.1	11.8
		Quality	55.7	11.8
		Relation	59.9	11.8
		Politeness	55.7	11.8
	Strong	First Quantity	51.3	19.4
		Second Quantity	57.0	19.4
		Quality	51.3	19.4
		Relation	79.0	19.4
		Politeness	68.0	19.4
PostOldCVT	Weak	First Quantity	53.1	11.1
		Second Quantity	47.3	11.1
		Quality	84.8	11.1
		Relation	66.0	11.1
		Politeness	70.2	11.1
	Strong	First Quantity	62.9	18.2
		Second Quantity	46.3	18.2
		Quality	62.9	18.2
		Relation	79.6	18.2
		Politeness	79.6	18.2
PostCombinedCVT	Weak	First Quantity	48.0	07.9
		Second Quantity	55.5	07.9
		Quality	70.5	07.9
		Relation	63.0	07.9
		Politeness	63.0	07.9
	Strong	First Quantity	58.7	13.0
		Second Quantity	52.1	13.0
		Quality	58.7	13.0
		Relation	78.7	13.0
		Politeness	73.7	13.0

Table 6.30: *Estimated marginal mean CVT percentage improvements and standard errors for the maxims (first quantity, second quantity, quality, relation and politeness) calculated using the post-training new items CVT measure and calculated using the post-training old items CVT measure for four-year-olds demonstrating weak and strong post-training 3 box shifting when individual differences in vocabulary score were taken into account.*

CVT measure	Shifting EF Level	Maxim	Mean % CVT score	Standard Error
ImproveNewCVT	Weak	First Quantity	24.4	16.9
		Second Quantity	24.6	16.9
		Quality	05.6	16.9
		Relation	-02.7	16.9
		Politeness	-08.9	16.9
	Strong	First Quantity	01.9	27.9
		Second Quantity	-03.8	27.9
		Quality	01.5	27.9
		Relation	35.2	27.9
		Politeness	07.2	27.9
ImproveOldCVT	Weak	First Quantity	24.2	13.4
		Second Quantity	09.6	13.4
		Quality	34.6	13.4
		Relation	03.3	13.4
		Politeness	05.5	13.4
	Strong	First Quantity	13.1	22.1
		Second Quantity	-14.5	22.1
		Quality	13.1	22.1
		Relation	35.5	22.1
		Politeness	18.8	22.1

Since BPVS performance was found to correlate with one of the DCCS measures and one of the CVT measures, participants were regrouped by BPVS scores. The age-group distinction was removed and all children who attained a standardised score of 104 or greater for the BPVS were assigned to a high BPVS group. The remaining children were assigned to a low BPVS group. The previous analyses minus the partialling out of BPVS scores were then rerun within these new groupings (see Tables 6.31-6.36). Tables 6.37-6.48 report the means and standard deviations for these analyses. As for the previous analyses, children's DCCS improvement was classed as strong if their scores increased by 6 from pre- to post-training levels and weak if their scores increased by less than 6. Again, weak and strong division criteria were selected on the basis that they enabled as equal a split as possible in the number of children assigned to the two DCCS improvement groups. Of the 13 children with low BPVS scores given DCCS training, 5 demonstrated weak DCCS improvement and 8 demonstrated strong DCCS improvement. Of the 16 children with high BPVS scores given DCCS training, 8 demonstrated weak DCCS improvement and 8 demonstrated strong DCCS improvement. Children's 2 box post-training DCCS performance was classed as strong if their scores were 6 and weak if their scores were less than 6, with weak and strong division criteria selected on the basis that they enabled as equal a split as possible in the number of children assigned to the two DCCS performance groups. Of 13 children with low BPVS scores, 3 demonstrated weak post-training 2 box performance and 10 demonstrated strong post-training 2 box performance. Of 16 children with high BPVS scores, 4 demonstrated weak post-training 2 box performance and 12 demonstrated strong post-training 2 box performance. Children's 3 box post-training DCCS performance was classed as strong if their scores were 9 and weak if their scores were less than 9, with weak and strong division criteria chosen to enable as equal a split as possible in the number of children assigned to the two DCCS performance groups. Of 13 children with low BPVS scores, 8 demonstrated weak 3 box performance and 5 demonstrated strong 3 box performance. Of 16 children with high BPVS scores, 10 demonstrated weak 3 box performance and 6 demonstrated strong 3 box performance.

As in the preceding analyses, neither pre-to-post-training DCCS improvement, nor post-training DCCS scores were found to affect post-training CVT scores or pre-to post-training CVT improvement. However, in line with the pattern present in the whole sample analysis, and the analysis split by age, N was found to be unevenly distributed between post-training 2 box DCCS score groups formed for both the high and low BPVS groups. Thus only tentative conclusions can be formed from these analyses. There was no correlation between pre-training DCCS and pre-training CVT scores for the low BPVS group, $r=0.118$, $p>0.300$. However, there was a significant positive correlation between pre-training DCCS and pre-training CVT scores for the high BPVS group, $r=0.490$, $p<0.05$, although pre-training DCCS was not found to correlate with scores for the individual maxims presented in the pre-training CVT: r's<0.349, p's>0.09.

Table 6.31: *Results for the low BPVS group: Standardised scores <104 of 2 (2 box DCCS post training improvement level: <6 point increase = weak vs. 6 point increase = strong) x 5 (Maxim type) ANOVAs on post-training and improvement CVT measures.*

		df	F	p	partial Eta²
DCCS Improvement:	PostNewCVT	1,55	0.000	0.995	0.000
	PostOldCVT	1,55	3.618	0.062	0.062
	PostCombinedCVT	1,55	2.424	0.125	0.042
	ImproveNewCVT	1,55	0.320	0.574	0.006
	ImproveOldCVT	1,55	4.860	0.032	0.081
Maxim:	PostNewCVT	4,55	0.224	0.924	0.016
	PostOldCVT	4,55	2.551	0.049	0.156
	PostCombinedCVT	4,55	1.407	0.244	0.093
	ImproveNewCVT	4,55	0.417	0.795	0.029
	ImproveOldCVT	4,55	1.123	0.355	0.076
DCCS Improvement x Maxim:	PostNewCVT	4,55	0.374	0.826	0.026
	PostOldCVT	4,55	0.788	0.538	0.054
	PostCombinedCVT	4,55	0.527	0.717	0.037
	ImproveNewCVT	4,55	0.326	0.859	0.023
	ImproveOldCVT	4,55	0.165	0.955	0.012

Table 6.32: *Results for the low BPVS group of 2 (post-training DCCS level: 2 box scores: <6 = weak vs. = 6 = strong) x 5 (Maxim type) ANOVAs on post-training and improvement CVT measures.*

		df	F	p	partial Eta²
2 box DCCS:	PostNewCVT	1,55	0.446	0.507	0.008
	PostOldCVT	1,55	1.405	0.241	0.025
	PostCombinedCVT	1,55	0.153	0.697	0.003
	ImproveNewCVT	1,55	0.012	0.914	0.000
	ImproveOldCVT	1,55	1.811	0.184	0.032
Maxim:	PostNewCVT	4,55	0.424	0.791	0.030
	PostOldCVT	4,55	1.492	0.217	0.098
	PostCombinedCVT	4,55	1.002	0.414	0.068
	ImproveNewCVT	4,55	1.113	0.360	0.075
	ImproveOldCVT	4,55	0.662	0.621	0.046
2 box DCCS x Maxim:	PostNewCVT	4,55	0.784	0.541	0.054
	PostOldCVT	4,55	0.344	0.847	0.024
	PostCombinedCVT	4,55	0.251	0.908	0.018
	ImproveNewCVT	4,55	1.358	0.260	0.090
	ImproveOldCVT	4,55	0.527	0.716	0.037

Table 6.33: *Results for the low BPVS group of 2 (post-training DCCS level: 3 box scores: <9 = weak vs.= 9 = strong) x 5 (Maxim type) ANOVAs on post-training and improvement CVT measures.*

		df	F	p	partial Eta²
3 box DCCS:	PostNewCVT	1,55	0.601	0.442	0.011
	PostOldCVT	1,55	0.076	0.783	0.001
	PostCombinedCVT	1,55	0.690	0.410	0.012
	ImproveNewCVT	1,55	1.338	0.252	0.024
	ImproveOldCVT	1,55	0.910	0.344	0.016
Maxim:	PostNewCVT	4,55	0.102	0.981	0.007
	PostOldCVT	4,55	1.826	0.137	0.117
	PostCombinedCVT	4,55	1.056	0.387	0.071
	ImproveNewCVT	4,55	0.345	0.846	0.024
	ImproveOldCVT	4,55	1.273	0.292	0.085
3 box DCCS x Maxim:	PostNewCVT	4,55	0.578	0.680	0.040
	PostOldCVT	4,55	1.209	0.317	0.081
	PostCombinedCVT	4,55	0.951	0.442	0.065
	ImproveNewCVT	4,55	0.500	0.736	0.035
	ImproveOldCVT	4,55	1.141	0.347	0.077

Table 6.34: *Results for the high BPVS group: Standardised scores ≥ 104 of 2 (2 box DCCS post training improvement level: <6 point increase = weak vs. 6 point increase = strong) x 5 (Maxim type) ANOVAs on post-training and improvement CVT measures.*

		df	F	p	partial Eta^2
DCCS Improvement:	PostNewCVT	1,70	0.311	0.579	0.004
	PostOldCVT	1,70	0.524	0.472	0.007
	PostCombinedCVT	1,70	0.450	0.505	0.006
	ImproveNewCVT	1,70	0.698	0.406	0.010
	ImproveOldCVT	1,70	0.807	0.372	0.011
Maxim:	PostNewCVT	4,70	0.261	0.902	0.015
	PostOldCVT	4,70	1.288	0.283	0.069
	PostCombinedCVT	4,70	0.321	0.863	0.018
	ImproveNewCVT	4,70	0.436	0.782	0.024
	ImproveOldCVT	4,70	0.220	0.927	0.012
DCCS Improvement x Maxim:	PostNewCVT	4,70	1.346	0.262	0.071
	PostOldCVT	4,70	1.238	0.303	0.066
	PostCombinedCVT	4,70	0.351	0.843	0.020
	ImproveNewCVT	4,70	0.717	0.583	0.039
	ImproveOldCVT	4,70	1.572	0.191	0.082

Table 6.35: *Results for the high BPVS group of 2 (post-training DCCS level: 2 box scores: <6 = weak vs. = 6 = strong) x 5 (Maxim type) ANOVAs on post-training and improvement CVT measures.*

		df	F	p	partial Eta^2
2 box DCCS:	PostNewCVT	1,70	0.166	0.685	0.002
	PostOldCVT	1,70	0.533	0.468	0.008
	PostCombinedCVT	1,70	0.946	0.334	0.013
	ImproveNewCVT	1,70	1.147	0.288	0.016
	ImproveOldCVT	1,70	1.897	0.173	0.026
Maxim:	PostNewCVT	4,70	1.511	0.208	0.079
	PostOldCVT	4,70	1.398	0.244	0.074
	PostCombinedCVT	4,70	0.179	0.948	0.010
	ImproveNewCVT	4,70	0.939	0.447	0.051
	ImproveOldCVT	4,70	0.680	0.608	0.037
2 box DCCS x Maxim:	PostNewCVT	4,70	2.829	0.031	0.139
	PostOldCVT	4,70	1.874	0.125	0.097
	PostCombinedCVT	4,70	1.823	0.134	0.094
	ImproveNewCVT	4,70	0.848	0.500	0.046
	ImproveOldCVT	4,70	0.908	0.464	0.049

Table 6.36: *Results for the high BPVS group of 2 (post-training DCCS level: 3 box scores: <9 = weak vs. = 9 = strong) x 5 (Maxim type) ANOVAs on post-training and improvement CVT measures.*

		df	F	p	partial Eta^2
3 box DCCS:	PostNewCVT	1,70	0.216	0.644	0.003
	PostOldCVT	1,70	0.223	0.638	0.003
	PostCombinedCVT	1,70	0.096	0.758	0.001
	ImproveNewCVT	1,70	0.038	0.846	0.001
	ImproveOldCVT	1,70	0.051	0.822	0.001
Maxim:	PostNewCVT	4,70	0.042	0.997	0.002
	PostOldCVT	4,70	1.124	0.352	0.060
	PostCombinedCVT	4,70	0.399	0.809	0.022
	ImproveNewCVT	4,70	0.139	0.967	0.008
	ImproveOldCVT	4,70	0.447	0.774	0.025
3 box DCCS x Maxim:	PostNewCVT	4,70	1.740	0.151	0.090
	PostOldCVT	4,70	0.429	0.787	0.024
	PostCombinedCVT	4,70	1.391	0.246	0.074
	ImproveNewCVT	4,70	3.551	0.011	0.169
	ImproveOldCVT	4,70	0.968	0.431	0.052

Table 6.37: *CVT percentage score means and standard deviations for the maxims (first quantity, second quantity, quality, relation and politeness) produced across the post-training new items, post-training old items and combined post-training CVT conditions by children with low BPVS scores (<104) demonstrating weak and strong improvement in shifting.*

CVT measure	Shifting EF improvement	Maxim	Mean % CVT score	Standard Deviation
PostNewCVT score	Weak	First Quantity	46.6	38.1
		Second Quantity	63.4	21.7
		Quality	56.8	15.2
		Relation	56.6	25.4
		Politeness	70.0	27.4
	Strong	First Quantity	60.5	19.9
		Second Quantity	60.4	26.7
		Quality	58.3	28.3
		Relation	58.3	37.9
		Politeness	56.3	32.0
PostOldCVT	Weak	First Quantity	80.0	27.4
		Second Quantity	39.8	36.6
		Quality	73.4	25.3
		Relation	46.6	38.1
		Politeness	80.0	27.4
	Strong	First Quantity	43.6	28.1
		Second Quantity	27.0	23.4
		Quality	68.8	38.3
		Relation	50.0	39.9
		Politeness	50.0	37.8
PostCombinedCVT	Weak	First Quantity	60.0	28.3
		Second Quantity	52.0	17.9
		Quality	64.0	08.9
		Relation	52.0	22.8
		Politeness	75.0	17.7
	Strong	First Quantity	50.0	15.1
		Second Quantity	42.5	16.7
		Quality	62.5	22.5
		Relation	52.5	28.2
		Politeness	53.1	24.8

Table 6.38: *CVT percentage improvement means and standard deviations for the maxims (first quantity, second quantity, quality, relation and politeness) calculated using the post-training new items CVT measure and calculated using the post-training old items CVT measure for children with low BPVS scores (<104) demonstrating weak and strong shifting improvement.*

CVT measure	Shifting EF Improvement	Maxim	Mean % CVT score	Standard Deviation
ImproveNewCVT	Weak	First Quantity	-06.8	48.0
		Second Quantity	26.8	32.7
		Quality	13.4	57.3
		Relation	33.4	44.1
		Politeness	20.0	27.4
	Strong	First Quantity	10.5	43.7
		Second Quantity	14.6	51.7
		Quality	10.3	47.3
		Relation	12.5	47.0
		Politeness	06.3	41.7
ImproveOldCVT	Weak	First Quantity	26.0	43.4
		Second Quantity	03.2	34.0
		Quality	30.0	44.7
		Relation	23.2	43.5
		Politeness	30.0	27.4
	Strong	First Quantity	-06.3	33.3
		Second Quantity	-18.6	20.7
		Quality	20.9	34.2
		Relation	04.1	50.9
		Politeness	00.0	53.5

Table 6.39: *CVT percentage score means and standard deviations for the maxims (first quantity, second quantity, quality, relation and politeness) produced by weak and strong post-training 2 box shifters who had low BPVS scores (<104), across the post-training new items, post-training old items and combined post-training CVT conditions..*

CVT measure	Shifting EF level	Maxim	Mean % CVT score	Standard Deviation
PostNewCVT score	Weak	First Quantity	33.3	33.5
		Second Quantity	50.0	00.0
		Quality	55.7	19.6
		Relation	66.7	28.9
		Politeness	66.7	28.9
	Strong	First Quantity	61.7	23.7
		Second Quantity	65.0	26.7
		Quality	58.3	25.3
		Relation	54.9	34.4
		Politeness	60.0	31.6
PostOldCVT	Weak	First Quantity	66.7	28.9
		Second Quantity	55.3	38.7
		Quality	83.3	28.9
		Relation	44.3	19.6
		Politeness	66.7	28.9
	Strong	First Quantity	54.9	34.4
		Second Quantity	24.9	22.5
		Quality	66.7	34.3
		Relation	50.0	42.4
		Politeness	60.0	39.4
PostCombinedCVT	Weak	First Quantity	46.7	23.1
		Second Quantity	53.3	23.1
		Quality	66.7	11.5
		Relation	53.3	11.5
		Politeness	66.7	14.4
	Strong	First Quantity	56.0	20.7
		Second Quantity	44.0	15.8
		Quality	62.0	19.9
		Relation	52.0	28.6
		Politeness	60.0	26.9

Table 6.40: *CVT percentage improvement means and standard deviations for the maxims (first quantity, second quantity, quality, relation and politeness) calculated using the post-training new items CVT measure and calculated using the post-training old items CVT measure for children with low BPVS scores (<104) demonstrating weak and strong post-training 2 box shifting.*

CVT measure	Shifting EF Level	Maxim	Mean % CVT score	Standard Deviation
ImproveNewCVT	Weak	First Quantity	-33.3	44.4
		Second Quantity	16.7	33.5
		Quality	05.7	67.7
		Relation	55.7	41.8
		Politeness	16.7	28.9
	Strong	First Quantity	15.0	39.7
		Second Quantity	20.1	48.5
		Quality	13.2	46.6
		Relation	10.0	42.6
		Politeness	10.0	39.4
ImproveOldCVT	Weak	First Quantity	00.0	00.0
		Second Quantity	22.0	19.1
		Quality	33.3	57.7
		Relation	33.0	00.0
		Politeness	16.7	28.9
	Strong	First Quantity	08.0	45.1
		Second Quantity	-19.9	21.9
		Quality	21.7	32.4
		Relation	05.0	52.7
		Politeness	10.0	51.6

Table 6.41: *CVT percentage score means and standard deviations for the maxims (first quantity, second quantity, quality, relation and politeness) produced by weak and strong post-training 3 box shifters who had low BPVS scores (<104), across the post-training new items, post-training old items and combined post-training CVT conditions..*

CVT measure	Shifting EF level	Maxim	Mean % CVT score	Standard Deviation
PostNewCVT score	Weak	First Quantity	58.4	34.6
		Second Quantity	64.6	22.6
		Quality	60.5	21.9
		Relation	52.0	20.9
		Politeness	68.8	25.9
	Strong	First Quantity	50.0	12.0
		Second Quantity	56.6	28.1
		Quality	53.2	27.5
		Relation	66.6	47.2
		Politeness	50.0	35.4
PostOldCVT	Weak	First Quantity	68.8	37.2
		Second Quantity	24.9	34.5
		Quality	77.1	25.1
		Relation	41.6	34.6
		Politeness	62.5	35.4
	Strong	First Quantity	39.8	09.3
		Second Quantity	43.2	09.3
		Quality	60.0	43.5
		Relation	60.0	43.5
		Politeness	60.0	41.8
PostCombinedCVT	Weak	First Quantity	60.0	23.9
		Second Quantity	45.0	20.7
		Quality	67.5	14.9
		Relation	47.5	21.2
		Politeness	65.6	22.9
	Strong	First Quantity	44.0	08.9
		Second Quantity	48.0	11.0
		Quality	56.0	21.9
		Relation	60.0	31.6
		Politeness	55.0	27.4

Table 6.42: *CVT percentage improvement means and standard deviations for the maxims (first quantity, second quantity, quality, relation and politeness) calculated using the post-training new items CVT measure and calculated using the post-training old items CVT measure for children with low BPVS scores (<104) demonstrating weak and strong post-training 3 box shifting..*

CVT measure	Shifting EF Level	Maxim	Mean % CVT score	Standard Deviation
ImproveNewCVT	Weak	First Quantity	12.5	51.8
		Second Quantity	33.5	36.8
		Quality	16.8	60.1
		Relation	16.8	48.1
		Politeness	12.5	35.4
	Strong	First Quantity	-10.0	28.1
		Second Quantity	-03.4	49.5
		Quality	03.0	27.4
		Relation	26.6	45.1
		Politeness	10.0	41.8
ImproveOldCVT	Weak	First Quantity	22.5	36.5
		Second Quantity	-06.3	30.6
		Quality	33.4	38.9
		Relation	06.1	54.8
		Politeness	06.3	49.6
	Strong	First Quantity	-20.0	29.9
		Second Quantity	-16.6	23.5
		Quality	10.0	32.3
		Relation	20.0	36.0
		Politeness	20.0	44.7

Table 6.43: *CVT percentage score means and standard deviations for the maxims (first quantity, second quantity, quality, relation and politeness) produced across the post-training new items, post-training old items and combined post-training CVT conditions by children with high BPVS scores (≥104) demonstrating weak and strong improvement in shifting.*

CVT measure	Shifting EF improvement	Maxim	Mean % CVT score	Standard Deviation
PostNewCVT score	Weak	First Quantity	77.1	36.7
		Second Quantity	52.0	24.5
		Quality	62.5	44.3
		Relation	45.8	35.4
		Politeness	56.3	32.0
	Strong	First Quantity	39.6	26.7
		Second Quantity	58.4	22.0
		Quality	62.5	36.5
		Relation	56.3	21.9
		Politeness	56.3	41.7
PostOldCVT	Weak	First Quantity	50.0	31.0
		Second Quantity	60.4	25.2
		Quality	56.3	23.7
		Relation	83.4	23.5
		Politeness	68.8	37.2
	Strong	First Quantity	58.3	38.9
		Second Quantity	58.4	20.0
		Quality	52.0	22.8
		Relation	52.1	10.9
		Politeness	75.0	37.8
PostCombinedCVT	Weak	First Quantity	62.5	16.7
		Second Quantity	55.0	17.7
		Quality	60.0	28.3
		Relation	62.5	24.9
		Politeness	62.5	26.7
	Strong	First Quantity	50.0	23.9
		Second Quantity	57.5	12.8
		Quality	57.5	12.8
		Relation	55.0	14.1
		Politeness	65.6	35.2

Table 6.44: *CVT percentage improvement means and standard deviations for the maxims (first quantity, second quantity, quality, relation and politeness) calculated using the post-training new items CVT measure and calculated using the post-training old items CVT measure for children with high BPVS scores (≥104) demonstrating weak and strong shifting improvement.*

CVT measure	Shifting EF Improvement	Maxim	Mean % CVT score	Standard Deviation
ImproveNewCVT	Weak	First Quantity	14.6	60.1
		Second Quantity	-12.6	30.6
		Quality	-02.3	58.8
		Relation	-08.3	42.9
		Politeness	-12.5	35.4
	Strong	First Quantity	-06.3	49.7
		Second Quantity	14.8	26.1
		Quality	20.8	57.8
		Relation	-02.1	27.5
		Politeness	-06.3	41.7
ImproveOldCVT	Weak	First Quantity	-12.5	39.5
		Second Quantity	-04.1	47.7
		Quality	-08.4	34.4
		Relation	29.1	36.5
		Politeness	00.0	46.3
	Strong	First Quantity	12.4	56.1
		Second Quantity	14.5	20.7
		Quality	10.4	26.5
		Relation	-06.3	25.0
		Politeness	12.5	44.3

Table 6.45: *CVT percentage score means and standard deviations for the maxims (first quantity, second quantity, quality, relation and politeness) produced by weak and strong post-training 2 box shifters who had high BPVS scores (≥104), across the post-training new items, post-training old items and combined post-training CVT conditions..*

CVT measure	Shifting EF level	Maxim	Mean % CVT score	Standard Deviation
PostNewCVT score	Weak	First Quantity	79.3	24.9
		Second Quantity	62.5	28.6
		Quality	75.0	28.9
		Relation	16.5	19.1
		Politeness	37.5	25.0
	Strong	First Quantity	51.4	38.0
		Second Quantity	52.8	21.3
		Quality	58.3	42.4
		Relation	62.5	21.6
		Politeness	62.5	37.7
PostOldCVT	Weak	First Quantity	33.3	27.4
		Second Quantity	70.8	34.5
		Quality	54.3	16.3
		Relation	79.3	24.9
		Politeness	50.0	40.8
	Strong	First Quantity	61.1	34.4
		Second Quantity	55.6	16.5
		Quality	54.1	24.9
		Relation	63.9	23.4
		Politeness	79.2	33.4
PostCombinedCVT	Weak	First Quantity	55.0	19.1
		Second Quantity	65.0	19.1
		Quality	65.0	10.0
		Relation	45.0	19.1
		Politeness	43.8	23.9
	Strong	First Quantity	56.7	22.3
		Second Quantity	53.3	13.0
		Quality	56.7	23.9
		Relation	63.3	18.7
		Politeness	70.8	29.8

Table 6.46: *CVT percentage improvement means and standard deviations for the maxims (first quantity, second quantity, quality, relation and politeness) calculated using the post-training new items CVT measure and calculated using the post-training old items CVT measure for children with high BPVS scores (≥104) demonstrating weak and strong post-training 2 box shifting.*

CVT measure	Shifting EF Level	Maxim	Mean % CVT score	Standard Deviation
ImproveNewCVT	Weak	First Quantity	12.5	45.9
		Second Quantity	-04.3	28.3
		Quality	12.3	34.5
		Relation	-41.8	21.6
		Politeness	-25.0	28.9
	Strong	First Quantity	01.4	58.5
		Second Quantity	02.8	32.7
		Quality	08.3	64.7
		Relation	07.0	29.9
		Politeness	-04.2	39.6
ImproveOldCVT	Weak	First Quantity	-33.3	47.1
		Second Quantity	04.3	67.1
		Quality	-08.3	16.5
		Relation	20.8	24.9
		Politeness	-12.5	62.9
	Strong	First Quantity	11.0	45.6
		Second Quantity	05.5	24.9
		Quality	04.1	34.8
		Relation	08.3	38.5
		Politeness	12.5	37.7

Table 6.47: *CVT percentage score means and standard deviations for the maxims (first quantity, second quantity, quality, relation and politeness) produced by weak and strong post-training 3 box shifters who had high BPVS scores (≥104), across the post-training new items, post-training old items and combined post-training CVT conditions..*

CVT measure	Shifting EF level	Maxim	Mean % CVT score	Standard Deviation
PostNewCVT score	Weak	First Quantity	65.1	30.9
		Second Quantity	56.6	26.5
		Quality	73.3	37.1
		Relation	39.9	29.7
		Politeness	55.0	36.9
	Strong	First Quantity	47.2	45.3
		Second Quantity	52.8	16.7
		Quality	44.5	39.0
		Relation	69.5	16.4
		Politeness	58.3	37.6
PostOldCVT	Weak	First Quantity	49.9	32.5
		Second Quantity	63.3	27.1
		Quality	58.3	25.3
		Relation	66.7	24.9
		Politeness	75.0	35.4
	Strong	First Quantity	61.2	39.0
		Second Quantity	52.8	06.9
		Quality	47.2	16.7
		Relation	69.5	24.5
		Politeness	66.7	40.8
PostCombinedCVT	Weak	First Quantity	56.0	20.7
		Second Quantity	58.0	17.5
		Quality	66.0	16.5
		Relation	52.0	21.5
		Politeness	65.0	31.6
	Strong	First Quantity	56.7	23.4
		Second Quantity	53.3	10.3
		Quality	46.7	24.2
		Relation	70.0	11.0
		Politeness	62.5	30.6

Table 6.48: *CVT percentage improvement means and standard deviations for the maxims (first quantity, second quantity, quality, relation and politeness) calculated using the post-training new items CVT measure and calculated using the post-training old items CVT measure for children with high BPVS scores (≥104) demonstrating weak and strong post-training 3 box shifting.*

CVT measure	Shifting EF Level	Maxim	Mean % CVT score	Standard Deviation
ImproveNewCVT	Weak	First Quantity	18.4	42.7
		Second Quantity	03.3	30.4
		Quality	21.5	49.9
		Relation	-26.8	19.5
		Politeness	-20.0	35.0
	Strong	First Quantity	-19.5	67.2
		Second Quantity	-02.7	34.3
		Quality	-11.2	68.3
		Relation	30.8	22.1
		Politeness	08.3	37.6
ImproveOldCVT	Weak	First Quantity	03.3	50.7
		Second Quantity	10.0	42.4
		Quality	06.6	20.9
		Relation	00.0	28.2
		Politeness	00.0	40.8
	Strong	First Quantity	-05.7	49.0
		Second Quantity	-02.8	26.6
		Quality	-08.3	44.4
		Relation	30.5	40.0
		Politeness	16.7	51.6

6.4 Discussion

This study was concerned with applying the results of Experiment 3. A training study was employed to enhance the shifting EF ability of young children to investigate whether this would lead to a corresponding increase in CU. The training program was found to lead to a significant increase in shifting EF, assessed and trained using the DCCS task. This finding is in line with the report of increased EF following training, from Kloo and Perner (2003), on whose procedure the EF training program was based. Furthermore, whilst pre-training DCCS performance had been at significantly below-chance levels, post-training levels were at significantly greater than chance levels, indicating the attainment of DCCS competence. However, because a control training condition was not included in the study, due to a lack of significant differences between pre- and post-training CVT scores, conclusions based on the effects of DCCS training should only be drawn tentatively. It is possible, albeit unlikely, that the significant improvement in DCCS scores could be attributable to general training effects such as goal focussing, which could have been imparted through a control training condition. Without a comparison of DCCS improvement across the two training conditions, such a possibility cannot be ruled out.

Nevertheless, in contrast to expectations and the results of Experiment 3, multiple varied analyses did not indicate that DCCS training resulted in a significantly enhanced CVT performance, that improvements in EF led to significant corresponding rises in CVT scores, that DCCS performance post-training affected CVT scores or improvement or even that pre-training DCCS scores were related to pre-training CVT performance once vocabulary ability was taken into account. This was the case when the entire sample was considered, but also when the performance of the 3- and 4-year-old age groups was considered separately and when the performance of children attaining high standardised vocabulary scores was considered separately from the performance of children attaining low standardised vocabulary scores. An exception was the single relationship found between pre-training DCCS and pre-training CVT scores for the group attaining high standardised vocabulary scores.

However, the effectiveness of the DCCS training entailed an unequal division of children across post-training DCCS ability groups with the majority performing at ceiling. This means that conclusions regarding the absence of DCCS-CVT relations in the post-training phase should be drawn with caution.

Vocabulary scores were entered as a covariate when children were not split into groups based on their vocabulary performance to ensure that any relationships found were not simply reflecting the relationships between BPVS and pre-training DCCS performance and between BPVS and performance on the post-training Old CVT. Since vocabulary scores reflect linguistic proficiency and CVT performance reflects CU, which is an aspect of linguistic proficiency, one might expect the positive correlation found between BPVS and CVT scores. Furthermore, the implication that linguistic ability plays a role in EF (Vygotsky, 1962 ; Singer & Bashir, 1999) accounts for the relationship found between BPVS scores and EF ability. Nevertheless, tests indicated that although pre-training CVT scores were merely at chance levels, post-training CVT scores were significantly better than chance. This is consistent with the theory that a degree of CVT improvement had followed DCCS training. The CVT improvement from chance levels prior to DCCS training to significantly above chance levels, post-training, provides some support for the suggestion that CVT performance improved following DCCS training. However, the fact that the pre- to post-training CVT improvements were not significant suggests that the influence of EF might be fairly restricted, so that a larger number of participants would be required to reveal a significant effect. Nevertheless, as indicated above, the lack of a control training condition means that any conclusions based on the effects of DCCS training can only be drawn tentatively. The possibility remains that any significant improvement in post-training CVT scores could be attributable to general training effects such as goal focussing, which could have been imparted through a control training condition. Without a comparison of CVT improvement across the two training conditions, such a possibility cannot be ruled out.

More fine-grained analyses investigating scores for groups of CVT items reflecting individual conversational maxims or rules, indicated that the potentially positive influence of DCCS training was restricted to the groups of CVT items reflecting the Maxims of Quality and Politeness. Although the difference between performance on items reflecting the different maxims was not itself significant, nor the difference between pre- and post-training performance for any of these maxims, improvement from pre-training chance performance to post-training above chance performance was only demonstrated for items reflecting the Quality and Politeness maxims. This complements the findings of Siegal, Iozzi and Surian (2009), which demonstrated that superior bilingual performance on the CVT did not extend to items representing the Maxim of First Quantity, providing support for their suggestion that the items intended to reflect the Maxim of First Quantity might have lacked validity (see Section 5.4). However, Siegal, Iozzi and Surian's suggestion refers only to the CVT items reflecting the Maxim of First Quantity and so does not explain why EF training did not appear to have an effect on performance on items reflecting the Maxims of Second Quantity and Relation. Thus, an argument put forward in Section 5.4 appears suitable to restate here: although it has been suggested that the effect of EF on CU is mediated by ToM, it is not clear that ToM is the factor identified in this instance, as it is not obvious as to why children would demonstrate unequal competence across the different maxims/conversational rules. It seems more likely that access to world knowledge is the factor involved here as the demands for world knowledge would seem much more likely to differ across the maxim-based groups of CVT items.

The absence of significant relationships between improvements in shifting EF and CVT scores in Experiment 4, seems to contrast with the findings of Experiment 3, which demonstrated a significant relationship between shifting EF ability and CVT performance. However, whilst 4- and 5-year-olds were recruited for Experiment 3, Experiment 4 used only 3- and 4-years-olds. The decision to restrict recruitment for Experiment 4 to 3- and 4-year-olds arose from the need to identify a large number of children who were unable to pass the DCCS task. These children could then be trained on the DCCS task. Since most children pass the DCCS task by 5 years of age (Zelazo, 2006), inclusion of 5-year-olds would have required a greater number of participants to be tested during the screening/pre-training phase, before a sufficient number of children would have been deemed eligible for the DCCS training program. Indeed, in Experiment 3, which had demonstrated the relationship between DCCS performance and CVT scores, only 11 of 60 participants aged 4 to 5 years, had attained less than perfect scores on the DCCS, of which 3 were 5 years old.

However, the very fact that DCCS performance improves with age (Perner & Lang, 2002) and thus that failing the DCCS at older ages is unusual, may indicate that the relationship between DCCS performance and CVT scores found in the older age group tested in Experiment 3, is merely incidental. It is conceivable that a third factor, which is related to the DCCS failure in older children, is also related to CVT performance. For example, if ToM is necessary for the development of EF, as some would argue (Hughes & Graham 2002), it could be that the older children failing the DCCS in Experiment 3, were doing so due to a ToM deficit. Since ToM is believed to be related to CU, the DCCS failures in Experiment 3 could be accompanied by poor CVT performance. However, in such a scenario, the DCCS failures and poor CU would not be directly related, merely indirectly linked by their dependence on ToM. It could further be, that DCCS failures in younger children are more reflective of a different cause such as limitations in working memory capacity. Since working memory capacity is more restricted at younger ages (Swanson, 1996), limitations in working memory might be the main cause of DCCS difficulties in the younger children. It is conceivable that, following age-related improvements in working memory, only those children with ToM deficits still experience problems with the DCCS task.

However, the suggestion that the relationship found between CU and shifting EF is age-dependent, demonstrated only by older children experiencing a deficit in another domain such as ToM, is undermined by the lack of evidence for a significant CU-EF relationship when children were separated into older and younger age-groups. The older group in Experiment 4, which comprised 4-year-olds, corresponded to the younger age-group tested in Experiment 3, which also comprised 4-year-olds.

Experiment 4 is the first examination of the effects of EF training on CU. Thus, there are no previous, directly-related findings in the literature, with which to compare the results of this study. However, it is worth noting that only one of the two previously mentioned EF training studies undertaken to promote ToM (Kloo & Perner, 2003; Fisher & Happé, 2005) produced convincing evidence that EF enhancements promoted ToM. Kloo and Perner (2003) found that following EF training, children performed significantly better on both EF and ToM tasks. However, Fisher and Happé (2005) reported that although EF trained children with autism performed significantly better in post-test and follow-up ToM tests relative to pre-test ToM tasks, they did not demonstrate significantly enhanced performance on EF tasks given post-test or follow-up. This result suggests that the improvements in ToM which were reported, cannot be attributed to enhanced EF. Since it has previously been suggested that CU is an aspect of ToM (see Section 1.6), one might attribute difficulties in promoting CU through EF training to difficulties promoting ToM in this manner.

However, it should be borne in mind, that the EF training program employed in this study followed closely the procedure of the very study which appeared to promote ToM through the enhancement of EF. Furthermore, the study which failed to find that EF-training promoted ToM, employed a clinical population: participants with autism. Thus, the absence of an EF training effect on ToM in the Fisher and Happé (2005) study could be attributable to a specific characteristic of people with autism, rather than a characteristic of the more general, non-clinical population.

Furthermore, Siegal, Iozzi and Surian (2009) investigated whether advanced shifting EF abilities could account for the bilingual advantage they found in CVT performance. Siegal et al. found that although Italian-Slovenian bilinguals and Slovenian monolinguals outperformed Italian monolinguals on shifting ability, entering shifting ability as a covariate did not remove the bilingual advantage in CVT performance.

Although consistent with the theory that EF promotes CU, the relationship found between shifting EF and CVT performance in Experiment 3 is also compatible with the suggestion that CU promotes EF. For this reason, initial attempts were made to incorporate CVT training into Experiment 4. However, piloting revealed that various forms of CVT training failed to improve the performance of 3- and 4-year-olds on the CVT. For this reason, the CVT training condition was left out of Experiment 4. Thus, it remains a possibility that the relationship found between shifting EF and CVT performance in Experiment 3 is due to the facilitating effect of CU on EF. It is therefore conceivable, that the employment of an effective CVT training program would have revealed the enhancing effect of CU on EF.

In line with expectations, EF training was not found to affect performance on the control task. However, only BPVS scores and performance in the pre-training control task appeared to improve significantly with age in Experiment 4. The children tested in this experiment, were specifically selected on the basis of their poor performance in the pre-training DCCS task, thus an age effect would not be expected for performance on the DCCS task. Since performance on the pre-training CVT was at chance levels, it seems reasonable that CVT performance did not demonstrate an age effect either.

6.5 Conclusions

Experiment 4 found that enhancing the shifting EF of 3- and 4-year-olds through training did not significantly improve their performance on the CVT measure of CU. Thus the current study was not able to provide strong support for the masked competence account of CU. Although CVT performance was not found to improve significantly following training, or to be significantly related to DCCS performance subsequent to training, it was nonetheless found to have changed from chance to significantly above chance levels, consistent with the proposal that shifting EF was promoting CU levels. However, a control training condition would be required to establish that this was not the consequence of general training effects. Experiment 5 was conducted to examine whether bilingual children, whom the literature suggests demonstrate advanced EF, would possess enhanced CU. Evidence that bilingual children demonstrate advanced CU would provide stronger support for the masked competence account of CU.