The cognitive underpinnings of referential abilities

Cécile De Cat
University of Leeds

Abstract

The aim of the chapter is to take stock and bring together fields of enquiry relevant to the investigation of preschool children's referential abilities. After an appraisal of the evidence as to what adults' referential abilities really are, the key cognitive prerequisites for unhindered performance in experimental tasks assessing referential abilities are identified. An estimation of when these cognitive prerequisites are in place in young children is then provided, on the basis of a review of the relevant developmental psychology literature. This leads to concrete predictions as to the emergence of referential abilities in preschool children, upon which one can define a programme of research to progress our understanding of children's referential abilities.

Keywords: executive function, theory of mind, egocentricity, working memory, perspective, visual context, common ground, salience, discourse

1 Introduction

The encoding of information in conversation is intrinsically an interactive process, in which the speaker has to evaluate the hearer's knowledge state, i.e., monitor which referents are new or available from their point of view, in order to determine how best to encode them in their own speech. To adequately map the information status of referents onto the appropriate structures and lexical items, children need (i) to acquire the required linguistic competence and (ii) to develop the cognitive ability to perceive differences in perspective between one's own and that of one's addressee. Research spanning four decades has shown that children's referential abilities remain patchy until at least five years of age, as evidenced (among other things) by their over-use of definite articles (1) and pronouns (2).[1]

(1) The boy is sitting on a chair. (Context: boy not previously mentioned)

(2) She wants to get out. (Context: referent not previously mentioned)

In both cases, the expression chosen is insufficient to disambiguate reference adequately from the addressee's point of view. The exact cause of these errors has been much debated

[1]Definite article and pronoun over-use by young children have been documented across languages, in a large body of literature. See De Cat (2011) for a review of studies on definiteness overuse and e.g., Karmiloff-Smith (1985) on pronoun overuse.

in the literature. One tradition in particular attributes children's definiteness errors to an egocentric stage in their cognitive development (Karmiloff-Smith, 1979), or its manifestation as an immature pragmatic principle (Schaeffer & Matthewson, 2005). This predicts a correlation between the production of definiteness errors and the relevant Theory of Mind component (Gundel, 2009; Schafer & de Villiers, 2000), but such a direct correlation may in fact not exist (De Cat, 2013).

In this contribution, I explore another cognitive avenue as a possible explanation for children's over-use of forms (such as definites and pronouns) that assume a high level of shared knowledge with their interlocutor, namely: executive function skills. Since Miyake and colleagues' seminal work (Friedman et al., 2006; Miyake et al., 2000) executive functions have been divided into three separable components: the inhibition of unwanted responses, the shifting between tasks and mental sets (i.e., flexibility), and the updating and monitoring of working memory representations. Some argue that executive function is necessary to develop mental state concepts (e.g., Carlson & Moses, 2001). Executive function skills have been shown to predict perspective-taking abilities in adults (e.g., Brown-Schmidt, 2009). I discuss the evidence available to date and lay out the predictions regarding the acquisition of reference by monolingual and bilingual children. The backdrop for this discussion is a broader exploration of the cognitive prerequisites of referential abilities.

After an appraisal of the evidence as to what adults' referential abilities really are (section 2), I list the key cognitive prerequisites for unhindered performance in experimental tasks assessing referential abilities (section 3) and then review the developmental psychology literature to estimate when these cognitive prerequisites are in place (section 4), ending with concrete predictions as to the emergence of referential abilities in preschool children.

The aim of the chapter is to take stock and bring together fields of enquiry relevant to the investigation of preschool children's referential abilities. Key concepts will therefore be introduced throughout the chapter by necessity, and the literature review cannot aim to be comprehensive — rather, it will tend to focus on seminal or cutting-edge studies. The argument developed here is not dependent on the adoption of a particular theoretical framework. The ambition of this contribution is to set out a programme of research to progress our understanding of children's referential abilities, through a better understanding of their cognitive underpinnings.

2 Adults' referential abilities

2.1 Components of perspective evaluation

Stalnaker (1974, 1978) distinguishes two levels of accessibility of referents: those pertaining to the common ground (i.e., shared by speaker and addressee) and those pertaining to the privileged ground (i.e., available only from the perspective of one of the interlocutors, be it the speaker *or* the addressee). In a conversation, the speaker is assumed to make available information from their privileged ground to the addressee, so that it becomes part of the common ground. What is part of the common ground is not limited to prior linguistic

mention: it can also be established by physical co-presence (e.g., in the visual field via joint attention), and experiences or world knowledge shared by speaker and hearer (Clark & Marshall, 1981). Accordingly, various factors determine the accessibility status of discourse referents: textual distance from a previous mention in the discourse (Ariel, 1990), information-structural status of the previous mention (e.g., focus or topic — (Reinhart, 1981; Gundel, Hedberg, & Zacharski, 1993)), ambiguity arising from the presence of a competitor in the discourse context (Arnold & Griffin, 2007) or the visual context (Fukumura, 2010), grammatical and thematic role in the previous sentence (Arnold, 2001; Arnold & Tanenhaus, 2011), animacy (Fukumura & Van Gompel, 2011), relevance for the current purpose (Gundel et al., 1993), interlocutor's attention state (Arnold, 2010),...

In light of this complexity, the distinction between common and privileged ground should be conceived more as a continuum than a strict dichotomy reflecting the cognitive status of discourse referents. In that spirit, Gundel et al. (1993) propose a further refinement, arguing that six cognitive statuses are necessary to account for the use of referring expressions in discourse. Their Givenness Hierarchy (3) predicts how referring expressions are mapped onto cognitive statuses (corresponding with the level of accessibility of a referent to the addressee), thereby regulating their use and interpretation.

(3)

in focus >	activated >	familiar >	uniquely > identifiable	referential >	type-identifiable
it	*this, that this N*	*that N*	*the N*	*indefinite this N*	*a N*

However, even in Gundel et al.'s (1993) fine-grained Givenness Hierarchy, there is no one-to-one relationship between type of referential expression and cognitive status. And the evaluation of cognitive status (e.g., whether a referent is considered activated or familiar) depends on how the discourse participants evaluate it with respect to the factors listed above (textual distance, relevance, etc). As pointed out by Heller, Grodner, and Tanenhaus (2009), information about interlocutors' knowledge states have to be deduced from different cues, with varying levels of certainty. This depends to a large extent on the goals of the interlocutors (Trueswell, Papafragou, & Choi, 2011)[2] and the focus of attention of participants (Garrod, 2011).

In spite of general agreement that the visual context plays an important role, there is no consensus on how it ranks with respect to the discourse context (especially in children). Garrod proposes that (in adults,) the visual context has a stronger influence on the situation model[3] than the discourse context, in that fixation on a particular referent makes other

[2]Quoting Trueswell et al. (2011): "The shape of the discourse and the goals of the interlocutors ought to be a far better predictor of the referential domain of a referential expression and hence a better predictor of the level of specificity needed for that expression" (p.86). "In order to capture this flexibility in interpreting referential expressions, any account of reference must include what is plausible given the goals of the interlocutors as a matter of course" (p.93).

[3]The situation model is a representation of the discourse world held within working memory. It includes "information about space, time causality, intentionality, and currently relevant individuals" (Garrod, 2011,

potential referents less accessible (Garrod, 2011, p. 283). However, as shown by Vogels, Krahmer, and Maes (2013b), the impact of perceptual factors on the evaluation of referent accessibility (by adults) does not determine the choice of referring expression as much as linguistic salience. Grodner, Dalini, Pearlstein-Levy, and Ward (2012) even argue that for typical adults, language is a more reliable source than visual context to establish the common ground.

2.2 Egocentricity in adults?

Most discourse models assume that speakers' choice of referring expressions depends on the referent's salience or prominence in the context (e.g., Almor & Nair, 2007; Ariel, 1990; Gundel et al., 1993), and that this choice is driven by an endeavour to facilitate comprehension. This has recently been disputed by Fukumura and Van Gompel (2012), who show that adult speakers rely more on their *own* perspective than that of their addressee when choosing referential expressions: In Fukumura and Van Gompel's experiments, speakers used more pronouns for referents mentioned in the immediately preceding sentence, irrespective of whether that sentence was shared with the addressee (via loudspeakers). Similarly, Vogels, Krahmer, and Maes (2013a) show that, under increased cognitive load, speakers use significantly more pronouns for referents that are not (visually) salient to their addressee.

The findings from these production studies are in line with those of some real-time comprehension studies, which have shown that adults sometimes fail to deploy the ability to distinguish their own representation of the discourse context from others'. Keysar, Barr, Balin, and Bruauner (2000) even suggest that adults' initial step in reference resolution is determined by their egocentric perspective. In that eye-tracking study, adult participants were asked to manipulate objects in cubbyholes, some of which they knew were hidden from the instruction-giver (who was sitting on the other side of the cubbyholes). The instruction-followers were found *not* to restrict their search for referents to the objects in the common ground (i.e., visible to both themselves and the instruction-giver). Keysar et al. (2000) see these egocentric interpretations as the result of a strategy to alleviate the cognitive cost associated with the monitoring of the common and privileged grounds. They conclude that the implications are either (i) that mutual knowledge is only used for error correction, and not for referent identification, or (ii) that mutual knowledge acts as a constraint on the probability of reference to objects (with non-shared objects less likely to be referred to). Keysar, Lin, and Barr (2003) go as far as contending that "elements of [...] theory of mind are not fully incorporated in the human comprehension system". The more radical interpretation of these findings is that the initial stage of processing is necessarily egocentric, and the integration of ground information is only triggered by circumstances highlighting the presence of ambiguity (Keysar et al., 2000) or for repair strategies (Pickering & Garrod, 2004).[4] In line with this interpretation, Epley, Morewedge,

p. 274). Whether it can be equated with the discourse model is open to debate, and beyond the scope of the present chapter.

[4]It is worth noting that a confound might have obscured the results of Keysar et al.'s experiments:

and Keysar (2004) argue that the essential difference between children and adults lies in the speed at which adults correct an initial egocentric interpretation, and that the likelihood of that initial egocentric interpretation is equally great in adults and children. They conclude that "egocentrism isn't outgrown as much as it is overcome each time a person attempts to adopt another person's perspective" (Epley et al., 2004, p. 765). The relevant automatic processes might therefore be identical in adults and children, but with different outcomes arising from differences in controlled mental processes.

From these production and comprehension studies, it emerges that adult speakers are somehow biased by their own perspective, and that the extent to which they can take their addressee's perspective into account is limited, and costly in terms of cognitive processing. Recent research suggests that the cognitive skills involved in perspective-taking are those of the executive function (e.g., Brown-Schmidt (2009); Grodner et al. (2012)).

2.3 The cognitive cost of perspective evaluation

The executive function (EF) encompasses a set of cognitive skills used in goal-directed behaviours in order to override more automatic or established thoughts or responses. EF can be viewed as a unitary construct encompassing partially dissociable components, i.e., components that have common and unique variance (Miyake et al., 2000). The main EF components are (i) information updating and monitoring (i.e., working memory), (ii) inhibition of prepotent responses, and (iii) mental set shifting (i.e., the ability to change from one mental set to another, conflicting one). All EF component operations require the involvement of a central attention system (Rothbart & Posner, 2001).

Brown-Schmidt (2009) demonstrates that lower levels of inhibition control correlate with occasional insensitivities to perspective in adults. Participants were involved in a dialogue-based game in which they had access to visual information that was partially different to that of the experimenter. This information consisted in a grid with each cell containing a picture of an animal wearing an accessory. Only a third of the pictures was available to both interlocutors, and the remainder were non-shared (a third was available only to the experimenter, and another third only to the participant). The participants were asked questions by the experimenter to identify an animal that was missing in their own (i.e., experimenter's) grid. Each question featured an early ambiguity point (e.g., *the cow* in (4)) as the first noun phrase had two possible referents in the participant's display. The ambiguity was resolved in the latter part of the question (because similar animals always wore different accessories).

(4) What is above the cow that's wearing shoes?

In half of the critical trials, one of the two possible referents had been verbally established

Heller et al. (2009) point out that in that design, competitor objects in the listener's privileged ground fitted the referring expressions used by the instructor better than the object in the common ground. In their own study, Heller et al. controlled for this and found evidence for listener's sensitivity to the speaker's knowledge and how it differed from their own.

in the common ground (and hence was known, from the experimenter's perspective). In the other half, communication breakdown between participant and experimenter prevented the grounding of that referent. Optimal interpretation of the question required participants to take into account the experimenter's perspective by honing in early on the referent that had *not* been grounded (i.e., was not available to the experimenter). This was measured using eye-tracking. Participants with lower scores on a verbal measure of inhibition control were less able to reduce the activation of a competitor referent that had been verbally established in the common ground. This indicates that inhibition of one's own perspective for the purpose of referent identification is cognitively costly. Importantly, however, participants relied on perspective as a cue in contexts that lacked global ambiguity, suggesting that adult on-line processing is not egocentric by default (*contra* Keysar et al., 2000).

In a series of experiments using the visual-world paradigm, Grodner et al. (2012) investigated the extent to which inhibition and social abilities contribute to perspective use in adults. Participants were instructed (by a naive confederate) to pick up a target object in a cubbyhole display in which three objects were visible to the participant only and three visible to both the participant and the confederate (sitting on the other side of the display). In control trials, there were no identical objects in the visual common ground. In the competitor trials, two of these objects were identical. Participants' eye-movements were recorded. Inhibition abilities were found to predict performance in the first half of the experiments involving a perspective-taking task (especially if the cognitive load was increased by competing demands). In the first half of trials only, egocentric perspective (measured as the number of fixations to the competitor, minus the number of fixations to the control) was negatively correlated with inhibitory control (measured by the Simon task). However, this effect was only found in experiments in which the common ground was visually determined (rather than discourse-based). The first half of such experiments was interpreted as a "learning" phase, in which the participants learn which cues in the environment correspond to perspective. In the second half (but not the first), social aptitude (measured by the Autism Spectrum Quotient — Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001) was negatively correlated with egocentricity. This was taken to reflect the fact that, once the critical cues for perspective were reliably extracted, social aptitude reliably predicted perspective-taking. Additional cognitive load (i.e., remembering a 7-digit sequence during either the first or the second half of the experiment) increased egocentricity in the first but not the second half.

Based on these findings, apparent egocentricity in adults does seem to stem from the cognitive cost of inhibiting their own perspective. However this cost is "affordable" for adults: (i) it does not result in a significant number of interpretation errors (e.g., in Grodner et al's experiments, the participants never fixated the competitor more than the target, and never picked up the competitor) and (ii) its effect is only reliably observed in complex tasks (e.g., in the learning phase of a task, or in the presence of competing cognitive demands).

2.4 Summary

Evaluating the accessibility and salience of referents requires weighing a number of linguistic and contextual cues (as modelled e.g., by Trueswell et al., 2011) and taking one's interlocutor's perspective into account. This evaluation is not straightforward, as the weighting of the cues might be variable (depending on the purpose of the exchange, situational variables, etc), and perspective-taking is cognitively demanding even for adults.

3 Prerequisites to the encoding of reference

When linguists subject young children to tasks investigating their ability to encode information (e.g., their use of definites vs. indefinites, or pronouns vs. full noun phrases, etc), they usually assume that the prerequisite cognitive abilities are in place. This section attempts to spell out what these cognitive abilities are (leaving out the most basic ones such as responding to joint visual attention — see (Moll & Kadipasaoglu, 2013; Mundy et al., 2007)).[5] Their development will be considered in section 4.

As the most fundamental prerequisite, the child needs to be able to engage with the task at hand. Typically, this requires 'reading' a picture-based story, or in a more constrained task, listening to a prompt describing a context-setting scene in order to answer a follow-up question (which may also require processing visual information) — see section 4.3. The child is expected to focus on the relevant stimulus and ignore distractors — whether internal or external to the task. This requires attentional resources (see section 4.1).

The information gleaned by the child needs to be organised into a coherent representation of the situation or unfolding events. In most experimental designs, this requires using visual information. If adult-like, the resulting representation can be viewed as a "situation model [that] 'mediate[s] between the language and the discourse world [...] whether or not there is a real or a visual world to which the expressions in the language refer. [...] In the context of a visual world, [the model] acts as an interface between the language and a visual scene" (Garrod, 2011, p. 290). In adults, the visual world captures attention differently to non-visual contexts: "attending to specific referents in the visual scene may affect the accessibility of alternative antecedents referents" (Garrod, 2011, p. 290). I will argue below that this is exacerbated in young children (see section 4.3).

In order to build an accurate situational model, the child has to be able to recognise characters when they reappear, which requires having identified them as individuals in the context. More generally, children also need to keep track of contextual information (i.e., where things take place, what objects are used and which ones are visible but not used, etc). This will require integrating visual and verbal information (see section 4.4). Furthermore, the situational model needs to be held in mind and updated, which requires working memory and attention control (see section 4.1). These cognitive abilities are known to be inter-related in adults (Garon, Bryson, & Smith, 2008, p. 40).

[5] I focus here on the cognitive prerequisites exclusively. For an overview of the linguistic competence required for successful encoding of topics and given vs. new information, see De Cat (2009, 2011).

The exchange of information is assumed to rest on the evaluation of what is available to the addressee (e.g., by virtue of being part of the common ground, or being salient in the context). Referent accessibility is related to the focus of attention, under the assumption that it is shared between interlocutors (see section 4.1). Building on this, what is required is sustained awareness and monitoring of differences in the addressee's perspective. However, as discussed in section 2.2, this process is not automatic, even in adults, and has been shown to be cognitively costly in terms of executive function and attentional resources — especially in children (see section 4.1).

Summing up, in addition to the relevant linguistic competence, reference encoding (and decoding) requires (i) attentional resources, (ii) working memory, (iii) the ability to *distinguish* one's perspective from that of one's addressee, (iv) the ability to *inhibit* one's perspective, (v) the ability to integrate visual and verbal information into a coherent situational model, and (vi) the ability to maintain and update that situational model. And these cognitive resources need to be sufficiently integrated for the child to be able to make use of them adequately. The next section is a review of the (cognitive and linguistic) developmental literature aiming to sketch the developmental trajectory of the above cognitive resources up to the end of preschool years.

4 Cognitive development

4.1 Executive function in the preschool years

It is unclear whether EF organisation is identical in young children (before 8 years of age) and adults (see Garon et al., 2008 for a review). Individual components have been shown to emerge within the first 3 years of life (see below). However, emerging skills are only progressively integrated into higher order systems, and this development is not linear: while EF develops throughout early childhood, a critical phase occurs from 3 to 5 years of age, with abrupt changes affecting complex skills that involve the coordination of simpler skills (Garon et al., 2008). Between 3 and 4 years, EF becomes increasingly coherent (as indicated by the increased correlation between EF tasks). Garon et al. (2008, p. 51) argue that this critical development is enabled by "development of the attention system and its connectivity with other brain areas underlying component EFs".

The maturation of the **attentional capacity** appears to be essential to the development of EF abilities in preschoolers, as it enables the resolution of conflicts during information processing, thereby affording children an increasing amount of control over their thoughts and behaviours (Rothbart & Posner, 2001). Between the ages of 2 and 6, the anterior attention system (which selects which information to process) becomes progressively more influential than the orienting attention system (which shifts attention to other internal or external information). Sustained attention improves dramatically between 3 and 5 years of age (Garon et al., 2008, p. 38). In parallel with the maturation of the attentional system, EF undergoes a growth spurt from 3 to 6 years of age.

Working memory is the first EF component to develop (Garon et al., 2008, p. 49).

Simple retention of information is mastered earlier than the updating of information held in mind, which continues developing during the preschool years (Gathercole, 1998).

Response inhibition has been extensively researched in preschoolers. Simple response inhibition is the ability to suppress a dominant response (e.g., a rewarding behaviour). Complex response inhibition requires holding an arbitrary rule in mind, responding according to this rule, and inhibiting a dominant response. Both types of response inhibition undergo development during the preschool years, with significant improvement in complex inhibition between 3 and 5 years of age, and with many children younger than 6 failing complex tasks (see Garon et al., 2008 for a review). However, performance improves if "hot" aspects of EF are reduced — i.e., cognitive and emotional factors, such as memory costs or the way rewards are presented (Hongwanishkul, Happaney, Lee, & Zelazo, 2005; Metcalfe & Mischel, 1999).

Set shifting is the most complex EF component. It consists in the ability to focus away from an original mental set (in which a particular stimulus and a response are associated) and onto another, conflicting mental set. The conflict can apply to (i) the perceptual stage or (ii) the response stage. In the first case, what is required is *attention shifting*, where attention needs to focus on different aspects of the stimulus. In the second case, what is required is *response shifting*, where the motor response to the stimulus has to change. Attention shifting develops later than response shifting; the ability to resolve conflict develops until approximately 6 years of age (Rueda, Posner, & Rothbart, 2005).

In summary, the preschool years (between 2 and 5 years of age) are pivotal in terms of executive function development — not just for the development of the individual components, but also for their integration into an efficient system.

4.2 Perspective taking, Theory of Mind, and Executive Function

Perspective taking has been studied in several domains: visual perspective (Piaget & Inhelder, 1948), emotional perspective (Flavell, Botkin, Fry, Wright, & Jarvis, 1968), cognitive perspective (Marvin, Greenberg, & Mossler, 1976) — see Perner (2000) for a review. In their seminal study of visual-attentional acts and abilities, Lempers, Flavell, and Flavell (1977) demonstrate that one-year-olds can already understand what others can(not) see; they also can gesture appropriately to provide information to knowledgeable vs. ignorant partners (Liszkowski, Carpenter, & Tomasello, 2008). From age 2, they are able to overcome egocentricity when orienting pictures for the benefit of another person and are able to take others' needs into consideration (Grosse, Moll, & Tomasello, 2010). By 2;6, they understand the role of others' eyes in seeing and are sensitive to to other people's visual or auditory perspective (Moll & Meltzoff, 2011; Moll, Carpenter, & Tomasello, 2014). Hiding ability is established by age 3. The understanding of visual perspective is however not sufficient for the understanding of cognitive perspective, and in particular Theory of Mind.

Theory of Mind (ToM) is the cognitive ability to attribute mental states (such as beliefs, desires,...) to others. A large body of research has established that, in spite of signs of earlier sensitivity to others' beliefs (Freeman, Lewis, & Doherty, 1991), children really start understanding belief around 4 years of age (usually measured by False Belief tasks —

Hogrefe, Wimmer, & Perner, 1986; Wimmer & Perner, 1983). Second-order ToM, i.e., the ability to reason about *other people's* mental states concerning other mental states, emerges even later, around age 6.

Neurological evidence (Aichhorn, Perner, Kronbichler, Staffen, & Ladurner, 2006) suggests that visual perspective tasks require only a specific part of ToM: that which relates to so-called *cold* facts about the mind, i.e., representing perspective and making behavioural predictions. *Hot* facts about the mind (such as the anticipation of emotional consequences of a predicted behaviour) are computed elsewhere in the brain.

Children's perspective-taking abilities have been found to be related to their executive function development, and in particular their performance on conflict inhibition tasks (Nilsen & Graham, 2009). More generally, ToM seems to be associated in complex ways with executive function — and with response inhibition in particular (Carlson & Moses, 2001; Henning, Spinath, & Aschersleben, 2010; Hughes & Ensor, 2007; Moses & Tahiroglu, 2010). In the current debate as to which of EF or ToM is a prerequisite for the other, evidence appears more supportive of the view that EF is required not only for successful *expression* of ToM abilities, but probably also for its *emergence*. In particular (and most relevant to the purpose of the present chapter), EF appears to play a special role in the understanding of epistemic states (Henning et al., 2010). And the reflective capacity to distance oneself from the immediate perceptual context requires being able to suppress one's own perspective, which requires some level of EF (Carlson, Claxton, & Moses, 2015).

Some degree of sensitivity to the interlocutor's perspective has been reported in 3-year-olds' productions (Matthews, Lieven, Theakston, & Tomasello, 2006), along with a significant improvement by age 4 (see also Salomo, Graf, Lieven, & Tomasello, 2011). In comprehension, 3- to 4-year-olds show some ability to inhibit their own perspective, proportionally to their inhibitory control skills (Nilsen & Graham, 2009). While at age 4 and 5, children are still reported to experience difficulties aligning their referential domain with that of their addressee (Trueswell et al., 2011), sensitivity to their interlocutor's perspective, including sensitivity to common-ground information from the very initial stages of real-time processing, has been demonstrated in 5- to 6-year-olds (Nadig & Sedivy, 2002).

Matthews et al. (2006) demonstrate a positive effect of training on 2-, 3- and 4-year-olds' performance in a task that required them to uniquely identify an object in a display so it could be picked by an ignorant addressee. The most efficient training consisted in providing feedback to under-informative directions provided by the child. Crucially for the present chapter, the training effect only translated into improved performance in a different communicative task in children above the age of 4. This suggests that training is only fully effective when the relevant cognitive underpinnings are in place.

4.3 Interpreting visual information

Tasks that rely on the use of pictures require the child to be able to understand the relationship between a pictorial representation and its referent. Without this understanding, the child would literally mistake appearance for reality and think that people and objects occur in multiple contexts (Doherty, 2009, p. 78). While the foundations for the under-

standing of picture-object relations are in place before 2 years (Ganea, Allen, Butler, Carey, & DeLoache, 2009), it is not before the age of 2;6 that children reliably understand that a pictured object can represent the actual object in reality (i.e., that an individual element in a picture can *correspond* to one in reality) — (DeLoache & Burns, 1994). Around age 3, children understand that representations correspond in systematic ways to their referents — i.e., that the toy chair in a model corresponds to the real chair in reality, etc (Newcombe & Huttenlocher, 2000). *Full* understanding of the representational relationship between a picture and its referent only emerges around 4 years of age, when children become able to report two interpretations of an ambiguous figure (Doherty & Wimmer, 2005) and to use a picture of a room as a map of that room (Blades & Cooke, 1994). Both require metarepresentation — which Doherty (2009) argues is a prerequisite for second-order Theory of Mind.

In addition to the above, the ability to grasp that multiple representations of the same referent do not equate to several ("real") instances of that referent but to different representations of the same referent (e.g., at different times in a narrative) is a prerequisite for the interpretation of visual narratives, which are often used to gauge children's referential abilities.

4.4 Integrating visual and verbal information

Narrative skills develop well into the school years. Temporal sequencing of events emerges between 3 and 5 years (Cattanach, 2008). Before that, young children cannot go beyond the description of objects and events. Causality and logical sequencing emerge from 5 years onward. Narrative cohesion emerges (cross-linguistically) around 6 years of age (Hickmann, 2000). Preschool children may therefore not be able to reliably interpret a sequence of pictures as a coherent sequence of events and may not be able to see continuity in the reappearance of a character on a different picture — especially if that character lacks salient characteristics that would make it unique, and if it does not fulfill a prominent role in the events depicted.

Preschoolers' assessment of the common ground seems to rely heavily on visual information. Experiments in which a child is asked to communicate information for the benefit of an addressee who cannot see the child's visual source of information (e.g., a picture book, scenes on a computer,...) have shown that preschoolers over-use definites (see De Cat, 2011 for a literature review) and demonstratives (e.g., Gundel, 2011). This may be due in part to a stronger reliance (compared with older children and adults) on the visual context as a source of information rather than an inability to evaluate their addressee's perspective.[6]

If the child is unable to *integrate* the knowledge that their addressee cannot see the prompt pictures with their assessment of the common ground, s/he is likely to treat visually available referents as sufficiently salient to be encoded as definite (i.e., unique in

[6]This is not inconsistent with research showing that, in their parsing of ambiguous structures, children initially rely more strongly on intrinsic lexical biases rather than the visual context (Weighall, 2008): the present discussion focuses on verbal information at the discourse level rather than at the lexical level — which does not require integration into a discourse model. See also Kidd, Stewart, and Serratrice (2011).

the context) even on first mention (in an elicitation task). S/he may also not be able to resolve conflicts between types of information or conflicts arising from differences in perspective. As argued by Trueswell et al. (2011, p. 100): "Children between the ages of 4 and 6 years understand that a definite NP (produced or heard) must apply maximally to the current referential domain, but this domain is skewed by their attentional state. Instances in which the child successfully computes the referential domain [...] arise from circumstances in which there is a clear and reliable indicator from the conversation about the shape of the current referential domain and the goal of the utterance."

4.5 Predictions

The major cognitive developments reported above are predicted to have a significant impact on children's performance in referential tasks (including, but not limited to, elicited narratives). Insightful research into young children's discourse competence will need to be able to disentangle linguistic competence from the cognitive mechanisms underpinning the implementation of that competence: "Errors" are predicted to persist long after the key linguistic distinctions are in place (e.g., definiteness, topichood,...). Three major areas are predicted to remain challenging for preschool children: the evaluation of salience, the maintenance of a discourse model, and the integrated representation of incompatible rules. I address them in turn, based on the arguments developed in preceding sections.

Children's evaluation of the salience of referents will tend to be poor in experimental designs using visual stimuli, because (i) they tend to rely on the visual context more than the discourse context for the establishment of the common ground and tend to assume joint attention from their interlocutor, and because (ii) the ability to distance themselves from the immediate perceptual context is demanding in terms of executive function abilities. Better performance is predicted in tasks that do not rely on visual stimulus (as there would be no visual cues to override verbal cues of salience), and in tasks requiring active monitoring of the interlocutor's perspective (especially if scaffolding is provided). Performance is predicted to be correlated with inhibition measures.

Children's ability to build and maintain a discourse model is predicted to be correlated with working memory and attention measures. Performance will be poorer in tasks requiring an elaborate discourse model (e.g., with many referents).

Young children's lack of an integrated representation of incompatible rules results in discrepancies between what they know and what they do (Zelazo, Qu, & Müller, 2005). Sensitivity to others' perspective is therefore likely to be evidenced earlier in comprehension tasks that do not require action (e.g., eye-tracking but not picture selection). As their cognitive control increases, especially toward the end of the preschool years, children are likely to do better in tasks in which the monitoring of their interlocutor's perspective is explicitly and repeatedly made relevant. Their ability to do so effectively is also predicted to correlate with their executive function abilities.

Compared with monolingual children, bilingual children have been claimed to be better able to ignore irrelevant information (Bialystok, 1986), especially where the relevant feature is perceptual, as opposed to semantic (Bialystok & Martin, 2004). They are better at

attentional control (Bialystok, 1999), and show a clear advantage in the resolution of conflicting, competing cues (Carlson & Meltzoff, 2008; Martin-Rhee & Bialystok, 2008). If the above predictions are correct, bilingual children should therefore be expected to perform better than monolingual children in referential tasks that require perspective monitoring (all other things being equal).

Further research is needed to determine the relative strength of the various cues children rely on to evaluate the degree of salience of referents, and to identify how this weighting of cues differs from that of adults.[7] In this chapter, I have proposed that, *for the purpose of evaluating the salience of referents*, children are over-reliant on the visual context, and in a way that is exacerbated in experimental tasks. This could have several explanations. If young children have trouble maintaining a discourse model (abstract by definition), the visual context would provide a more reliable source for common ground evaluation. This would predict a correlation with working memory measures (consistent with the findings of van Rij, 2012). An alternative explanation could be that preschool children have not yet fully *integrated* the verbal cues to referent salience and/or are not able to resolve conflicts between cues from different sources. Recall that Grodner et al. (2012) found a correlation between inhibition and 'egocentricity' only in the first half of their experiment, i.e., the learning phase: the cognitively demanding one. If the integration of perceptual cues and linguistic rules is demanding for young children, their performance in referential tasks should be correlated with their inhibition abilities, especially in the absence of visual cues.

5 Conclusion

Many studies have found that children over-use linguistic forms that imply accessibility of the referent to their addressee. Does this mean they rely on a different discourse representation to adults? I have argued that great care needs to be taken to distinguish linguistic competence from the cognitive mechanisms required for its implementation. It could be that children and adults rely on the same type of information and encode it in the same way, but with different weighting of cues and/or with different cognitive limitations. Ultimately, if children and adults rely on the same representation and processes, calibrating the cognitive demands of tasks so that they are equally challenging for children and adults should result in similar performance.

Acknowledgments

This research was partly funded by a grant from the Leverhulme Trust (RPG-2012-633), which is gratefully acknowledged. Many thanks to the editors and the anonymous reviewer for their helpful suggestions.

[7] This could be implemented in a variety of approaches, such as for instance Trueswell et al.'s (2011) probabilistic processing model, Pickering and Garrod's (2004) interactive alignment model, or Optimality Theory (Hendriks, de Hoop, Krämer, de Swart, & Zwarts, 2010; Hendriks, 2015 (in press)).

References

Aichhorn, M., Perner, J., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Do visual perspective tasks need theory of mind? *Neuroimage*, *30*, 1059-1068.

Almor, A., & Nair, V. A. (2007). The form of referential expressions in discourse. *Language and Linguistic Compass*, *1*, 84-99.

Ariel, M. (1990). *Accessing noun-phrase antecedents*. London: Routledge.

Arnold, J. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes*, *31*, 137-162.

Arnold, J. (2010). How speakers refer: The role of accessibility. *Language and Linguistic Compass*, *4*, 187-203.

Arnold, J., & Griffin, Z. (2007). The effect of additional characters on choice of referring expression: Everyone counts. *Journal of Memory and Language*, *56*, 521-536.

Arnold, J., & Tanenhaus, M. (2011). Disfluency effects in comprehension: How new information can become accessible. In N. Perlmutter & E. Gibson (Eds.), *The processing and acquisition of reference* (p. 197-217). Cambridge, MA: MIT Press.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient: Evidence from Asperger syndrome/high-functioning autism, males and females, scientists, and mathematicians. *Journal of Autism and Developmental Disorders*, *31*, 5-17.

Bialystok, E. (1986). Factors in the growth of linguistic awareness. *Child Development*, *57*, 498-510.

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

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

Blades, M., & Cooke, Z. (1994). Young children's ability to understand a model as a spatial representation. *Journal of Genetic Psychology*, *115*, 201-218.

Brown-Schmidt, S. (2009). The role of executive function in perspective-taking during on-line language comprehension. *Psychonomic Bulletin and Review*, *16*, 893-900.

Carlson, S., Claxton, L., & Moses, L. (2015). The relation between executive function and theory of mind is more than skin deep. *Journal of Cognition and Development*, *16*, 186-197.

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

Carlson, S., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, *72*, 1032-1053.

Cattanach, A. (2008). *Narrative approaches in play with children*. London: Jessica Kingsley Publishers.

Clark, H., & Marshall, C. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. Webber, & I. Sag (Eds.), *Elements of discourse understanding* (p. 10-63). Cambridge: CUP.

De Cat, C. (2009). Experimental evidence for preschoolers' mastery of "topic". *Language Acquisition*, *16*, 224-239.

De Cat, C. (2011). Information tracking and encoding in early L1: Linguistic competence vs. cognitive limitations. *Journal of Child Language*, *38*, 828-860.

De Cat, C. (2013). Egocentric definiteness errors and perspective evaluation in preschool children. *Journal of Pragmatics*, *56*, 58-69.

DeLoache, J., & Burns, N. (1994). Early understanding of the representational function of pictures. *Cognition*, *52*, 83-110.

Doherty, M. J. (2009). *Theory of mind. How children understand others' thoughts and feelings*. Hove, UK: Psychology Press.

Doherty, M. J., & Wimmer, M. C. (2005). Children's understanding of ambiguous figures: Which cognitive developments are necessary to experience reversal? *Cognitive Development*, *20*, 407-421.

Epley, N., Morewedge, C., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology*, *40*, 760-768.

Flavell, J., Botkin, P., Fry, C., Wright, J., & Jarvis, D. (1968). *The development of role-taking and communication skills in children*. New York: Wiley.

Freeman, N. H., Lewis, C., & Doherty, M. J. (1991). Preschoolers' grasp of a desire for knowledge in false-belief prediction: Practical intelligence and verbal report. *British Journal of Developmental Psychology*, *9*, 139-157.

Friedman, N., Miyake, A., Corley, R., Young, S., DeFries, J., & Hewitt, J. (2006). Not all executive functions are related to intelligence. *Psychological Science*, *17*, 172-179.

Fukumura, K. (2010). *Choosing referring expression* (PhD). University of Dundee.

Fukumura, K., & Van Gompel, R. (2011). The effects of animacy in the choice of referring expressions. *Language and Cognitive Processes*, *26*, 1472-1504.

Fukumura, K., & Van Gompel, R. (2012). Producing pronouns and definite noun phrases: Do speakers use the addressee's discourse model? *Cognitive Science*, *36*, 1289-1311.

Ganea, P. A., Allen, M. L., Butler, L., Carey, S., & DeLoache, J. S. (2009). Toddlers' referential understanding of pictures. *Journal of Experimental Child Psychology*, *104*, 283-295.

Garon, N., Bryson, S., & Smith, I. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, *134*, 31-60.

Garrod, S. (2011). Referential processing in monologue and dialogue with and without access to real-world referents. In N. Perlmutter & E. Gibson (Eds.), *The processing and acquisition of reference* (p. 273-293). Cambridge, MA: MIT Press.

Gathercole, S. (1998). The development of memory. *Journal of Child Psychology and Psychiatry*, *39*, 3-27.

Grodner, D., Dalini, M., Pearlstein-Levy, S., & Ward, A. (2012). *Factors that contribute to the use of perspective in referent identification* (Paper presented at the City University of New York conference on human sentence processing).

Grosse, G., Moll, H., & Tomasello, M. (2010). 21-month-olds understand the cooperative logic of requests. *Journal of Pragmatics*, *42*, 3377-3383.

Gundel, J. (2009). Children's use of referring expressions: What can it tell us about theory of mind? *Cognitive Critique*, *1*, 73-100.

Gundel, J. (2011). Child language, theory of mind, and the role of procedural markers in identifying referents of nominal expressions. In V. Escandell-Vidal, M. Leonetti, & A. Ahern (Eds.), *Procedural meaning: Problems and perspectives* (p. 205-231). Bingley: Emerald Group Publishing Limited.

Gundel, J., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, *69*, 274-307.

Heller, D., Grodner, D., & Tanenhaus, M. K. (2009). The real-time use of information about common ground in restricting domains of reference. In U. Sauerland & K. Yatsushiro (Eds.), *Semantics and pragmatics. From experiment to theory* (p. 228-248).

Hendriks, P. (2015 (in press)). Optimal acquisition: Why children's language production can exceed their comprehension. *Proceedings of GALA 2013*,, .

Hendriks, P., de Hoop, H., Krämer, I., de Swart, H., & Zwarts, J. (2010). *Conflicts in interpretation*. London: Equinox.

Henning, A., Spinath, F., & Aschersleben, G. (2010). The link between preschoolers' executive function and theory of mind and the role of epistemic states. *Journal of Experimental Child Psychology*, *108*, 513-531.

Hickmann, M. (2000). Le développement de l'organisation discursive. In M. Kail & M. Fayol (Eds.), *Acquisition du langage volume 2: Le langage en développement* (Vol. 2, p. 83-115). Paris: PUF.

Hogrefe, G., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in attribution of epistemic states. *Child Development*, *57*, 567-582.

Hongwanishkul, D., Happaney, K. R., Lee, W., & Zelazo, P. D. (2005). Hot and cool executive function: Age-related changes and individual differences. *Developmental Neuropsychology*, *28*, 617-644.

Hughes, C., & Ensor, R. (2007). Executive function and theory of mind: Predictive relations from ages 2 to 4. *Developmental Psychology*, *43*, 1447-1459.

Karmiloff-Smith, A. (1979). *A functional approach to child language. A study of determiners and reference*. Cambridge: CUP.

Karmiloff-Smith, A. (1985). Language and cognitive process from a developmental perspective. *Language and Cognitive Processes*, *1*, 65-85.

Keysar, B., Barr, D., Balin, J., & Bruauner, J. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*, 32-38.

Keysar, B., Lin, S., & Barr, D. (2003). Limits on theory of mind use in adults. *Cognition*, *89*, 25-41.

Kidd, E., Stewart, A., & Serratrice, L. (2011). Children do not overcome lexical biases where adults do: The role of the referential scene in garden-path recovery. *Journal of Child Language*, *38*, 222-234.

Lempers, J., Flavell, E., & Flavell, J. (1977). The development in very young children of tacit knowledge concerning visual perception. *Genetic Psychology Monographs*, *95*, 3-53.

Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition*, *108*, 732-739.

Martin-Rhee, M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *11*, 81-93.

Marvin, R., Greenberg, M., & Mossler, D. (1976). The early development of conceptual perspective taking: Distinguishing among multiple perspectives. *Child Development*, *47*, .

Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2006). The effect of perceptual availability and prior discourse on young children's use of referring expressions. *Applied Psycholinguistics*, *27*, 403-422.

Metcalfe, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: Dynamics of willpower. *Psychological Review*, *106*, 3-19.

Miyake, A., Friedman, N., Emerson, M., Witzki, A., Howerter, A., & Wager, T. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49-100.

Moll, H., Carpenter, M., & Tomasello, M. (2014). Two- and 3-year-olds know what others have and have not heard. *Journal of Cognition and Development*, *15*, 12-21.

Moll, H., & Kadipasaoglu, D. (2013). The primacy of social over visual perspective-taking. *Frontiers in Human Neuroscience*, *7*, article 588.

Moll, H., & Meltzoff, A. N. (2011). How does it look? Level 2 perspective-taking at 36 months of age. *Child Development*, *82*, 661-673.

Moses, L., & Tahiroglu, D. (2010). Clarifying the relation between executive function and children's theories of mind. In J. Carpendale, G. Iarocci, U. Mueller, B. Sokol, & A. Young (Eds.), *Self- and social-regulation: Exploring the relations between social interaction, social cognition, and the development of executive functions* (p. 218-233). Oxford: OUP.

Mundy, P., Block, J., Delgado, C., Pomares, Y., Hecke, A. V. V., & Parlade, M. V. (2007). Individual differences and the development of joint attention in infancy. *Child Development*, *78*, 938-954.

Nadig, A., & Sedivy, J. C. (2002). Evidence of perspective-taking constratins in children's on-ine reference resolution. *Psychological Science*, *13*, 329-336.

Newcombe, N., & Huttenlocher, J. (2000). *Making space: The development of spatial representation and reasoning.* Cambridge, MA: MIT Press.

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

Perner, J. (2000). Theory of mind. In M. Bennett (Ed.), *Developmental psychology: Achievements and prospects.* Hove, East Sussex: Psychology Press.

Piaget, J., & Inhelder, B. (1948). *The child's conception of space.* London: Routledge and Kegan Paul.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*, 169-226.

Reinhart, T. (1981). Pragmatics and linguistics: An analysis of sentence topics. *Philosophica*, *27*, 53-94.

Rothbart, M., & Posner, M. (2001). Mechanism and variation in the development of attentional networks. In C. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (p. 353-363). Cambridge: CUP.

Rueda, M. R., Posner, M. I., & Rothbart, M. K. (2005). The development of executive attention: Contributions to the emergence of self-regulation. *Developmental Neuropsychology*, *28*, 573-594.

Salomo, D., Graf, E., Lieven, E., & Tomasello, M. (2011). The role of perceptual availability and discourse context in young children's question answering. *Journal of child language*, *38*, 918-931.

Schaeffer, J., & Matthewson, L. (2005). Grammar and pragmatics in the acquisition of article systems. *Natural language and Linguistic Theory*, *23*, 53-101.

Schafer, R., & de Villiers, J. (2000). Imagining articles: What *a* and *the* can tell us about the emergence of DP. In S. C. Howell, S. Fish, & T. Keith-Lucas (Eds.), *Proceedings of BUCLD 24* (p. 609-620). Sommerville, MA: Cascadilla Press.

Stalnaker, R. (1974). Pragmatic presuppositions. In M. K. Munitz & P. K. Unger (Eds.), *Semantics and philosophy* (p. 197-213). New York: New York University Press.

Stalnaker, R. (1978). Assertion. In P. Cole (Ed.), *Pragmatics: Syntax and semantics.* New York: Academic Press.

Trueswell, J. C., Papafragou, A., & Choi, Y. (2011). Referential and syntactic processes: What develops? In N. Perlmutter & E. Gibson (Eds.), *The processing and acquisition of reference* (p. 65-108). Cambridge, MA: MIT Press.

van Rij, J. (2012). *Pronoun processing: Computational, behavioral, and psychophysiological studies in children and adults* (Vol. 108). Groningen: Groningen Dissertations in Linguistics.

Vogels, J., Krahmer, E., & Maes, A. (2013a). Cognitive load does not decrease pronoun use when speaker's and addressee's perspectives are dissociated. In *Proceedings of COGSCI 2013* (p. 3681-3868).

Vogels, J., Krahmer, E., & Maes, A. (2013b). Who is where referred to how, and why? the influence of visual saliency on referent accessibility in spoken language production. *Language and Cognitive Processes*, *28*, 1323-1349.

Weighall, A. R. (2008). The kindergarten path effect revisited: Children's use of context in processing structural ambiguities. *Journal of Experimental Child Psychology*, *99*, 75-95.

Wimmer, & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103-128.

Zelazo, P. D., Qu, L., & Müller, U. (2005). Hot and cool aspects of executive function: Relations in early development. In W. Schneider, R. Schumann-Hengsteler, & B. Sodian (Eds.), *Young children's cognitive development: Interrelationships among executive functioning, working memory, verbal ability, and theory of mind* (p. 71-93). Mahwah, NJ: Erlbaum.