MIT Libraries | DSpace@MIT

MIT Open Access Articles

Pragmatic markers: the missing link between language and Theory of Mind

Massachusetts Institute of Technology

THE CULTURAL EVOLUTION OF HUMAN SOCIAL COGNITION

Pragmatic markers: the missing link between language and Theory of Mind

Paula Rubio-Fernandez[1,2]

Abstract

Language and Theory of Mind come together in communication, but their relationship has been intensely contested. I hypothesize that pragmatic markers connect language and Theory of Mind and enable their co-development and co-evolution through a positive feedback loop, whereby the development of one skill boosts the development of the other. I propose to test this hypothesis by investigating two types of pragmatic markers: demonstratives (e.g., 'this' vs. 'that' in English) and articles (e.g., 'a' vs. 'the'). Pragmatic markers are closed-class words that encode non-representational information that is unavailable to consciousness, but accessed automatically in processing. These markers have been associated with implicit Theory of Mind because they are used to establish joint attention (e.g., 'I prefer that one') and mark shared knowledge (e.g., 'We bought the house' vs. 'We bought a house'). Here I develop a theoretical account of how joint attention (as driven by the use of demonstratives) is the basis for children's later tracking of common ground (as marked by definite articles). The developmental path from joint attention to common ground parallels language change, with demonstrative forms giving rise to definite articles. This parallel opens the possibility of modelling the emergence of Theory of Mind in human development in tandem with its routinization across language communities and generations of speakers. I therefore propose that, in order to understand the relationship between language and Theory of Mind, we should study pragmatics at three parallel timescales: during language acquisition, language use, and language change.

Keywords Theory of Mind · Pragmatics · Demonstratives · Definite articles · Joint attention · Common ground · Language change

✉ Paula Rubio-Fernandez
paula.rubio-fernandez@ifikk.uio.no

[1] Department of Philosophy, Classics, History of Art and Ideas, University of Oslo, Oslo, Norway

[2] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, USA

Springer

1 Introduction

Mastering communication takes more than mastering a language. Imagine that you and your partner are going to the theatre: you may say 'I forgot the tickets' to imply that you need to go back home. This example illustrates how, as speakers, we trust our listeners to read between the lines, and as listeners, we are willing to go beyond the literal to infer what the speaker intended to convey (Grice 1975). Theoretical work on the nature of communication has long argued that communication requires a *Theory of Mind*: an ability to reason about other people's mental states, such as their beliefs and intentions (Sperber and Wilson 1986; Levinson 2006; Tomasello 2008; Scott-Phillips 2014). In this paper, I will discuss the relationship between language and Theory of Mind and put forward the new hypothesis that pragmatic markers are a linchpin for Theory of Mind. Going back to our example, by saying 'the tickets', you would be signaling that these tickets are in your common ground with your partner—otherwise they would respond 'which tickets?'

A critical question regarding the connection between language and Theory of Mind is whether children's Theory of Mind development is dependent on their language abilities. Developmental studies have indeed shown a correlation between language and Theory of Mind (for a meta-analysis, see Milligan et al. 2007). Correlational studies normally use syntax and vocabulary scores as measures of linguistic ability, while Theory of Mind is assessed through *false-belief tasks*: a classic paradigm where a protagonist is mistaken about the location of an object, for example, and the child has to predict where the protagonist will look for the object, without defaulting to their own knowledge (Wimmer and Perner 1983). From a theoretical perspective, it has been proposed that false-belief understanding (as measured by standard tasks) emerges from children's mastery of sentential complement syntax (de Villiers 1999, 2007; cf. Hacquard and Lidz 2019). Parallel to the distinction between the child's true belief and the protagonist's false belief in a Theory of Mind task, understanding 'Sally thinks that the marble is in the box' requires appreciating that the sentence may be true even though the marble is in the basket. Supporting the view that complement syntax is related to false-belief understanding, developmental studies have shown that training children on subordinate clauses improves their performance in Theory of Mind tasks (Lohmann and Tomasello 2003).

Syntax-based accounts of the relationship between language and Theory of Mind suffer from two limitations: first, their assessment of Theory of Mind is confined to false-belief tasks, failing to account for more basic forms of Theory of Mind. Second, by focusing on sentential complement syntax, they also fail to account for other grammatical elements that require perspective taking (e.g., the use of pronouns). As a result, syntax-based accounts leave three fundamental questions unanswered, each related to a distinct timescale in the study of human language:

1. Regarding language acquisition, children do not pass standard false-belief tasks (or acquire complement clause syntax) before their 4th birthday (Wimmer and Perner 1983; Rakoczy 2017), so how does Theory of Mind develop up until age 4?

2. Regarding language use, once sentential complement syntax has been mastered, how do proficient speakers use their Theory of Mind in everyday communication?

3. Regarding language evolution, not all languages express mental states via subordinate clauses (Mithun 1984; Evans 2006a), so how did Theory of Mind emerge across languages and cultures?

While generally sympathetic to syntax-based accounts, in this paper I will propose to address these open questions through the study of pragmatic markers: a functional class of linguistic devices that structure discourse and mark *intersubjectivity* (i.e. the speaker's assumptions about the degree to which the listener shares their attention or knowledge). I hypothesize that pragmatic markers connect language and Theory of Mind and enable their co-development in ontogeny and co-evolution in diachrony and phylogeny through a *positive feedback loop*, whereby the development of one skill boosts the development of the other. To test this new hypothesis, I propose to investigate children's acquisition and adults' use of two kinds of pragmatic markers: demonstratives (e.g., 'this' vs. 'that' in English) and articles (e.g., 'a' vs. 'the'); as well as their cultural evolution (i.e. their diachronic change through processes of learning and use).

Demonstratives and articles are closed-class words that encode *procedural meanings*: non-representational information that is unavailable to consciousness and therefore implicit, but accessed automatically during processing (Blakemore 1987). This explains why a competent user of English would understand that 'We bought the house' refers to a familiar house, but would find it difficult to define the meaning of 'the' (Gundel and Johnson 2013). By contrast, conceptual meanings are conveyed by open-class words (such as nouns and verbs), which encode information that is representational and explicit, and therefore more accessible to introspection, but less automatic. The distinction between procedural versus conceptual meanings has been linked to that between implicit versus explicit Theory of Mind (Gundel et al. 2007). For example, Japanese encodes certainty and evidentiality in high-frequency, closed-class sentence-final particles (e.g., 'tte' marks hearsay), as well as in low-frequency mental state verbs (e.g., 'shitteru', to know). Matsui et al. (2006) showed that 3- to 6-year-old Japanese-speaking children understand the epistemic information encoded in sentence-final particles before they understand mental state verbs. Moreover, children's epistemic vocabulary correlated with their performance in standard false-belief tasks, whereas their understanding of sentence-final particles expressing the same meanings did not. Matsui et al. concluded that Japanese children's understanding of speakers' epistemic states as communicated by sentence-final particles paves the way for their later, fully-representational understanding of belief.

By focusing on procedural meanings, I will construe Theory of Mind broadly: as a form of social cognition that comprises not only belief understanding, but also more basic skills such as monitoring other people's attention or keeping track of shared knowledge (both of which involve some understanding of mental states and are recruited in communication). My proposal will therefore have a wider scope than previous work on the relationship between language and Theory of Mind, which has mainly focused on children's understanding of belief (see Tompkins et al. 2019). This also means that the present account does not hinge on the ongoing debate in the

Theory of Mind literature about whether the concept of belief is innate or develops during childhood (Onishi and Baillargeon 2005; Heyes 2014). In fact, this work should be relevant to both nativist and developmental accounts of belief since both need to explain how children *learn to use* Theory of Mind in interaction. By moving away from discussions of belief nativism, I will focus on communication as the natural arena for Theory of Mind development (Rubio-Fernandez 2017, 2019; Rubio-Fernandez et al. 2019).

2 The three timescales of evolutionary pragmatics

A growing body of work in cognitive science defends that human language is a learned product of cultural evolution, rather than being biologically endowed (Christiansen and Kirby 2003; Beckner et al. 2009; Evans and Levinson 2009; Heyes 2018; Smith 2018). In this view, language is a cultural artefact, together with our concepts, counting systems and social institutions, all of which change over historical time shaped by human interaction (Dediu et al. 2013; Christiansen and Chater 2016). While not committed to any particular view of the origin of Theory of Mind in human phylogeny or ontogeny, here I will adopt the cultural evolution view of language with a focus on the acquisition of pragmatics, and argue that children develop their Theory of Mind in the process of acquiring and using language. I will further propose that in order to understand the relationship between language and Theory of Mind, we must approach pragmatics from three parallel timescales: during language acquisition, language use, and language change.

These timescales have been previously used to investigate the origins of human language as a product of cultural evolution (i.e. through processes of learning and use; Kirby et al. 2008, 2014; Fedzechkina et al. 2012, 2017; Dediu et al. 2013; Culbertson and Adger 2014; Christiansen and Chater 2016). By adopting the same multi-scale approach, I propose to open a new research field within cultural evolution research: *evolutionary pragmatics*. Interestingly, even those researchers who defend the cultural evolution view and reject nativist accounts of human language (e.g., the idea that humans are endowed with a Universal Grammar; Chomsky 1965) nonetheless assume that the Theory of Mind abilities involved in human communication are innate (e.g., Tomasello et al. 2005; Levinson 2006; Scott-Phillips 2014). Similarly, Heyes and Frith (2014) have recently proposed that explicit Theory of Mind (as measured by false-belief tasks) is a learned, culturally inherited skill, but infants are endowed with an implicit Theory of Mind (see also Tomasello 2018). In my view, these accounts may be challenged on two grounds: first, none of them have systematically explored the possibility that Theory of Mind and language may have co-evolved (cf. Malle 2002; Woensdregt et al. 2020; Moore, under review), although it is generally agreed that language must play a role in Theory of Mind development. Second, even if we assume that Theory of Mind is innate, we still need to explain how children learn to use these early skills in communication (a process that takes years of maturation and has not been fully explained).

According to the positive feedback loop hypothesis, language acquisition boosts Theory of Mind development, and vice versa. For example, acquiring the seman-

tic meaning of 'here' and 'there' in English requires learning that these words encode relative distance from the speaker and are contrastive. The pragmatics of these demonstratives, however, require perspective taking: when used in a specific context, what is 'here' for the speaker may be 'there' for the listener, and vice versa. Since children acquire language through exposure and use, the process whereby young children acquire demonstratives like 'here' and 'there' requires that they develop their perspective-taking skills as part of the same process. In putting forward this view, I am not assuming that either language or Theory of Mind are prior, and will focus instead on their interdependency during human development.

A reader with a nativist incline may argue that if acquiring the meaning of 'here' and 'there' requires perspective taking, that presupposes that the young child must have a Theory of Mind to learn these words in the first place. Such a counterargument, however, obviates an important insight: children make mistakes that reveal insufficient perspective taking when learning demonstratives and other pragmatic markers (for a discussion of children's perspectival errors with 'here' and 'there', see Clark (1978), Clark and Sengul (1978), and Sect. 5.3 below). In fact, one of the best attested errors in the language acquisition literature are young children's *pronoun reversals*: their use of pronouns 'I' and 'you' to mistakenly refer to the listener and the speaker, respectively (e.g., Mom: 'I'm going to get you Teddy now, and you're going to sleep'; Child: 'No, you don't wanna sleep, I sleep!', pointing at the mother; Dale and Crain-Thoreson 1993; Loveland 1984). Young children's perspectival errors are an ideal illustration of the ways in which acquiring pragmatic markers can boost Theory of Mind development in a positive feedback loop—rather than Theory of Mind being a prerequisite for the acquisition and use of perspectival terms.

Previous studies on the relationship between language and Theory of Mind have relied on correlations between tasks that measure language and Theory of Mind separately (see Milligan et al. 2007). However, testing the positive feedback loop hypothesis would require that language and Theory of Mind be studied together, as they are *jointly used in communication*. Only such an investigation of language and Theory of Mind could reveal whether their joint use affects their acquisition in development and their change in cultural evolution, as predicted by the positive feedback loop hypothesis. This hypothesis therefore introduces a new way to understand the relationship between language and Theory of Mind as one of *co-dependence*: human language and Theory of Mind may have co-evolved in diachrony and phylogeny, and co-develop in ontogeny through the acquisition, use and cultural evolution of pragmatic markers.

While obvious on a moment's reflection, it may be worth noting that not all forms of Theory of Mind depend on the acquisition of pragmatic markers. For example, understanding the difference between 'Sally thinks that the marble is in the box' versus 'Sally knows that the marble is in the box' requires an understanding of factivity, as marked by mental state verbs 'think' versus 'know'. Likewise, coming to understand the connection between seeing and knowing, and developing a suitable heuristic (e.g., assume that if X has witnessed Y, X knows that Y) do not depend on the acquisition of pragmatic markers either. The positive feedback loop hypothesis is therefore intended to cover all instances of Theory of Mind development that could depend on (or benefit from) language acquisition and use, while leaving out of its scope those forms of

Theory of Mind development that may not depend on (or even benefit from) linguistic interaction—assuming there are any of the latter kind.

The advantage of using the acquisition, use and evolution of pragmatic markers as a testbed for the positive feedback loop hypothesis is that it offers a reasonable starting point for the co-evolution of language and Theory of Mind. Thus, rather than trying to speculate in an empirical vacuum about whether humans could have evolved languages without having a Theory of Mind, the starting point of my investigation will be the earliest linguistic form to require the use of Theory of Mind, both in diachrony and ontogeny; namely, demonstratives. The question of whether Theory of Mind emerged earlier than language (or whether human language could have emerged without a Theory of Mind) is beyond the scope of this proposal.

3 Aims, scope and working hypotheses

The aim of this paper is to put forward a new hypothesis about the relationship between language and Theory of Mind that could explain (1) the development of early forms of Theory of Mind through language acquisition, (2) their use and automatization in adult communication, and (3) their co-evolution with language in diachrony. The scope of the paper will not go beyond an outline for a large-scale research program, and therefore, all the issues discussed, as well as the details of the main proposal will need further theoretical refinement and empirical investigation. Tentatively then, I will start by putting forward three related hypotheses:

3.1 Hypothesis 1: Pragmatic markers in language acquisition

The acquisition of demonstratives (e.g., 'I want that cupcake'), which are often accompanied by a pointing gesture, builds on and buttresses young children's ability to engage in joint attention (i.e. sharing their focus of attention with others). Depending on the language, demonstratives may indicate not only the distance, but also the altitude, familiarity, position, reachability or visibility of a referent, from the perspective of the speaker, the listener, or both. Since demonstratives encode different relational values and require shifting perspectives, their acquisition should help the development not only of early joint attention, but also of later perspective-taking skills. This hypothesis lends itself to the prediction of cross-linguistic differences: the development of perspective taking follows different paths depending on the relational values and perspectives encoded in the demonstrative system(s) that the child is learning.

3.2 Hypothesis 2: Pragmatic markers in language use

Discourse demonstratives (e.g., 'John and Judy met in 1996. That was a good year.') and definite articles (e.g., 'We bought the house.') mark a more sophisticated form of common ground than gestural demonstratives: one that goes beyond the here-and-now and ranges over conversations and past shared experiences. Acquiring these pragmatic markers requires a broader, more abstract record of what is shared between interlocu-

tors, as well as greater memory capacity. I predict that the use of demonstratives and definite articles trains speakers in monitoring their interlocutor's attention and in managing common ground, resulting in the automatization of these processes over time, with potential cross-linguistic differences.

3.3 Hypothesis 3: Pragmatic markers in language change

Children's acquisition of the above pragmatic markers (ranging from demonstratives and pointing gestures to definite reference) reveal a developmental trajectory in Theory of Mind, which is instantiated not only in language acquisition but also in language change: the historical record shows that gestural demonstratives (or *exophoric demonstratives*, in linguistics jargon) give rise to discourse demonstratives (or *endophoric demonstratives*), which in turn give rise to definite articles. The parallels across language acquisition and language change open the possibility of modelling Theory of Mind development not only across childhood (as it has been done traditionally), but also across generations of speakers, driven by and in turn driving the evolution of pragmatic markers.

Testing these three hypotheses would require an ambitious experimental program of cross-linguistic research. As a modest first step, the remainder of this paper will focus on demonstratives, as the first pragmatic marker that children acquire across languages. The discussion will be divided in three parts. First, demonstratives will be characterized from a grammatical (Sect. 4.1), developmental (Sect. 4.2) and interactive (Sect. 4.3) perspective. The next part will include a review of cross-linguistic studies of demonstratives, also from three complementary perspectives: linguistic typology (Sect. 5.1), psycholinguistics (Sect. 5.2) and language acquisition (Sect. 5.3). The last part will focus on the evolution of demonstrative forms into definite articles and the implications of language change for Theory of Mind use and evolution. This last part will discuss the expansion of common ground (Sect. 6.1), the notion of *pragmatic relativity* (Sect. 6.2) and the power of procedural knowledge (Sect. 6.3).

4 Demonstratives: a universal tool for joint attention

4.1 From grammar to acquisition

Demonstratives are *deictic expressions*, also known as *directives* because they are primarily used to orient the listener's attention towards an element in the speech situation, normally one that was not currently in the listener's focus of attention (Diessel 1999, 2003; see Table 1 for the English demonstrative categories). It is because of their directive function that demonstratives are often used with a pointing gesture. Drawing on evidence from linguistic typology and historical linguistics, Diessel (2006, 2012a, b) has shown that demonstratives constitute a unique class of linguistic expressions that serve two closely related functions: (1) they indicate the location of a referent relative to the *deictic center* (e.g., the speaker's position in English), and (2) they coordinate

Table 1 Demonstrative categories in English

Category	Demonstratives	Example
Pronoun	that (one)/this (one)	*I prefer this one to that one*
Determiner	that/this	*I bought this coat in Paris*
Adverb	here/there	*Go over there and wait for me*

the interlocutors' joint focus of attention. The latter, Diessel argues, is one of the most basic functions of language, which explains many features of demonstratives.

In his introduction to a recent volume on demonstratives from a cross-linguistic perspective, Levinson (2018) also talks about the importance of demonstratives:

'[They are] a kind of ideal model system for the study of language use: a single word and gesture can function as a full referring act, with all the complexities of the joint attention, common ground, multimodality and pragmatic integration involved in more complex utterances' (p. 2).

If we understand pragmatics as the study of language as it is used in context, the relevance of demonstratives for pragmatics seems obvious from the above references. However, in order to propose that the acquisition of demonstratives and their grammaticalization into definite articles may be used to study not only developmental pragmatics, but also Theory of Mind development, a broader perspective must be adopted. For example: do all languages have demonstratives? And where do demonstratives come from, in terms of language evolution? Or thinking of language acquisition, at what age do children learn demonstratives, and when do they start using them like adults? Diessel (2006, 2012b, 2013) offers an exhaustive analysis of demonstratives that addresses all these questions:

1. Demonstratives are universal: they occur in all languages across the world (Levinson 2018).
2. Demonstratives are often accompanied by a pointing gesture, which is a universal communicative device that is used in all cultures to establish joint attention (Kita 2003).
3. Demonstratives emerge very early in language acquisition, being often the first non-content words that children learn together with their early use of pointing gestures (Clark 1978).
4. Demonstratives are so old that their roots are not etymologically analyzable. That is, the origins of demonstrative forms cannot be traced back to other types of expressions. This suggests that demonstratives emerged very early in the evolution of language, probably because of their basic communicative function to coordinate the interlocutors' joint attention (Diessel 2003).

Given their universal scope and their fundamental role in communication and language acquisition, it seems safe to assume that if there was a class of grammatical expressions linked to the emergence and development of Theory of Mind in humans, that would be demonstratives. It must be noted, however, that the connection between Theory of Mind and grammar is not limited to demonstrative expressions: Evans and colleagues have coined the term *grammar of engagement* to refer to those grammatical

means by which languages encode intersubjectivity (Evans 2006b; Evans et al. 2018a, b; compare, e.g., 'We bought a house' vs. 'We bought the house'). The proposal to study Theory of Mind development through the acquisition and use of demonstratives falls within the scope of Evans et al.'s grammar of engagement.

Demonstratives are acquired early on in development together with the use of pointing gestures to establish joint attention. These pragmatic markers are therefore a 'model case' for the study of early Theory of Mind in communication. Joint attention has been extensively studied in developmental psychology because of its fundamental role in language acquisition and communication (Baldwin 1995; Moore et al. 1995; Carpenter et al. 1998; Tomasello 1999): in order to communicate successfully, speakers and listeners must coordinate their focus of attention, for which the speaker may direct the listener's attention to an intended referent in the physical environment by using gaze, gesture and/or language (Diessel 2006). This ability does not emerge until the first year: infants' interaction with the world is at first dyadic, focusing their attention either on a person or an object, but not yet sharing their attention focus with another person. Children start engaging in triadic interactions at around 9 months, when they begin to follow another person's head movement and eye gaze, followed by their first pointing gestures at around 12 months, soon to be combined with the use of demonstratives. According to Clark (1978), the demonstratives 'this', 'that', 'here' and 'there' are amongst the first ten words that English-speaking children produce and are initially always accompanied by a pointing gesture. According to Diessel (2006, 2013), the early emergence of demonstratives is motivated by their communicative function and their relationship to deictic pointing: the combination of demonstratives and pointing gestures makes for a powerful expressive tool that allows the child to refer to any entity in their physical environment before they learn the corresponding word.

Toddlers' early productions of pointing gestures and demonstratives are one of the earliest manifestations of Theory of Mind use in human interaction. Reinforcing their connection to Theory of Mind development, demonstratives are often impaired in young children with Autism Spectrum Disorder (Friedman et al. 2019). Given their universal communicative function and cross-cultural significance, theoretical models of Theory of Mind development should account for the acquisition of demonstratives. For example: what Theory of Mind capacity is necessary in order to be able to engage in triadic interaction with others? Or to put it differently, what changes in the preverbal period between 6 and 12 months of age that enables the emergence of gaze following and deictic pointing? And further still: what role does the acquisition and use of demonstratives play in bootstrapping toddlers' Theory of Mind? It must be noted that, while the large majority of developmental research in Theory of Mind has focused on the emergence of false-belief understanding, none of these fundamental questions would be answered if a false-belief study convincingly showed that 12-month infants have a concept of belief. Therefore, all our theoretical and experimental efforts in understanding Theory of Mind development should be spread across the first years of life, and aim to explain not only false-belief understanding, but also the use of Theory of Mind in naturalistic interaction (see Shatz et al. 1983; Bartsch and Wellman 1995; Harris 1996, 1999).

4.2 Demonstratives in cognitive development

Moll and Meltzoff (2011a, b) have proposed a developmental trajectory in children's understanding of perspectives that starts in joint attention and peaks at false belief understanding, with young children going through three levels (and five stages within those three levels) between the ages of 1 and 4;6 years (see also Carpenter and Liebal 2011). At *Level 0 perspective-taking*, infants do not yet understand perspectives but can share them in joint attention. Between 12 and 18 months, infants reach *Level 1* experiential *perspective-taking* and become able to keep track of what others have experienced in joint attention with them. For example, Tomasello and Haberl (2003) had 1;0- and 1;6-year-old infants play with two objects together with one experimenter, and then play with a third object together with another experimenter. When the first experimenter returned and showed surprise, the infants understood that she was referring to the third object that they had not shared and handed it to her when she asked for 'it' (see also Moll and Tomasello 2006; Moll et al. 2007, 2008). While this early ability does not necessarily require understanding propositional knowledge, it does at the very least require monitoring what other people are familiar with from our shared experiences.

At about 2 years, children go from recognizing and monitoring others' attention to knowing what others can and cannot see from their viewpoint—what Moll and Meltzoff (2011a, b) refer to as *Level 1* visual *perspective-taking* (see Flavell 1992). A year later, this ability develops into *Level 2 perspective-taking*, whereby 3-year-olds are able to recognize how another person sees something, even if she sees it differently from how they see it. Moll and Meltzoff (2011c) presented 3-year-olds with two objects of the same color, while an experimenter in the room saw one of the objects through a tinted filter that changed its color. Even though the two objects looked blue to the children, when the adult asked for 'the green one', 3-year-olds systematically selected the object that looked green to the adult. Level 2 perspective-taking evolves a year or so later into an even more sophisticated ability: 4;6 year-old children are able not only to take, but also to *confront* different perspectives. Such an ability is required to pass standard false-belief tasks (Wimmer and Perner 1983), in which the child must inhibit their own knowledge of the situation and respond to the test question from the protagonist's perspective.

How does the acquisition of demonstratives and other pragmatic markers fit into this picture? The earliest uses of demonstratives, which tend to be accompanied by deictic pointing, would require Level 1 experiential perspective-taking, with later uses requiring Level 1 visual perspective-taking once the child starts monitoring the adult's focus of attention. Language acquisition studies have shown that young children show an earlier sensitivity to what is in the adult's focus of attention from what the adult has just said, than from what the adult has or has not seen (for a review, see Allen et al. 2015). Campbell et al. (2000), Matthews et al. (2006) and Rozendaal and Baker (2010) investigated the effect of prior mention and perceptual availability on young children's choice of referential expression and observed that prior mention already had an effect at 2 years (e.g., if asked 'What was the clown doing?', 2-year-old children were more likely to respond using the pronoun 'he' than if the question had not mentioned the

agent, as in 'What happened?'). However, it was not until age 3 that children started showing sensitivity to what the adult had or had not witnessed when describing an episode. Serratrice (2008) observed an even later use of visual perspective-taking, with 3-year-old children only showing sensitivity to prior mention (i.e. whether or not the subject had been made explicit in the question), while 5- and 6-year-olds revealed some sensitivity to perceptual availability, but were not yet able to integrate both cues at adult levels (e.g., 6-year-olds identified the subject unambiguously 60% of the time when their interlocutor was ignorant, whereas adults did so 97% of the time).

Relative to the results of Tomasello and Haberl (2003) and Moll and Tomasello (2006), Moll et al. (2007, 2008), Moll and Meltzoff (2011a, b), Level 1 perspective-taking seems to be observed earlier in behavioral studies (where toddlers in their first year have shown experiential perspective-taking) than in language acquisition studies (where perceptual availability does not affect children's verbal responses until age 3). This delay suggests that children are first able to track what is old or new for another person, before they can use that ability to inform their choice of referential expression. As Matthews et al. (2006) put it: 'Knowing that things can be given and new for other people in general and knowing how this is expressed in language are two different matters' (p. 419).

A common pattern observed in the language acquisition literature is that young children tend to omit referents and use pronouns for new or inaccessible referents (either perceptually or from previous discourse), resulting in ambiguous reference (Allen et al. 2015). However, Skarabella and Allan (2002) and Sakarabella 2007) observed that children aged 2;0–3;6 would omit referents and use demonstrative forms when the intended referent was in joint attention with their interlocutor, a tendency also observed in adults. Skarabella et al. (2013) further observed that these children's choice of demonstrative form was also informed by joint attention, with clitics being preferred in situations of joint attention, whereas full demonstratives were used when joint attention had not yet been established.

The results of language acquisition studies therefore suggest that joint attention is one of the earliest cues that young children rely on in their use of demonstratives and other pragmatic markers. Interestingly, a closer look at the pragmatics of demonstratives across different languages suggests that the mastery of demonstrative systems may require, depending on the grammar of engagement of the particular language (Evans et al. 2018a, b), up to Level 2 perspective-taking.

4.3 Demonstratives in interaction

Unlike content words, demonstratives and other deictic expressions establish a direct referential link between language and the world, rather than evoking a lexical concept (Diessel 2012b). Deictic expressions therefore rely strongly on pragmatics, since their use and interpretation are entirely determined by the context (e.g., the personal pronouns 'I' and 'you' refer to the speaker and the listener, respectively, but they pick up different referents during the course of a conversation). More importantly for the aim of this paper, the production and comprehension of deictic expression (including demonstratives) involves a particular viewpoint, or deictic center. The deictic center

is the zero-point in an evoked coordinate system (Hanks 2011): the pivot relative to which the referent is to be identified (e.g., in English, the difference in distance suggested by the utterances 'I prefer this one' versus 'I prefer that one' is established from the speaker's perspective).

The deictic center of a demonstrative does not always correspond with the speaker: languages like Japanese or Spanish have demonstratives that differentiate between referents near the speaker, referents near the listener, and referents away from both the speaker and the listener (Diessel 2012b; De Cock 2013).[1] Such a system requires shifting the deictic center when using different demonstrative forms (see Hanks 2011). Moreover, in addition to distance relative to the deictic origin, demonstratives may indicate whether the referent is visible or out of sight, at a higher or lower elevation, uphill or downhill, or in a particular location along the coastline (Diessel 1999, 2012b). Such *relational values* between the deictic center and the referent are also perspectival. However, Hanks (2011) notes that there are important relational values that are non-spatial and are also encoded directly in the semantics of demonstratives. Some of those relational values should be of interest to Theory of Mind researchers as they require monitoring the listener's focus of attention.

In an influential study on Turkish demonstratives, Özyürek (1998) showed that the forms 'bu' and 'o' seem to be used analogously to English 'this' and 'that', distinguishing entities that are close and far away from the speaker, respectively. However, the third demonstrative form 'şu' can be used to refer to objects at any distance from the speaker as long as joint attention has not yet been established. Evans et al. (2018a) gloss the Turkish deictic routine as follows: 'use a combination of pointing plus 'şu' until you have achieved mutual attention on the object at issue, then proceed by using 'bu' or 'o' according to the distance to the referent' (p. 18). This routine suggests that Turkish demonstratives encode interactive distinctions as part of their basic semantic meaning, with the form 'şu' serving two main functions that tap into social cognition, rather than spatial representations: (1) introducing a new referent in the discourse, and (2) directing the listener's attention to important referents in directives, questions and answers (Özyürek 1998). The interactive distinctions marked in the Turkish demonstrative system seem to require more sophisticated Theory of Mind abilities than simply establishing a referent's distance to the speaker.[2]

Along similar lines, Levinson (2018) points out that demonstrative systems encode *proximal* and *distal* zones, yet what counts as proximal and distal varies across languages and can be affected not just by physical distance but also (or rather) by interactive factors. Hanks (2011: p. 327) lists the following relational values encoded in the world demonstrative systems: *relative immediacy* (in space or time), *interiority* (inside, outside, lateral), *location* versus *trajectory*, *perception* (visual or other) and several varieties of *cognitive access* (e.g., reference to prior discourse, or relative salience). This is perhaps the kind of grammatical classification that would drive

---

[1] Some semantic analyses of the Spanish demonstrative system have characterized it as distance-based (e.g., Kemmerer 1999; Levinson 2004), while others situate it among person-oriented systems (e.g., Cifuentes-Honrubia 1989; Eguren 1999), or a combination of both (Jungbluth 2003).

[2] Küntay and Özyürek (2006; Küntay 2012; Küntay et al. 2014) refer to an unpublished study by Özyürek and Kita where they proposed a parallel analysis of the Turkish demonstrative 'şu' and the Japanese demonstrative 'so' (see also Levinson 2004).

an expert on Theory of Mind away from linguistics (and back to false-belief tasks), yet the marking of some of these distinctions bears on social cognition and deserves explanation not only as grammatical phenomena, but also as Theory of Mind abilities that are deployed in everyday social interaction.

Peeters and Özyürek (2016) have recently proposed that the production and comprehension of demonstratives are not primarily driven by the physical proximity of a referent to the speaker, but rather by the *psychological proximity* of a referent to both speaker and listener. In this social account, speaker and listener jointly establish which referents are psychologically proximal, relying on features of the referent such as its visibility, familiarity and ownership. Levinson (2018) also notes that the notion of *accessibility* is not only physical (commonly referred to as *reachability*) but also conceptual, marking whether a referent is or is not in the interlocutors' focus of attention. Defined in these terms, monitoring *cognitive accessibility* is a mindreading ability that is recruited by the different grammars of engagement of the world languages.[3]

5 Studies of demonstratives: different perspectives on Theory of Mind use

5.1 Typological studies

Traditionally, semantic analyses of demonstratives have posited that these expressions encode a distance relation to the speaker, which served as the basis for an egocentric, body-oriented representation of space in language and cognition (Diessel 2014). Accumulating data from linguistic fieldwork and experimental work with European languages suggest that the distance relation between the speaker and the referent may not always be the most basic relation encoded in demonstrative systems, with the status of the listener's attention to the referent being more basic in some languages.

Turkish demonstratives have already been described as encoding a two-term distance distinction relative to the speaker, with a third form marking those referents that are not yet in the interlocutors' joint focus of attention (Özyürek 1998). Burenhult (2003) investigated the attentional characteristics of 'ton', a nominal demonstrative in Jahai (Mon-Khmer, Malay Peninsula) that had previously been analyzed as marking spatial proximity to the listener. Burenhult describes the Jahai demonstrative system as the 'mirror image of the Turkish demonstrative system as re-analyzed by Özyürek (1998)' (2003: p. 377). Thus, 'ton' does not encode spatial information but rather marks that the referent is known to the listener, or already in their focus of attention. The remaining demonstrative forms in Jahai encode whether the referent is accessible to the speaker or to the listener, while having the opposite function to 'ton': namely, to draw the listener's attention to a new referent.

---

[3] While an abstract discussion of how languages mark cognitive accessibility may seem exotic or remote to a non-linguist, it must be noted that the English grammar requires that speakers mark whether a referent is inside or outside their common ground with the listener. For example, the utterance 'We're celebrating that I passed the exam' presupposes that the listener knows about the exam; otherwise, the utterance should be 'We're celebrating that I passed an exam'.

In a study of the use of demonstratives in Yucatec Maya, Bohnemeyer (2018) used an elicitation questionnaire and observed a systematic contrast between simple forms used with a pre-established focus of attention, and augmented forms used for attention direction (for a different analysis based on data from spontaneous interactions, see Hanks 2005). Bohnemeyer explains the importance of attention-direction in demonstrative forms: rather than providing a description of the referent, exophoric demonstratives provide information about where to find a referent. It is therefore not surprising that some languages use attention-calling forms to alert the listener to a new referent, and joint attention forms for those referents already in the interlocutors' focus of attention. In the case of the Yucatec demonstrative system, Bohnemeyer (2018) distinguishes two functions, which are encoded separately in the language: *deictic anchoring*, which is marked by the simple forms and distinguishes referents that are accessible or inaccessible to the speaker, and *attention calling*, which is marked by the augmented forms and distinguishes referents that are easily identifiable in the visual field from those that are not.

In a study of 'this', 'that' and 'it' in American English, Strauss (2002) proposes that speakers establish referent accessibility according to the degree of attention that the listener should pay to the referent, with 'this' marking high focus, 'that' marking medium focus and 'it' marking low focus. According to Strauss, this *gradient focus of attention* is determined by two factors: (1) the sharedness (or presumed sharedness) of the information, and (2) the relative importance of the referent for the speaker. Strauss (2002) presents this model as a dynamic alternative to the traditional proximal/distal distinction, arguing that the traditional analysis fails to explain how demonstratives are used in spoken English. Jarbou (2010) has recently proposed a similar analysis of spoken Jordanian Arabic in terms of *accessibility*, understood as the perceived ease of identification of the referent for the listener, regardless of its physical proximity.

In the introduction to his study of demonstratives in Yucatec Maya, Bohnemeyer (2018) describes traditional semantic analysis of demonstratives as seeking to determine their context-invariant meaning by eliminating all context dependencies. Similar views have been expressed in the other works reviewed in this section (see Özyürek 1998; Strauss 2002; Burenhult 2003; Hanks 2005; Jarbou 2010). Moving forward, Bohnemeyer makes the following proposal: "What is needed in order to study the use of demonstratives for exophoric spatial reference is a methodology that allows one to keep track of the interactional parameters of the speech context in which these forms are used. This includes the participants, their locations in real and in social space, and the location of the reference object (or *denotatum*) in these co-ordinate systems; e.g., the attention sharing among the speech act participants and the information status of the referent in discourse, and also possession of the object referred to by one of the participants" (2018: p. 177). All these interactional parameters could in principle be encoded in the semantics of demonstrative expressions, making their pragmatics dependent on Theory of Mind use. Experimental studies in psycholinguistics have considered some of these interaction parameters when investigating the use of demonstratives in adult interaction.

5.2 Psycholinguistic studies

Diessel (2005) found that the most common distinction encoded by demonstrative systems is a binary proximal/distal distinction: of a sample of 234 languages, more than half marked such a distinction. However, recent psycholinguistic studies have revealed a more nuanced picture of the parameters affecting a speaker's choice of demonstrative form, even for those systems traditionally analyzed as distance-based (for neuroscientific evidence, see Bonfiglioli et al. 2009; Stevens and Zhang 2013, 2014; Peeters et al. 2015). In a laboratory experiment comparing the use of demonstratives in English and Spanish, Coventry et al. (2008) observed that both language groups reduced their use of proximal forms (i.e. 'this' in English and 'este' in Spanish) when the target object was moved outside the speaker's peripersonal space, supporting the traditional analysis. However, when English and Spanish participants were given a long tool that allowed them to reach the target object beyond their normal reach, their peripersonal space was extended, together with their use of proximal demonstratives.

Coventry et al. (2008) also observed that not only spatial but also interactive factors affected demonstrative choice in English and Spanish: both language groups were sensitive to whether the participant or the experimenter had placed the target object in its location. English speakers used the proximal form 'this' more often when they had manipulated the object themselves, and Spanish speakers used the medial and distal forms ('ese' and 'aquel', respectively) more often when the object had been manipulated by the experimenter. In a follow up study looking at the mapping between linguistic and non-linguistic representations of space, Coventry et al. (2014) observed that three other interactive factors affected demonstrative choice in English: visibility, ownership and familiarity. Importantly, these variables did not interact with relative distance, suggesting that demonstrative choice in English is determined by more than a single space parameter.

In line with the pragmatic analysis proposed by Strauss (2002) for American English, Piwek et al. (2008) hypothesized that, when Dutch speakers use demonstratives accompanied by a pointing gesture, they use the proximal demonstrative for *strong indicating*, and the distal form for *neutral indicating*. In other words, the proximal demonstrative 'dit' would be used when the referent is not in the listener's focus of attention (low cognitive accessibility) and the distant form 'dat' would be used when the referent is already in the interlocutors' joint attention (high cognitive accessibility). The results of an unscripted interactive task supported this hypothesis. However, Piwek et al. (2008) did not control for the relative distance between the interlocutors and the target objects, leaving open the possibility that the effect of cognitive accessibility may have been modulated by space considerations.

In a follow-up study with Dutch speakers, Peeters et al. (2014) observed that distal demonstratives were used more often when both interlocutors were jointly attending to the referent, supporting Piwek et al.'s (2008) hypothesis that the Dutch distal demonstrative is used in situations of high cognitive accessibility. However, Peeters et al. also observed that participants were sensitive to space considerations, showing a preference for the proximal form when the referent was near the speaker and for the distal form when it was far away. Therefore, the distal form in Dutch seems to

be used both in a speaker-anchored way (indicating far-away referents) and also in a listener-anchored way (indicating referents in joint attention). These results suggest that, as in the case of English and Spanish, not only spatial but also interactive factors affect the use of demonstrative forms in Dutch.

In a recent study with native speakers of Danish, Rocca et al. (2018) observed that the use of proximal demonstratives increased not only as the target object was closer to the speaker, but also when it was closer to other objects in the physical context. These results suggest that the search space is organized as a contrastive space rather than being based on a simple peripersonal/extrapersonal distinction. Interestingly, Rocca et al. also observed a right-lateralized bias in the use of proximal demonstratives in Danish: participants used the proximal form 'den her' ('this one') more frequently when the referent object was closer to their right hand. This bias suggests that proximal demonstratives are more likely to be used for referents affording easier manual manipulation. Finally, like earlier studies, Rocca et al. (2018) also observed an effect of interactive factors on demonstrative choice: Danish speakers shifted their proximal space towards their shared space with the listener when they were actively collaborating on the task, but not when the other person was merely present (see also Rocca et al. 2019).

In summary, the results of several psycholinguistic studies reveal that demonstrative choice in European languages is affected not only by space considerations mapping onto the proximal/distal distinction, but also by interactive factors potentially requiring the use of social cognition abilities such as visual perspective-taking and attention monitoring. This nuanced picture leaves open several questions for the acquisition of demonstratives. For instance, when do children start using demonstratives in an adult-like fashion? Are young children initially sensitive to all factors potentially affecting demonstrative choice (e.g., the listener's focus of attention or the visibility of the referent), or do they first establish the basic proximal/distal distinction and only later become sensitive to interactive factors? Also, when it comes to learning interactive parameters, does it matter whether the child's language has specific forms for attention monitoring (as in Turkish, Jahai or Yucatec Maya), or do children start taking into account these factors around the same age independently of the language that they are acquiring? Unfortunately, cross-linguistic studies on the acquisition of demonstratives have not yet addressed all these questions.

5.3 Acquisition studies

Early studies on the acquisition of English deictic terms were conceived as tests of spatial egocentrism following Piaget's stage analysis. de Villiers and de Villiers (1974) observed that 3-year-olds were able to use 'this' and 'that' correctly as spatial deixis terms, but later work confirmed that the good performance reported in that study was dependent on the specific methodology used. Webb and Abrahamson (1976), Clark (1978) and Clark and Sengul (1978) found that young children had difficulties with perspective switching (e.g., understanding that 'here' refers to a different space depending on who is talking).

Deictic terms like 'this' and 'that' or 'here' and 'there' are usually present in child speech by age 2;6, but their comprehension suggests an immature understanding of the encoded contrasts. Clark and Sengul (1978) proposed two principles that children need to learn in order to master these deictic terms: the *Speaker principle*, according to which the speaker is the reference point, and the *Distance principle*, according to which deictic pairs such as 'here' versus 'there' or 'this' versus 'that' mark a distance contrast. Clark (1978) argues that in the process of learning deictic terms, young children test a series of hypothesis that allow them to refine the meaning of these words in three stages. At the *No-contrast stage*, children start using only one member of a deictic pair combined with a pointing gesture in order to indicate objects at any distance (average age: 3;3). At the *Partial-contrast stage*, children start using both terms in a pair, but have not yet mastered their contrastive meaning (3;10). For example, they may appreciate that 'here' and 'there' indicate different locations but not that they mark relative distance from the speaker's position. In the *Full-contrast stage*, children have adjusted their initial hypothesis and master both the Distance and Speaker principles (4;0).

Early developmental studies were conducted in English with a focus on the acquisition of spatial semantic contrasts. However, they did not investigate whether children were sensitive to any of the interactive factors that have been shown to affect the use of demonstratives in recent psycholinguistic studies with adults (e.g., familiarity, ownership or focus of attention). In the first study to investigate the acquisition of demonstratives in a language that encodes interactive aspects in their basic semantic meaning, Küntay and Özyürek (2006) collected conversational data from 4- and 6-year-old and adult speakers of Turkish. Their results showed that Turkish-speaking adults used demonstratives more frequently than children, and in different patterns. Children made appropriate use of the forms encoding a distance contrast, but revealed differences in their use of the form 'şu', which adults reserved to introduce new referents not yet in joint attention. Children used this form less frequently than adults and often used the form 'bu' instead. Küntay and Özyürek argue that 'even though demonstrative pronouns in early speech might be employed for getting attention (Clark and Sengul 1978), the ability to monitor and manipulate the participants' attentional states with the differential choice of demonstratives in conversation might develop much later' (2006: p. 308).

Rozendaal and Baker (2008) examined the acquisition of reference to persons and objects with indefinite and definite-demonstrative determiners by 2- to 3-year-old children acquiring Dutch, English and French. Their results revealed cross-linguistic differences in children's speed of acquisition, with French children being the fastest to acquire their determiner system, followed by English and Dutch children. These cross-linguistic differences are related to the frequency of determiners in the input: bare nouns were rare in the French input, whereas they were more frequent in Dutch than in English. This means French children have a strong cue in the input signaling the need to have a grammatical element precede nouns, while in English and Dutch this cue is less strong, slowing down children's acquisition of determiners as a result.

The pragmatic function to mark discourse-given entities was very frequent in both the child and adult samples investigated. Rozendaal and Baker (2008) argue that this pragmatic function is learned through an association with definite-demonstrative deter-

miners and a dissociation with indefinite determiners. Most relevant for the present study, children showed adult-like form–function associations once they started using a determiner to mark specificity and the new/given distinction (e.g., whether a character or entity is new or given in the conversation), but not for mutual knowledge. According to Rozendaal and Baker, errors in marking mutual knowledge result from children's lack of perspective-taking skills, which depend on their developing Theory of Mind. Importantly, however, lack of mutual knowledge was rarely marked in their samples, both in children's and adult speech. This suggests that familiar adults and children do not often rely on this pragmatic function in their exchanges, probably because of their extensive common ground and the limited scope of their conversations. The authors conclude that children need frequent contexts involving the expression of different pragmatic functions to build up the appropriate form–function associations.

More recently, Chu and Minai (2018) compared children's comprehension of demonstrative forms in English and Mandarin Chinese, both of which encode a two-term proximal/distal distinction. More specifically, this recent study investigated the relationship between demonstrative comprehension, Theory of Mind and Executive Function in 3- to 6-year-old children, with a focus on perspective switching. The results confirmed that children's comprehension of those demonstrative forms that required switching to the speaker's perspective correlated with their performance in a Theory of Mind task where children had to attribute knowledge to one of two puppets (a 'knower' or a 'guesser'), as well as with their performance in an Inhibitory Control task. Importantly, these correlations were not mediated by the children's language, suggesting a similar developmental path in English and Mandarin Chinese.

In summary, early studies on children's acquisition of demonstratives focused on their mastery of the proximal/distal distinction and their ability to switch perspectives with the listener. The most recent study by Chu and Minai (2018) confirmed that children's ability to switch perspectives in demonstrative comprehension is related to the development of their Theory of Mind and Executive Function. In their cross-linguistic study, Rozendaal and Baker (2008) argued that young children's errors with marking common knowledge (or a lack thereof) result from their immature perspective-taking skills and the low frequency of certain pragmatic functions in their input. While recent studies have provided valuable cross-linguistic data, only one study to date has looked at the acquisition of demonstrative forms that require both an understanding of spatial relations and monitoring the interlocutor's focus of attention, with the later ability emerging later than the former (Küntay and Özyürek 2006). In their discussion of children's protracted use of the Turkish demonstrative 'şu', Küntay and Özyürek (2006) draw an interesting parallel with children's mastery of the indefinite article, which is also used to introduce new referents not yet in common ground (e.g., 'We saw a fireman today') and has been shown to lag in development until 6 or 7 years of age, not revealing adult patterns until age 10 (Küntay 2002). The parallel between the acquisition of demonstrative and article systems is particularly interesting from an evolutionary perspective, as it mirrors language change.

6 Language change: implications for Theory of Mind

6.1 Expanding common ground

Grammaticalization is defined as the process whereby lexical words, such as nouns and verbs, develop into grammatical markers (Diessel 2007). Interestingly, grammaticalization processes tend to have a common source and follow *universal pathways*. The evolution of demonstratives into definite articles is one such universal pathway (Greenberg 1978), with the definite article 'the' in Modern English having its source in the Old English 'se' paradigm (Lyons 1999). In their most basic exophoric function, demonstratives have the same role as a pointing gesture: both indicate the location of a physical referent relative to the deictic center (e.g., 'Look at that!'; Diessel 2006). When they start being used for text-internal reference, exophoric demonstratives often develop into anaphoric demonstratives (e.g., 'John and Judy met in 1996. That year they got married'). Anaphoric demonstratives are not accompanied by pointing gestures because discourse referents are not visible, but both demonstrative forms have the same function: directing the listener's attention to a referent in the context, either physical or linguistic (Diessel 2006, 2012b).

Demonstratives that are routinely used to refer to linguistic elements in discourse provide a common historical source for definite articles (gloss: 'I bought that house you told me about' > 'I bought the house'; Diessel 2006, 2007, 2012b). Anaphoric demonstratives are normally used to refer to antecedents that are somewhat unexpected, contrastive or emphatic (Diessel 1999). However, when anaphoric demonstratives develop into definite articles, they start being used with all kinds of referents in the preceding discourse, losing their referential function and becoming formal markers of definiteness. In his seminal study, Greenberg (1978) described this process of language change as follows: 'The point at which a discourse deictic becomes a definite article is where it becomes compulsory and has spread to the point at which it means "identified" in general, thus including typically things known from context, general knowledge, or as with 'the sun' in non-scientific discourse, identified because it is the only member of its class' (pp. 61–62). Along similar lines, Diessel (2012a) describes definite articles as a *reference tracking device* that allows interlocutors to keep track of familiar referents.

Once again, the discussion of the evolution of demonstratives into definite articles seems to be taking us away from the realm of cognitive psychology and into the technical jargon of linguists and typologists. After all, the evolution of these forms marks a change in their semantics. However, I want to argue that this particular instance of language change has clear conceptual parallels in pragmatics and Theory of Mind (see Table 2). From a pragmatics viewpoint, the use of exophoric demonstratives relies on monitoring the physical context or what is *co-present* for speaker and listener (Clark and Marshall 1981). The use of anaphoric demonstratives, on the other hand, requires keeping track of previous discourse, while definite articles can be used to signal referents in earlier common ground. In terms of Theory of Mind development, joint attention is built and trained on the physical space shared by interlocutors, with this early ability developing into more advanced forms of experiential and visual perspective-taking

Table 2 Conceptual parallels across language, pragmatics and Theory of Mind in the evolution of gestural demonstratives into discourse demonstratives and definite articles

Theory of Mind	Joint attention	Perspective taking	Epistemic reasoning	
Pragmatics	Co-presence		Discourse	Common ground
Language	Exophoric demonstratives	Anaphoric demonstratives		Definite articles
Examples	*Look at that house.*	*That year they met.*		*We bought the car.*

(Moll and Meltzoff 2011a, b). At the start of this paper, I hypothesized that the acquisition of demonstrative systems plays a key role in the development of joint attention and perspective taking across languages and cultures, with exophoric demonstratives and pointing gestures serving as universal tools for joint attention. Building on these early developmental milestones, the acquisition and use of anaphoric demonstratives and definite articles depend on more sophisticated Theory of Mind abilities: monitoring ongoing discourse and earlier common ground require, at a minimum, to be able to keep a record of what has been said and previously shared and, once fully developed, an understanding of what is *known* to the interlocutors in a conversation. Therefore, the use of anaphoric demonstratives and definite articles ultimately requires the development of *epistemic reasoning* (e.g., deciding whether the listener knows the person you want to talk about, or you first need to introduce them in the conversation).

While it might not seem like a remarkable communicative feat for a competent native speaker, mastering the definite/indefinite distinction requires drawing rather sophisticated Theory of Mind inferences, as illustrated by the following exchanges:

[1] A: Did I tell you that we bought a house?
 B: The one you showed me the other day?
 A: Oh, I forgot I had showed you the house! Yes, that's the one we bought.

[2] A: Did I tell you that we bought the house?
 B: Which house?
 A: Sorry, I thought I had told you we wanted to buy a house.

Scenarios [1] and [2] illustrate 'misuses' of the definite and indefinite articles given what is common ground between speakers A and B: in scenario [1], speaker A fails to identify the house they bought as part of their common ground with speaker B, which leads B to infer that perhaps A bought a different house. Conversely, in scenario [2], speaker A wrongly presupposes that the house they bought was part of their common ground with speaker B, which results in a communication breakdown.

Here it is also worth highlighting the speed and flexibility with which adult speakers draw epistemic inferences during conversation, which has led me to argue that conversation is the natural arena to investigate belief reasoning in everyday interaction—rather than false-belief tasks (Rubio-Fernadez 2017, 2019; Rubio-Fernandez et al. 2019). Gundel et al. (2007, 2013) analyze the kind of inferences that can be drawn from speakers using a definite or indefinite article as a type of *scalar implicature* (Geurts 2010): in scenario [1], speaker A used the weak description 'a house', from which speaker B infers that the stronger alternative, 'the house' does not apply. This type of pragmatic reasoning, Gundel and colleagues argue, emerges later in child
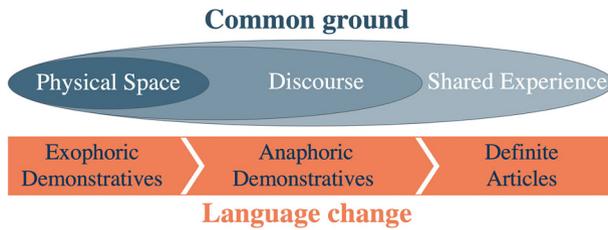
Fig. 1 A diachronic view of common ground expansion during language change

development than the mere acquisition of the article system because it requires more sophisticated Theory of Mind abilities.

Diessel (2006) characterized the evolution of demonstrative forms into anaphoric demonstratives and definite articles as an evolution of the corresponding communicative functions: 'Deictic > Anaphoric > Definite' (p. 477). Comparing the deictic function of exophoric demonstratives with the anaphoric function of later forms, Diessel (2013: p. 246) refers to the latter as 'disembodied uses' since discourse referents no longer have a physical substrate, unlike the co-present referents identified by exophoric demonstratives and pointing gestures. Here I want to propose a *diachronic view of common ground* whereby this pathway of language change marks a three-step expansion of the speakers' notion of common ground, starting with the shared physical space, and abstracting away to their ongoing discourse representation, and further still, to earlier experiences and world knowledge shared by the interlocutors (see Fig. 1).

Importantly, this three-step expansion of the notion of common ground characterizes not only language change but also language acquisition. Thus, developmental research in a number of areas has shown that young children start by relying on their shared physical space to build a common ground with their interlocutors, before they can form reliable discourse representations or engage in epistemic reasoning (for a review, see Moll and Kadipasaoglu 2013). Young children's over-reliance on co-presence has been argued to explain some of their so called 'egocentric' communicative behaviors: for example, their use of definite articles to introduce new characters (de Cat 2011, 2013), or even omitted forms if the referent is in joint attention (Skarabella et al. 2013; see also Gundel et al. 2013).

Language acquisition studies have repeatedly shown that young children rely on prior mention to formulate appropriate referential expressions before they rely on perceptual availability (Allen et al. 2015). This might seem to suggest that children are sensitive to discourse representations before they are sensitive to co-presence. However, young children's sensitivity to prior mention has been argued to be a form of *discourse alignment* which does not necessarily require perspective taking (Matthews et al. 2006). In this view, the child can rely on ad hoc strategies based on their linguistic knowledge (e.g., the answer to the question 'What did X do?' must be [pronoun/null reference + verb], whereas the answer to the question 'What happened?' must include a full noun; Serratrice 2008). Therefore, monitoring perceptual availability requires perspective taking and lags behind in pragmatic development, whereas prior mention allows for an immediate computation of joint attention, resulting in an early form of common ground (see Table 2).

Studies with older children looking at their narrative skills in the absence of shared knowledge with their interlocutor reveal that children under 5–6 years rarely use pronouns to introduce a new character in a story, revealing an emerging sensitivity to common ground constraints (Hickmann et al. 2015). However, the same children use both definite and indefinite articles to introduce new characters, as well as a substantial number of pronouns and omissions to re-introduce an old character, often resulting in ambiguity. The acquisition of discourse narrative functions is most delayed in the case of re-introductions, which are not mastered until age 7–10 years. This developmental pattern suggests that reference maintenance is easier than reference introduction and re-introduction, supporting the view that managing common ground through discourse representation is more demanding than monitoring what is already in joint attention.

In this view, the development of the child's notion of common ground is parallel to the development observed in the historical record across a vast number of languages that evolved definite articles from their demonstrative forms. This parallel development should allow us to model the development of Theory of Mind use in communication not only during childhood, but also during language change. Here I want to hypothesize that the expansion of common ground from the shared physical environment to the ongoing discourse representation, and our shared experiences and knowledge requires not only increasing memory capacity, but also *specialized forms of social attention*. While we could easily keep a mental record the eye color of our interlocutors or what they were wearing yesterday, we are much more likely to remember what we talked about, or whom they know, and use that information efficiently in our future conversations. Thus, in order to manage our multiple common grounds with our interlocutors (be those family, friends, acquaintances or strangers) we must focus our attention on what matters for communication: namely, *knowledge* and *sharedness* (or who knows what *with* us, or *for* us).

6.2 Pragmatic relativity

In their review of the linguistic relativity hypothesis, Wolff and Holmes (2011) discuss various ways in which language has been argued to affect thought. What they call *thinking before language* (or in Slobin's terms, *thinking for speaking* (1996)) has been the least controversial evidence in the long and heated debate on the *linguistic relativity hypothesis* (see Pinker 2007; Enfield 2015). For example, when describing a motion event, psycholinguistic studies have shown that English speakers attend to manner, whereas Greek speakers focus on path in order to encode either feature in their motion verbs, as required by their respective grammars (Papafragou et al. 2008).

Here I have hypothesized that the acquisition of pragmatic markers, such as demonstratives and definite articles, scaffold the development of early forms of Theory of Mind, leaving room for cross-linguistic differences in the development of social cognition. From the perspective of linguistic relativism, this hypothesis would be a form of thinking for speaking: if your grammar requires that you monitor your interlocutor's focus of attention in order to choose one demonstrative form or another (as in Turkish or Japanese, for example), your attentional resources are going to be allocated accordingly. A counterargument to such an effect of language on thought might

be that neurotypical speakers of all languages are able to monitor their interlocutor's focus of attention, regardless of the type of demonstrative system that they use in their language. That much is true, of course, and would be true of other forms of thinking for speaking: native speakers of English are obviously able to describe a motion event in terms of path, just as native speakers of Greek can do the same in terms of manner. However, what the experimental record shows is that when describing a motion event, these speakers *automatically focus* on those aspects of motion that are encoded by their grammar (Papafragou et al. 2008). Similarly, in my view, what the grammars of engagement of the world languages do for both children acquiring the language and adults using it proficiently is to train them to allocate their attentional resources in such a way as to monitor their interlocutor's focus of attention and build a common ground during interaction.

Another possible counterargument to the pragmatic relativity hypothesis may come from a linguistics angle: languages that do not have articles may mark the *new/given distinction* (e.g., 'We bought a house' vs. 'We bought the house') by relying on multi-functional morphemes such as case marking, the use of a numeral, or word order. What the experimental record shows, in this case, is that when definite/indefinite markers are obligatory in a language, children learn their function earlier (Küntay et al. 2014). At a first pass, this suggests that acquiring the new/given distinction through global markers such as word order, case marking or optional lexical items is harder than through local markers, such as definite and indefinite articles. The upshot of this difference is that acquiring local pragmatic markers may train children in their use of common ground, first by frequently exposing them to an explicit marking of the new/given distinction, and then by their own use of these markers in child speech.

A third counterargument to the pragmatic relativity hypothesis may come from the viewpoint of language evolution. One could imagine that the development of an article system might mark a milestone in the evolution of our Theory of Mind abilities; perhaps even an endpoint when our epistemic reasoning abilities (the hallmark of having a Theory of Mind) get recruited by our grammars and start being deployed automatically in our everyday social interactions. However, the historical record seems to contradict this evolutionary story: Greenberg (1978) has shown that the use of definite articles may spread from definite nouns to both definite and indefinite nouns, with their grammaticalization path continuing until they turn into gender or noun class markers, before they eventually disappear. As Lyons (1999) so fittingly put it in the closing line to his volume on definiteness: 'Not only can languages acquire the category of definiteness; they can also lose it' (p. 340). If the evolution of demonstrative markers into definite articles mirrored the evolution of our Theory of Mind abilities, the loss of the definite article should also mark the decay of the social cognition abilities of a speech community, and that does not seem plausible.

Lyons (1999) refers to the loss of the category of definiteness as the most intriguing point in the progression of the demonstrative: 'It is far from obvious why a formative with an important discourse function should lose it, and in many cases cease to have any grammatical or semantic function' (p. 339). It is equally unclear what role (if any) Theory of Mind could play in the loss of the definite article in some languages. However, regarding the *implications* of this loss for Theory of Mind use, we should bear in mind that other referential expressions may require similar mindreading abilities as

the use of the definite article. For example, choosing between a full noun or a pronoun often relies on the same new/given distinction as the use of a definite or indefinite article (compare, e.g., 'We saw Kenny this morning. He seemed very happy' vs. 'We saw a new Tesla this morning. The driver seemed very happy'). In this view, the loss of the definite article would reduce the frequency with which speakers need to mark common ground in their language production and track it in their language comprehension. A direct result of this process of language change would be that speakers of languages undergoing this change may eventually stop tracking and marking common ground automatically. However, the same mindreading abilities that are deployed in using a definite article may be deployed in the use of other linguistic forms, not necessarily being lost altogether with the decay of the definite article.[4]

The pragmatic relativity hypothesis is in line with Enfield's (2015) work on linguistic relativity as a hypothesis about the social reality of speakers. According to Enfield, 'If a language makes fine distinctions in meaning in some domain, people who speak that language will be subject to a different normative background for interpretation and accountability than they would be in the context of a language that does not make the same fine distinctions' (p. 217). Thus, a speaker of a language with definite and indefinite articles is *accountable* for their marking of common ground, such that if they misuse definite articles to refer to entities outside their common ground with the listener (or omit the marking of entities in common ground), the listener is entitled to call them out (as illustrated in Scenarios [1] and [2] above). In their seminal work on the pragmatics of language change, Hopper and Traugott (1993) showed how inference drives grammaticalization. Yet as Enfield (2015) points out, the inferences that people make are those available in their linguistic system. Thus, having an article system in one's language not only enable speakers to monitor and mark common ground precisely, it also enables common ground inferences to shape communication and eventually, drive language change.

In summary, according to the pragmatic relativity hypothesis, languages with demonstrative forms that require monitoring the listener's focus of attention will train their users in perspective taking, just as learning to use a definite article will train children to manage and mark their common ground with others. This may seem quite uncontroversial. However, as with the old linguistic relativity hypothesis (Whorf in Carroll 1956), the real question is how deep these cross-linguistic differences run (Pinker 2007; Enfield 2015). Ultimately, it is an empirical question whether Turkish- or Japanese-speaking children are better at monitoring their interlocutors' focus of attention than children whose language does not have an attention calling demonstrative. A more interesting question that may not be as easily testable (although it may be possible to model the process using computational methods) is whether the mindreading abilities of a language community change when they start using definite (or indefinite) articles and these forms crystallize into a new grammatical marker. Are

---

[4] Similar arguments could be made about languages without article systems. Ultimately, it is an empirical question whether having definite/indefinite articles results in a different marking of common ground, or simply in the automatization of certain mindreading processes common to all languages. Pointing in the direction of a conceptual difference, extensive research suggests that acquiring articles in a second language is considerably harder if the learner's first language does not have an article system (Ionin et al. 2009; Dayal 2018).

English or Spanish speakers more sensitive to the new/given distinction than speakers of article-less languages such as Polish, Russian or Hindi? A relevant finding that has been repeatedly documented in second language acquisition is that native speakers of article-less languages rarely master the use of articles in a new language. As Dayal (2018) concedes: 'The statement that adult learners of a language with articles never quite master the system if their L1 lacks articles is almost a truism' (p. 23). While not necessarily a reflection of the mindreading requirements of marking common ground, the difficulties of second language learners to use articles appropriately if they lack them in their mother tongue certainly deserve investigation from a pragmatic/interactive perspective.

6.3 The ineffable power of procedural meanings

Heyes and Frith (2014) have recently argued that mindreading is a culturally transmitted ability. As Heyes (2018) explains, expert mindreaders pass on their social cognition skills by 'communicating mental state concepts, and ways of representing those concepts, to novices' (p. 168). Here, I have argued that language and Theory of Mind may co-develop through the acquisition and use of pragmatic markers, which might fit, in some respects, Heyes' description of how adults use communication to pass on their social skills to children. However, there is a fundamental difference between these two accounts: Heyes and Frith focus on *explicit* Theory of Mind (as measured by traditional false-belief tasks and evidenced by children's use of mental state verbs), whereas the social cognition skills that sustain a grammar of engagement (including the use and comprehension of pragmatic markers) would be better described as *implicit*. In the last section of this paper I want to argue that implicit forms of mindreading are as fundamental to human social cognition as its more explicit counterparts—if not more.

But a caveat is in order first: to argue that implicit forms of Theory of Mind are important for social cognition might seem controversial at this point, given the ongoing debate around false-belief studies with infants and whether babies have a Theory of Mind (e.g., Heyes 2014; Ruffman 2014; Rakoczy 2017). However, highlighting the importance of implicit forms of Theory of Mind is only problematic if we mistakenly equate having a Theory of Mind with passing a false-belief task. If we understand Theory of Mind to comprise abilities more basic than false-belief reasoning (e.g., attributing goals or establishing joint attention) as well as more sophisticated skills (e.g., reminding someone of something they seem to have forgotten, or preempting a misunderstanding), then we can simply appreciate implicit forms of Theory of Mind for being an integral part of the machinery behind our everyday social interactions (see Jara-Ettinger et al. 2016). We must also bear in mind that implicit forms of Theory of Mind may emerge earlier in development, but are not superseded by the explicit forms that emerge later: joint attention, for instance, continues to be fundamental to communicative success even after children start attributing beliefs to others. Therefore, the question is not whether implicit forms of mindreading are "real Theory of Mind" (as it has been formulated in recent studies with infants), but when and how exactly implicit and explicit forms of mindreading work *in tandem* to make efficient social interaction possible, including communication.

As mentioned in the introduction, Matsui et al. (2006) tested Japanese-speaking 3 to 6 year-olds on their understanding of speaker certainty and evidence for an expressed opinion. Part of its grammar of engagement, Japanese encodes certainty and evidentiality in high-frequency, closed-class, sentence-final particles, as well as in low-frequency, mental state verbs. Matsui and colleagues showed that Japanese-speaking children are able to make use of the epistemic information encoded in sentence-final particles earlier than the information encoded in mental state verbs. Supporting Heyes and Frith's cultural evolution of mindreading, children's epistemic vocabulary correlated with their performance in standard false-belief tasks. However, their understanding of sentence-final particles expressing the same meanings did not correlate with their understanding of false belief. Matsui et al. concluded that Japanese children's understanding of speakers' epistemic states as communicated by sentence-final particles precedes their later, fully-representational understanding of belief.

Following Matsui et al. (2006), Gundel et al. (2007) and Gundel and Johnston (2013) have argued that the mindreading skills involved in appropriately using determiners in spontaneous conversation (including demonstratives and definite articles) are also implicit, and crucially different from the explicit mindreading abilities required to pass false-belief tasks. This distinction is based on the linguistic dichotomy between procedural and conceptual meaning: closed-class words (such as particles or articles) normally encode non-representational, procedural information that is more implicit and automatic, whereas open-class words (such as mental state verbs) are declarative, representational and explicit. Procedural meanings are normally understood as specifications on how to manipulate conceptual meanings: for example, procedural terms such as 'however' or 'moreover' do not contribute to a propositional representation, but rather encode instructions for processing the propositional representations they introduce (Blakemore 1987).

The results of Matsui et al. (2006) suggest that, in language acquisition, implicit forms of mindreading lay the foundation for the development of explicit forms. Similarly, the developmental path from the use of demonstratives in joint attention, to later referential uses requiring perspective taking, and even later epistemic reasoning also seem to transition from implicit to explicit forms of Theory of Mind. However, I want to stress that these implicit forms of mindreading continue to be fundamental to human social cognition in later stages in life. While a mature native speaker might find it hard to explain the meaning of words like 'the' or 'a' (Gundel and Johnson 2013), they are certainly able to quickly derive epistemic inferences from the use of articles (as Scenarios [1] and [2] above illustrate). Thus, by the very nature of their procedural meaning, closed-class words enable the routinization and automatization of mindreading processes in adult communication, resulting in extremely fast and highly sophisticated social interactions that would simply not be possible on the basis of explicit mindreading alone.

As Enfield (2015) notes: 'Of special interest for Whorf and many since was the encoding of concepts in grammatical as opposed to lexical forms, given that the former are maximally requisite, tacit, and practiced, and thus maximally habitual' (Footnote 2; p. 210). In this view, the grammaticalization and evolution of pragmatic markers leads to the routinization of implicit forms of Theory of Mind, such as attention monitoring and common ground marking, which in turns contributes to the development and

evolution of Theory of Mind through communication—as predicted by the positive feedback loop hypothesis.

I have looked at demonstratives from a grammatical, developmental and interactive perspective. However, demonstratives are also a crucial part of what is known in linguistics and cognitive science as *information structure*: 'the interface between the structure and meaning of linguistic utterances, on the one hand, and the interlocutors' mental representations of information, discourse referents, and the overall universe of discourse, on the other' (Zimmermann 2016). Efficient information transfer between interlocutors depends on information structure so that interlocutors can rapidly and easily update their mental models of the world and establish common ground. Thus, marking information structure (by using a definite article or a marked intonation, for example) facilitates information update, including the interlocutors' belief states.

Different theoretical traditions have tried to explain information structural effects in language (for a review, see Arnold et al. 2013). Categorical approaches within linguistic theory describe the nature of the information itself by drawing distinctions such as *topic* versus *focus*, or *focus* versus *presupposition* (e.g., Reinhart 1981). The functional linguistics literature has proposed gradient representations of information status that vary along a hierarchy of specificity, from unstressed pronouns to heavily modified noun phrases (e.g., Ariel 1990). However, since information status also reflects the social/communicative dimensions of language use, other approaches have focused on the different sources of knowledge that make up common ground, including social, cultural and discourse background (e.g., Clark and Marshall 1981; Prince 1992), as well as the role of memory processes (Horton and Gerrig 2005). Finally, according to the optimal system approach, which is based on Information Theory, the most efficient way to maximize information transfer in a noisy communication channel, such as natural speech, is to maintain uniform information transmission over time (Shannon 1948).

It is difficult to envisage what kind of empirical evidence could settle the debate on whether human language could have evolved without Theory of Mind, even though language acquisition starts years before children are able to pass standard false-belief tasks (Sperber 2000; Malle 2002; Levinson 2006; Tomasello 2008; Scott-Phillips 2014; Woensdregt et al. 2020; Moore, under review). For what is worth, I want to conclude that, regardless of the final outcome of the chicken versus egg dilemma, the fact that languages have grammars of engagement, that information structure must be marked in order to make human communication as fast and efficient as it is today, are evidence that language and Theory of Mind are *co-evolving*, *as we speak*. Thus, the very existence of pragmatic markers of intersubjectivity speak to the mutual dependency between language and Theory of Mind. How far back in phylogeny that interdependency goes is beyond the scope of this paper, but I hope to have convincingly argued that if pragmatics has made it to the grammars of the world languages, that can only be the result of the joint use of language and Theory of Mind in everyday communication.

7 Concluding remarks

Christiansen and Chater (2016) argue that language is created across three parallel timescales: as we use it, both in language production and comprehension; that is the *timescale of the utterance*. Language is also created as we acquire it, not only during childhood but also across the lifespan; that is the *timescale of the individual*. And thirdly, language is created through cultural transmission by generations of language-learning individuals; that is the *timescale of language evolution*. In their book, Christiansen and Chater argue that the key to the origin and shape of human language lies in the relationship between these three timescales.

In this paper, I have argued that the study of pragmatics during language acquisition, language use and language change also promises to unlock the deep relationship and tight dependencies between language and Theory of Mind. Typological, psycholinguistic and developmental studies of demonstratives, all have called for a more rigorous investigation of the interactive factors underlying the use of these forms. More importantly, perhaps, they have also called for more cross-linguistic studies looking at the acquisition and use of demonstratives in non-Western societies (for some desiderata for the future study of language evolution, see Dediu et al. 2013). I want to conclude by echoing the need for these studies and insisting that the relationship between language and Theory of Mind is encoded in today's grammars of engagement. We just need to rise to the empirical challenge.

References

Allen, S., Hughes, M., & Skarabela, B. (2015). The role of cognitive accessibility in children's referential choice. In L. Serratrice & S. Allen (Eds.), *The acquisition of reference* (pp. 123–155). Amsterdam: Palgrave.

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

Arnold, J. E., Kaiser, E., Kahn, J., & Kim, L. (2013). *Information structure: Linguistic, cognitive, and processing approaches*. WIREs Cognitive Science. New York: Wiley. https://doi.org/10.1002/wcs.1234.

Baldwin, D. A. (1995). Understanding the link between joint attention and language. In C. Moore & P. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 131–158). Hillsdale, NJ: Erlbaum.

Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. Oxford: Oxford University Press.

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., et al. (2009). Language is a complex adaptive system: Position paper. *Language Learning, 59,* 1–26.

Blakemore, D. (1987). *Semantic constraints on relevance*. Oxford: Blackwell.

Bohnemeyer, J. (2018). Yucatec demonstratives in interaction: Spontaneous versus elicited data. In S. C. Levinson, S. Cutfield, M. Dunn, N. Enfield, & S. Meira (Eds.), *Demonstratives in cross-linguistic perspective* (pp. 176–205). Cambridge: Cambridge University Press.

Bonfiglioli, C., Finocchiaro, C., Gesierich, B., Rositani, F., & Vescovi, M. (2009). A kinematic approach to the conceptual representations of *this* and *that*. *Cognition, 111,* 270–274.

Burenhult, N. (2003). Attention, accessibility, and the addressee. *Pragmatics, 13,* 363–379.

Campbell, A. L., Brooks, P., & Tomasello, M. (2000). Factors affecting young children's use of pronouns as referring expressions. *Journal of Speech, Language, and Hearing Research, 43,* 1337–1349.

Carpenter, M., & Liebal, K. (2011). Joint attention, communication, and knowing together in infancy. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 159–181). Cambridge, MA: MIT Press.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63,* 1–143.

Carroll, J. B. (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. Cambridge, MA: MIT Press.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Christiansen, M. H., & Chater, N. (2016). *Creating language: Integrating evolution, acquisition, and processing*. Cambridge, MA: MIT Press.

Christiansen, M. H., & Kirby, S. (Eds.). (2003). *Language evolution*. Oxford: Oxford University Press.

Chu, C. Y., & Minai, U. (2018). Children's demonstrative comprehension and the role of non-linguistic cognitive abilities: A cross-linguistic study. *Journal of Psycholinguistic Research, 47,* 1343–1368.

Cifuentes-Honrubia, J. L. (1989). *Lengua y espacio: Introducción al problema de la deixis en español*. Alicante: Universidad de Alicante.

Clark, E. V. (1978). From gesture to word: On the natural history of deixis in language acquisition. In J. S. Bruner & A. Garton (Eds.), *Human growth and development* (pp. 85–120). Oxford: Oxford University Press.

Clark, H. H., & Marshall, C. R. (1981). Definite knowledge and mutual knowledge. In A. Joshi, B. Webber, & I. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge: Cambridge University Press.

Clark, E. V., & Sengul, C. J. (1978). Strategies in the acquisition of deixis. *Journal of Child Language, 5,* 457–475.

Coventry, K. R., Griffiths, D., & Hamilton, C. J. (2014). Spatial demonstratives and perceptual space: Describing and remembering object location. *Cognitive Psychology, 69,* 46–70.

Coventry, K. R., Valdés, B., Castillo, A., & Guijarro-Fuentes, P. (2008). Language within your reach: Near–far perceptual space and spatial demonstratives. *Cognition, 108,* 889–895.

Culbertson, J., & Adger, D. (2014). Language learners privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences, 111,* 5842–5847.

Dale, P. S., & Crain-Thoreson, C. (1993). Pronoun reversals: Who, when, and why? *Journal of Child Language, 20*(3), 573–589.

Dayal, V. (2018). (In)definiteness without articles: Diagnosis, analysis, implications. In G. Sharma & R. Bhatt (Eds.), *Trends in Hindi linguistics*. Berlin: Mouton de Gruyter.

de Cat, C. (2011). Information tracking and encoding in early L1: Linguistic competence versus 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.

de Cock, B. (2013). Entre distancia, discurso, e intersubjetividad: Los demostrativos neutros en español. *Anuario de Letras: Lingüística y Filología, 1,* 7–36.

de Villiers, J. G. (1999). On acquiring the structural representations for false complements. In B. Hollebrandse (Ed.), *New perspectives on language acquisition*. Amherst, MA: UMOP.

de Villiers, J. G. (2007). The interface of language and Theory of Mind. *Lingua, 117,* 1858–1878.

de Villiers, P. A., & de Villiers, J. G. (1974). On this, that, and the other: Non-egocentrism in very young children. *Journal of Experimental Child Psychology, 18,* 438–447.

Dediu, D., Cysouw, M., Levinson, S. C., Baronchelli, A., Christiansen, M. H., Croft, W., et al. (2013). Cultural evolution of language. *Cultural evolution: Society, technology, language, and religion. Strüngmann forum reports* (Vol. 12, pp. 303–332). Cambridge, MA: MIT Press.

Diessel, H. (1999). The morphosyntax of demonstratives in synchrony and diachrony. *Linguistic Typology, 3,* 1–49.

Diessel, H. (2003). The relationship between demonstratives and interrogatives. *Studies in Language, 27,* 635–655.

Diessel, H. (2005). Distance contrasts in demonstratives. In M. Haspelmath, M. Dryer, D. Gil, & B. Comrie (Eds.), *World atlas of language structures* (pp. 170–173). Oxford: Oxford University Press.

Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics, 17,* 463–489.

Diessel, H. (2007). Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology, 25,* 108–127.

Diessel, H. (2012a). Where do grammatical morphemes come from? On the development of grammatical markers from lexical expressions, demonstratives, and question words. In K. Davidse, T. Breban, L. Brems, T. Mortelmans, B. Cornillie, H. Cuyckens, & T. Leuschner (Eds.), *Grammaticalization and language change: New reflections*. Amsterdam: John Benjamins.

Diessel, H. (2012b). Deixis and demonstratives. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.), *An international handbook of natural language meaning* (Vol. 3, pp. 2407–2431). Berlin: Mouton de Gruyter.

Diessel, H. (2013). Where does language come from? Some reflections on the role of deictic gesture and demonstratives in the evolution of language. *Language and Cognition, 5,* 239–249.

Diessel, H. (2014). Demonstratives, frames of reference, and semantic universals of space. *Language and Linguistics Compass, 8,* 116–132.

Eguren, L. J. (1999). Pronombres y adverbios demonstrativos. Las relaciones deícticas. In I. Bosque & V. Demonte (Eds.), *Gramática Descriptiva de la Lengua Española* (Vol. 1, pp. 929–974). Madrid: Espasa.

Enfield, N. J. (2015). Linguistic relativity from reference to agency. *Annual Review of Anthropology, 44,* 207–224.

Evans, N. (2006a). Who said polysynthetic languages avoid subordination? Multiple subordination strategies in Dalabon. *Australian Journal of Linguistics, 26,* 31–58.

Evans, N. (2006b). View with a view: Towards a typology of multiple perspective. *Berkeley Linguistics Society, 31,* 93–120.

Evans, N., Bergqvist, H., & San Roque, L. (2018a). The grammar of engagement I: Framework and initial exemplification. *Language and Cognition, 10,* 110–140.

Evans, N., Bergqvist, H., & San Roque, L. (2018b). The grammar of engagement II: Typology and diachrony. *Language and Cognition, 10,* 141–170.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences, 32,* 429–448.

Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences, 109,* 17897–17902.

Fedzechkina, M., Newport, E. L., & Jaeger, T. F. (2017). Balancing effort and information transmission during language acquisition: Evidence from word order and case marking. *Cognitive Science, 41*(2), 416–446.

Flavell, J. H. (1992). Perspectives on perspective-taking. In H. Beilin & P. B. Pufall (Eds.), *Piaget's theory: Prospects and possibilities* (pp. 107–139). Hillsdale, NJ: Erlbaum.

Friedman, L., Lorang, E., & Sterling, A. (2019). The use of demonstratives and personal pronouns in fragile X syndrome and autism spectrum disorder. *Clinical Linguistics & Phonetics, 33*(5), 420–436.

Geurts, B. (2010). *Quantity implicatures*. Cambridge: Cambridge University Press.

Greenberg, J. H. (1978). How does a language acquire gender markers. *Universals of Human Language, 3,* 47–82.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics* (Vol. 3, pp. 41–58)., Speech acts New York: Academic Press.

Gundel, J. K., & Johnson, K. (2013). Children's use of referring expressions in spontaneous discourse: Implications for Theory of Mind development. *Journal of Pragmatics, 56,* 43–57.

Gundel, J. K., Ntelitheos, D., & Kowalsky, M. (2007). Children's use of referring expressions: Some implications for Theory of Mind. *ZAS Papers in Linguistics, 48,* 1–21.

Hacquard, V., & Lidz, J. (2019). Children's attitude problems: Bootstrapping verb meaning from syntax and pragmatics. *Mind and Language, 34,* 73–96.

Hanks, W. F. (2005). Explorations in the Deictic field. *Current Anthropology, 46,* 191–220.

Hanks, W. F. (2011). Deixis and indexicality. In W. Bublitz & N. R. Norrick (Eds.), *Foundations of pragmatics* (pp. 315–346). Berlin: De Gruyter Mouton.

Harris, P. L. (1996). Desires, beliefs and language. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (pp. 200–220). Cambridge: Cambridge University Press.

Harris, P. L. (1999). Acquiring the art of conversation: Children's developing conception of their conversation partner. In M. Bennett (Ed.), *Developmental psychology: Achievements and prospects* (pp. 89–105). Philadelphia: Psychology Press.

Heyes, C. M. (2014). False belief in infancy: A fresh look. *Developmental Science, 17,* 647–659.

Heyes, C. M. (2018). *Cognitive gadgets: The cultural evolution of thinking*. Cambridge, MA: Harvard University Press.

Heyes, C. M., & Frith, C. D. (2014). The cultural evolution of mind reading. *Science, 344*(6190), 1243091.

Hickmann, M., Schimke, S., & Colonna, S. (2015). From early to late mastery of reference: Multifunctionality and linguistic diversity. In L. Serratrice & S. Allen (Eds.), *The acquisition of reference* (pp. 181–211). Amsterdam: John Benjamins.

Hopper, P. J., & Traugott, E. C. (1993/2003). *Grammaticalization*. Cambridge: Cambridge University Press.

Horton, W. S., & Gerrig, R. J. (2005). Conversational common ground and memory processes in language production. *Discourse Process, 40,* 1–35.

Ionin, T., Zubizarreta, M. L., & Philippov, V. (2009). Acquisition of article semantics by child and adult L2-English learners. *Bilingualism: Language and Cognition, 12,* 337–361.

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences, 20,* 589–604.

Jarbou, S. O. (2010). Accessibility versus physical proximity: An analysis of exophoric demonstrative practice in spoken Jordanian Arabic. *Journal of Pragmatics, 42,* 3078–3097.

Jungbluth, K. (2003). Deictics in the conversational dyad: Findings in Spanish and some cross-linguistic outlines. In F. Lenz (Ed.), *Deictic conceptualisation of space, time and person* (pp. 13–40). Amsterdam: John Benjamins Publishing Company.

Kemmerer, D. (1999). 'Near' and 'far' in language and perception. *Cognition, 73,* 35–63.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences, 105,* 10681–10686.

Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology, 28,* 108–114.

Kita, S. (Ed.). (2003). *Pointing: Where language, culture, and cognition meet*. Mahwah, NJ: Erlbaum.

Küntay, A. C. (2002). Development of the expression of indefiniteness: Presenting new referents in Turkish picture-series stories. *Discourse Processes, 33,* 77–101.

Küntay, A. C. (2012). Crosslinguistic research. In E. Hoff (Ed.), *The Blackwell guide to research methods in child language* (pp. 287–299). Malden, MA: Wiley-Blackwell.

Küntay, A. C., Nakamura, K., & Ates Sen, B. (2014). Crosslinguistic and crosscultural approaches to pragmatic development. In D. Matthews (Ed.), *Pragmatic development in first language acquisition* (pp. 317–341). Amsterdam: John Benjamins.

Küntay, A. C., & Özyürek, A. (2006). Learning to use demonstratives in conversation: What do language specific strategies in Turkish reveal? *Journal of Child Language, 33,* 303–320.

Levinson, S. C. (2004). Deixis and pragmatics. In L. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 97–121). Oxford: Blackwell.

Levinson, S. C. (2006). Cognition at the heart of human interaction. *Discourse Studies, 8,* 85–93.

Levinson, S. C. (2018). Demonstratives: Patterns in diversity. In S. C. Levinson, S. Cutfield, M. Dunn, N. Enfield, & S. Meira (Eds.), *Demonstratives in cross-linguistic perspective* (pp. 1–42). Cambridge: Cambridge University Press.

Lohmann, H., & Tomasello, M. (2003). The role of language in the development of false belief understanding: A training study. *Child Development, 74,* 1130–1144.

Loveland, K. A. (1984). Learning about points of view: Spatial perspective and the acquisition of 'I/you'. *Journal of Child Language, 11*(3), 535–556.

Lyons, C. (1999). *Definiteness*. Cambridge: Cambridge University Press.

Malle, B. F. (2002). The relation between language and Theory of Mind in development and evolution. In T. Givón & B. F. Malle (Eds.), *The evolution of language out of pre-language* (pp. 265–284). Amsterdam: Benjamins.

Matsui, T., Yamamoto, T., & McCagg, P. (2006). On the role of language in children's early understanding of others as epistemic beings. *Cognitive Development, 21,* 158–173.

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.

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and Theory of Mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development, 78,* 622–646.

Mithun, M. (1984). How to avoid subordination. *Berkeley Linguistic Society, 10,* 493–509.

Moll, H., Carpenter, M., & Tomasello, M. (2007). Fourteen-month-olds know what others experience only in joint engagement. *Developmental Science, 10,* 826–835.

Moll, H., & Kadipasaoglu, D. (2013). The primacy of social over visual perspective-taking. *Frontiers in Human Neuroscience, 7,* 558.

Moll, H., & Meltzoff, A. N. (2011a). Perspective-taking and its foundation in joint attention. In N. Eilan, H. Lerman, & J. Roessler (Eds.), *Perception, causation, and objectivity. Issues in philosophy and psychology* (pp. 286–304). Oxford: Oxford University Press.

Moll, H., & Meltzoff, A. N. (2011b). Joint attention as the fundamental basis of perspectives. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 393–413). Cambridge, MA: MIT Press.

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

Moll, H., Richter, N., Carpenter, M., & Tomasello, M. (2008). Fourteen-month-olds know what "we" have shared in a special way. *Infancy, 13,* 90–101.

Moll, H., & Tomasello, M. (2006). Level 1 perspective-taking at 24 months of age. *British Journal of Developmental Psychology, 24,* 603–613.

Moore, R. (under review). Extended social cognition and the cultural evolution of mind-modelling.

Moore, C., & Dunham, P. J. (1995). *Joint attention: Its origins and role in development*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs. *Science, 308*(5719), 255–258.

Özyürek, A. (1998). An analysis of the basic meaning of Turkish demonstratives in face-to-face conversational interaction. In S. Santi, I. Guaitella, C. Cave, & G. Konopczynski (Eds.), *Oralite et Gestualite: Communication multimodale, interaction*. Paris: L'Harmattan.

Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition, 108,* 155–184.

Peeters, D., Azar, Z., & Ozyurek, A. (2014). The interplay between joint attention, physical proximity, and pointing gesture in demonstrative choice. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 1144–1149).

Peeters, D., Hagoort, P., & Özyürek, A. (2015). Electrophysiological evidence for the role of shared space in online comprehension of spatial demonstratives. *Cognition, 136,* 64–84.

Peeters, D., & Özyürek, A. (2016). *This* and *that* revisited: A social and multimodal approach to spatial demonstratives. *Frontiers in Psychology, 7,* 222.

Pinker, S. (2007). *The stuff of thought: Language as a window into human nature*. New York: Viking.

Piwek, P., Beun, R. J., & Cremers, A. (2008). 'Proximal' and 'distal' in language and cognition: Evidence from deictic demonstratives in Dutch. *Journal of Pragmatics, 40,* 694–718.

Prince, E. (1992). The ZPG letter: Subjects, definiteness, and information-status. In S. Thompson & W. Mann (Eds.), *Discourse description: Diverse analyses of a fund-raising text* (pp. 295–325). Philadelphia/Amsterdam: John Benjamins.

Rakoczy, H. (2017). In defense of a developmental dogma: Children acquire propositional attitude folk psychology around age 4. *Synthese, 194,* 689–707.

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

Rocca, R., Wallentin, M., Vesper, C., & Tylén, K. (2018). *This* and *that* back in context: Grounding demonstrative reference in manual and social affordances. In *Proceedings of the 40th Annual Meeting of the Cognitive Science Society* (pp. 1522–1527).

Rocca, R., Wallentin, M., Vesper, C., & Tylén, K. (2019). This is for you: Social modulations of proximal versus distal space in collaborative interaction. *Scientific Reports, 9,* 14967.

Rozendaal, M. I., & Baker, A. E. (2008). A cross-linguistic investigation of the acquisition of the pragmatics of indefinite and definite reference in two-year-olds. *Journal of Child Language, 35*(4), 773–807.

Rozendaal, M., & Baker, A. (2010). The acquisition of reference: Pragmatic aspects and the influence of language input. *Journal of Pragmatics, 42,* 1866–1879.

Rubio-Fernandez, P. (2017). The director task: A test of Theory-of-Mind use or selective attention? *Psychonomic Bulletin & Review, 24*(4), 1121–1128.

Rubio-Fernandez, P. (2019). Theory of Mind. In C. Cummins & N. Katsos (Eds.), *The Oxford handbook of experimental semantics and pragmatics*. Oxford: OUP.

Rubio-Fernandez, P., Mollica, F., Ali, M. O., & Gibson, E. (2019). How do you know that? Automatic belief inferences in passing conversation. *Cognition, 193,* 104011.

Ruffman, T. (2014). To belief or not belief: Children's Theory of Mind. *Developmental Review, 34,* 265–293.

Scott-Phillips, T. (2014). *Speaking our minds*. London: Palgrave Macmillan.

Serratrice, L. (2008). The role of discourse and perceptual cues in the choice of referential expressions in English preschoolers, school-age children, and adults. *Language Learning and Development, 4,* 309–332.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27,* 623–656.

Shatz, M., Wellman, H. M., & Silber, S. (1983). The acquisition of mental verbs: A systematic investigation of the first reference to mental state. *Cognition, 14,* 301–321.

Skarabela, B. (2007). Signs of early social cognition in children's syntax: The case of joint attention in argument realization in child Inuktitut. *Lingua, 117,* 1837–1857.

Skarabela, B., & Allen, S. E. (2002). The role of joint attention in argument realization in child Inuktitut. In *Proceedings of the 26th annual Boston University conference on language development* (pp. 620–630). Somerville, MA: Cascadilla Press.

Skarabela, B., Allen, S. E., & Scott-Phillips, T. C. (2013). Joint attention helps explain why children omit new referents. *Journal of Pragmatics, 56,* 5–14.

Slobin, D. I. (1996). From "thought and language" to "thinking for speaking". In J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge, MA: Cambridge University Press.

Smith, K. (2018). The cognitive prerequisites for language: Insights from iterated learning. *Current Opinion in Behavioral Sciences, 21,* 154–160.

Sperber D. (Ed.). (2000). Metarepresentations in an evolutionary perspective. In *Metarepresentations: A multidisciplinary perspective* (pp. 117–137). Oxford, UK: Oxford University Press.

Sperber, D. & Wilson, D. (1986/1995). *Relevance: Communication and cognition*. Oxford: Basil Blackwell.

Stevens, J., & Zhang, Y. (2013). Relative distance and gaze in the use of entity-referring spatial demonstratives: An event-related potential study. *Journal of Neurolinguistics, 26,* 31–45.

Stevens, J., & Zhang, Y. (2014). Brain mechanisms for processing co-speech gesture: A cross-language study of spatial demonstratives. *Journal of Neurolinguistics, 30,* 27–47.

Strauss, S. (2002). *This*, *that*, and *it* in spoken American English: A demonstrative system of gradient focus. *Language Sciences, 24,* 131–152.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.

Tomasello, M. (2018). How children come to understand false beliefs: A shared intentionality account. *Proceedings of the National Academy of Sciences, 115*(34), 8491–8498.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences, 28*(5), 675–691.

Tomasello, M., & Haberl, K. (2003). Understanding attention: 12-and 18-month-olds know what is new for other persons. *Developmental Psychology, 39,* 906–912.

Tompkins, V., Farrar, M., & Montgomery, D. E. (2019). Speaking your mind: Language and narrative in young children's Theory of Mind development. *Advances in Child Development and Behavior, 56,* 109–140.

Webb, P. A., & Abrahamson, A. A. (1976). Stages of egocentrism in children's use of 'this' and 'that': A different point of view. *Journal of Child Language, 3,* 349–367.

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

Woensdregt, M., Cummis, C., & Smith, K. (2020). A computational model of the cultural co-evolution of language and mindreading. *Synthese.* Special issue on *The cultural evolution of human social cognition.*

Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews: Cognitive Science, 2,* 253–265.

Zimmermann, M. (2016). Information structure. In M. Aronoff (Ed.), *Oxford bibliographies in linguistics.* https://doi.org/10.1093/obo/9780199772810-0130.