

# Pragmatics and Language Evolution

Marieke Woensdregt & Kenny Smith

August 15, 2016

Word count: 9448

## Contents

<b>1</b>	<b>Main Essay</b>	<b>2</b>
1.1	Pragmatics and Mindreading . . . . .	2
1.2	What is pragmatic competence? . . . . .	4
1.3	Psychological mechanisms underlying pragmatic competence . . . . .	6
1.3.1	Minimal requirement for ostensive-inferential communication . . . . .	8
1.4	Pragmatic competence in great apes . . . . .	10
1.4.1	Mental state representations . . . . .	10
1.4.2	Intentional and ostensive communication in great apes . . . . .	12
1.5	The biological evolution of human pragmatic skills . . . . .	13
1.6	The cultural evolution of human pragmatic skills . . . . .	16
1.7	Have language and theory of mind co-evolved? . . . . .	18
1.8	Biological, cultural, and co-evolution of pragmatic competence . . . . .	19

## Summary

Pragmatics is the branch of linguistics that deals with language use in context. It looks for instance at the meaning linguistic utterances can have beyond their literal meaning (implicature), but also at presupposition and turn-taking in conversation. Thus, pragmatics lies on the interface between language and social cognition.

Doing pragmatics has widely been argued to require reasoning about the minds of others. A speaker has to think about what knowledge they share with the hearer in order to choose what information to explicitly encode in their utterance and what to leave implicit. A hearer has to make inferences about what the speaker meant based on the context, their knowledge about the speaker, and their knowledge of general conventions in language use. This ability to reason about the minds of others (usually referred to as ‘mindreading’ or ‘theory of mind’) is a cognitive capacity that is uniquely highly developed in humans compared to other animals.

This article will review what we know about how pragmatics (and the underlying ability to make inferences about others’ minds) has evolved. We will discuss the two main processes that can lead to the development of a complex behaviour over generations: biological evolution and cultural evolution, and explore to what extent they account for what we know about pragmatics.

In biological evolution, changes happen as a result of natural selection on traits that are transmitted genetically. In cultural evolution, on the other hand, modifications and improvements accumulate as skills are transmitted across generations through social learning. Many hypotheses have been put forward about the role that natural selection may have played in the evolution of social and communicative skills in humans (as a result of e.g. changes in food sources, foraging strategy or group size). The role of social learning and cumulative culture, however, has been often overlooked. This omission is particularly striking in the case of pragmatics, as language itself is a prime example of a culturally transmitted skill, and there is solid evidence that the pragmatic capacities which are so central to language use may themselves be partially shaped by social learning.

This article reviews the potential contributions of both evolutionary mechanisms (biological and cultural) to the evolution of pragmatics, in the light of empirical findings from comparative, developmental and experimental research. It also discusses the possibility of a co-evolution between the socio-cognitive skills underlying pragmatics and language (i.e. the vocabulary and grammar) itself.

## Keywords

– pragmatics, — ostensive-inferential communication, — primate communication, — theory of mind, — biological evolution, — cultural evolution, — co-evolution

## 1 Main Essay

### 1.1 Pragmatics and Mindreading

Being a competent language user does not just involve having access to a vocabulary and a grammar which are shared with others. It also involves knowing how to deploy those linguistic tools in order to achieve your communicative goals. This requires you to keep track of what your interlocutor knows and doesn't know, how their view on the world differs from your own and what is appropriate to say in a given situation. In other words, we take into account the context in which communication occurs and exploit its affordances to get our message across. The word 'context' here refers not only to the situation and physical surroundings, but also to the 'mental' context of the communicators, e.g. what they can see at this moment and also what they are likely to know or be interested in. The field of pragmatics is concerned with how we use such context when producing and interpreting linguistic utterances.

Deploying communicative signals flexibly depending on context is not unique to human communication. For instance, captive chimpanzees have been found to use modality flexibly based on the orientation and attentional state of their audience. Leavens et al. (2010) found that if a human experimenter was facing a chimpanzee, the latter would use gestures in order to request a specific food item; if, however, the experimenter was facing away, they used vocalisations to attract the experimenter's attention first. Chimpanzees in the wild have also been observed to adapt their signalling behaviour according to the composition of their audience. When attacked by a group member chimpanzees will normally scream in response, and the acoustic properties of their scream reflect the severity of the aggression, a correlation which nearby group members use to determine whether they should intervene in the fight or not. However, victims will also exaggerate the length and frequency of

their screams in response to mild aggression if they know that there's a high-ranking group member around who will be likely to help (Slocombe and Zuberbühler, 2007).

Thus, the ability to flexibly adapt signal choice and the way in which signals are used based on the context is a skill we share with at least our closest living relatives, and therefore presumably reflects cognitive capacities already present in our last common ancestor with chimpanzees. However, pragmatics in language involves a cognitive capacity that is more restricted in its distribution, and possibly unique to humans: the ability to adapt signal use based on knowledge or inferences of what goes on *inside the minds* of others. This is the part of pragmatics that is concerned with implicature and inference, which make use not of observable features of the physical or linguistic context, but of unobservable mental states.

An example of such implicature is the fact that the sentence “Ella got the car to stop” brings with it the implication that Ella did not simply hit the brakes, but got the car to stop in some more unusual fashion. This implicature is known as a *manner implicature* as it arises from Grice’s ‘Maxim of Manner’, which states that speakers should “Be perspicuous” and “be brief (avoid unnecessary prolixity)” (Grice, 1975, p. 308). The simplest way to say that Ella stopped the car by hitting the brakes would be to say “Ella stopped the car”. Since the speaker said something more elaborate (“Ella got the car to stop”), and thus would be violating the manner maxim if their intended meaning was that Ella hit the brakes, the hearer can infer that the speaker intended to communicate a more complex meaning: that Ella got the car to stop in an unusual way.

Understanding and using such implicature is qualitatively different from adapting one’s signal use to the composition of one’s audience or their attentional state, because it requires both speaker and hearer to reason about each other’s mental states, which are not directly observable but have to be inferred. Several researchers have argued that even the simplest exchanges in human language require several levels of embedded reasoning about mental states (Sperber and Wilson, 1986/95; Scott-Phillips, 2015a), and that this is what makes human language special when compared to the communication systems of other animals (Scott-Phillips, 2015b). Although this analysis of what everyday language use consists of is a matter of debate (see e.g. Moore, 2014, 2015, 2016a), it is not contested that natural linguistic exchanges between humans *can* involve complex inference-making and that this requires the ability to reason about the content of other’s minds (e.g. Moore, 2015). This ability is often referred to as *theory of mind*, *mindreading*, *metapsychology* or *mentalising* (see e.g. Baron-Cohen et al., 1985) — in this article we will use the terms *theory of mind* (abbreviated ToM) and *mindreading* interchangeably, simply because these are the most commonly used.

Humans are more proficient mindreaders than any other species. How has this pragmatic competence evolved? Is it a biological adaptation, and if so, what selection pressure has it evolved in response to? Or is it a product of cultural evolution, where skills are transmitted from generation to generation through social learning, accumulating improvements as they go? Or have culture and biology worked together to produce this unique capacity? Have the socio-cognitive abilities that underlie pragmatic competence in humans evolved for the purpose of language, or did they initially evolve for other purposes? Or have language and social cognition co-evolved, the one skill building on the other?

In order to answer these questions, we will start by providing an analysis of what human pragmatic competence consists of (section 1.2), followed by a breakdown of the psychological mechanisms involved (section 1.3). We will then go on to explore to which extent these psychological mechanisms are shared between humans and other primates (section 1.4), in order to identify which parts of pragmatic competence have evolved exclusively in

the *Homo* lineage. Subsequently, we will turn to theories of the evolution of the human-specific components of pragmatic competence. We will first review explanations involving biological adaptation (section 1.5), followed by explanations drawing on cultural evolution (section 1.6). Finally, we will discuss the possibility that the socio-cognitive skills underlying pragmatics have co-evolved with language itself (i.e. the conventional code) (section 1.7).

## 1.2 What is pragmatic competence?

Pragmatic competence is what allows an individual to look beyond the literal meaning of an utterance in order to determine the *speaker meaning*. Where literal meaning refers to the semantic concepts that are associated with the words and structure of a sentence, speaker meaning refers to the goal that the speaker has when they produce that sentence. This can be a goal to inform (“the entrance is on the other side of the building”); a request (“could you open the window?”) or general social bonding (“so sunny today!”).

The ability to infer a speaker’s intention behind an utterance obviously comes into play when interpreting deliberately non-literal language use, such as metaphors or sarcasm. But it is also necessary for interpreting a straightforward utterance such as “I’m tired”. Depending on the context, this sentence could mean anything from “Let’s have a coffee break” to “I don’t feel like talking about it” to “I’m thinking of quitting my job”, and so on. Thanks to this flexibility in use and interpretation, there may be an infinite set of potential speaker meanings for any given utterance in human language. This phenomenon is known as linguistic *underdeterminacy* (Carston, 2002, section 1.2). A hearer can resolve part of this underdeterminacy based on the context and the preceding conversation, the remainder must be disambiguated based on knowledge and inferences about the speaker’s mind.

The phenomenon of linguistic underdeterminacy illustrates that to analyse human communication we must go beyond what is known as the *code model* of communication (Shannon, 1948). In the code model, communication consists of a signaller encoding a message into a signal and a receiver decoding it to uncover the message (often by doing the inverse of the encoding operation). Communication systems that are sufficiently described by this model, sometimes known as *natural codes*, simply consist of pairs of associations, where the signaller has associations between states of the world and signals, and the receiver has associations between signals and responses. Many of the communication systems we find in nonhuman animals can be analysed in this way (Wharton, 2003). If the encoding and decoding operations in a natural code are properly tuned and there is no noise in transmission, the message that goes in at one end should be the same as what comes out the other. What this model cannot account for is the underdeterminacy of human language — where the same signal can have many different interpretations depending on the situational context, the linguistic context, the manner of delivery etc. A natural code is based on associations between signals and relevant phenomena in the world. A conventional code (like language) on the other hand, is made possible by associations between signals and inferred speaker meanings (see Wheeler’s commentary on Scott-Phillips, 2015b, p. 74)

The ability to make inferences about speaker intention is therefore an essential part of human language and our pragmatic competence. This requires theory of mind both on the part of the hearer and the part of the speaker. Building on an initial proposal by Grice (1957), Sperber and Wilson (1986/95) argue that any linguistic utterance contains the following two intentions:

“*Informative intention*: to inform the audience of something;

*Communicative intention*: to inform the audience of one’s informative intention”  
(Sperber and Wilson, 1995, p. 29)

The informative intention contains *what* the speaker wants to communicate, and the communicative intention contains *that* they want to communicate. Not every instance of language use involves an intention to share information however. Examples of this are “Stop tickling me!” (an intention to induce a certain behaviour) or “Look, an eagle!” (an intention to attract attention/share experience). To emphasise this point, Moore (2015) reformulates the two intentions of the speaker as follows (conceded by Sperber and Wilson (1986/95)):

1. an intention to produce a particular response in the hearer/audience  
and
2. an intention that the hearer/audience recognises intention 1  
(Moore, 2015)

Sperber and Wilson (1986/95) use the term *ostensive behaviour* or simply *ostension* to describe communicative behaviour that involves both these intentions; i.e. “behaviour which makes manifest an intention to make something manifest” (1995, p. 49). To capture both the ostension on the side of the speaker and the inference on the side of the hearer in a unified model of pragmatics, they coined the term *ostensive-inferential communication*. This model describes the type of communication we find in humans, as opposed to communication systems which can be described by the code model. Other models of communication have also been proposed (e.g. Gärdenfors, 2003), but the contrast between the code model and the ostensive-inferential model of communication suffices to outline the questions that this article is concerned with.

At this point there are two important things to note. Firstly, ostensive-inferential communication is something humans also do in non-linguistic communication. A tilt of the head or roll of the eyes are examples of ostensive behaviour that can make the receiver look for an informative intention (such as “Look, Uncle Steve is getting drunk again”) — and even completely novel, non-conventional gestures can be used to communicate ostensively, given that signaller and receiver share sufficient background. Secondly, the content of an informative intention can be recovered by a hearer even without recognising the encompassing communicative intention. This is especially the case in non-linguistic and non-conventionalised ostensive behaviour, such as moving someone’s phone into their line of sight to make sure they don’t forget it. The receiver of this signal may fulfil the signaller’s goal even without realising that the phone was moved there with an intention to signal something. However, the ability to recognise communicative intentions does make communication more efficient, because it points a hearer towards potentially relevant information. An act of ostension makes a receiver look for an informative intention — even if they do not directly see what the content of the informative intention is, recognising that there is a communicative intention will motivate them to spend cognitive resources on inferring it (Csibra, 2010). This is what Sperber and Wilson (1986/95) refer to as the ‘principle of relevance’.

Although in theory this type of ostensive-inferential communication could be highly standardized and code-like (see e.g. Csibra, 2010), in practice we see that humans can improvise ostensive signals on the fly and interpret utterances even if they are ambiguous and unexpected (Sperber and Wilson, 2002). This makes it highly likely that human communication involves some level of mental state attribution and thus theory of mind (ToM).<sup>1</sup> In order to

---

<sup>1</sup>Note however that Sperber and Wilson (2002) propose that humans have evolved a ‘comprehension

answer the question of how this pragmatic competence *evolved* however, we need a theoretical framework for analysing exactly what psychological processes are involved, and what the precursors of these might be.

### 1.3 Psychological mechanisms underlying pragmatic competence

A good place to start when trying to identify the requirements for pragmatic competence and their precursors is Dennett’s intentionality framework, which classifies the different levels of intentionality that can be ascribed to an organism (Dennett, 1983). A *zero-order* intentional system is in fact not an intentional system, because there are no mental states (such as beliefs and desires) behind the signal that the organism sends. The signal still counts as a signal however, because it is an adaptation that has evolved for the purpose of altering a receiver’s behaviour in a way that increases the sender’s fitness (Maynard Smith and Harper, 1995). An example of this kind of signal is *aposematism* (warning colouration), which we find for instance in poisonous frogs that have evolved a salient skin colour which warns predators of their toxicity: although this signal has a clear ‘message’ for the predator (“Don’t eat me!”), there is no intentionality on the side of the signaller (Summers and Clough, 2001).

A *first-order* intentional system is an organism that, in the words of Dennett (1983) “has beliefs and desires (etc.), but no beliefs and desires *about* beliefs and desires”. For communication this means that there is a mental representation underlying the signal, but no intention to modify another individual’s mental state. Signals that are sent with such first-order intentionality are often referred to as *functionally referential signals*. This term was coined in order to accommodate the fact that although signallers and receivers behave as if these signals refer to specific objects or events in the same way that human words do, the mental processes underlying the production and reception of these signals may be very different from those involved in human language (Scarantino, 2013). The classic example of this type of signalling system are the alarm calls of vervet monkeys (although many species have similar systems of alarm calls). Vervet monkeys have different calls for different predators, and on hearing a call group members will produce the corresponding evasive behaviour (Seyfarth et al., 1980). However, current consensus is that these calls are most likely produced as a direct response to observing a predator, rather than with the intention to inform others (Zuberbühler, 2013), i.e. they are more like a natural code than an instance of ostensive-inferential communication.

A *second-order* intentional system then is a system that also has beliefs and desires *about* the beliefs and desires of others. Dennett’s orders of intentionality can go up even further (*third-order* intentionality could be evidenced for instance in sentences such as “Steve wants Ella to believe that he did not know about the surprise party”), and every order from second-order intentionality upwards involves the ability to entertain *metarepresentations*, that is, to have representations *about* representations. This is something humans are remarkably good at — O’Grady et al. (2015) showed that adults can keep track of mental state representations up to seven levels deep. How many levels of metarepresentation are required to do ostensive-inferential communication is a question currently under debate, which we will discuss in section 1.3.1.

All levels of intentionality exceeding first-order intentionality require an ability to represent the mental states of others (beliefs about beliefs) and thus a ToM. Levels of ToM can

---

module’ that is dedicated directly to inferring informative intentions once a communicative intention is recognized, which may reduce the amount of mindreading required.

be counted in the same way as the orders of intentionality described above: first-order ToM is the ability to represent beliefs, second-order ToM the ability to have beliefs about beliefs, etc. (e.g. Baron-Cohen et al., 1997). A particularly well-studied kind of belief about belief is so-called *false belief* understanding (i.e. holding the belief that someone else has a belief which you know is not true). False belief understanding is special because it requires an understanding that other minds contain *representations* of the world that can be different from reality (Wellman et al., 2001). It thus requires the individual to represent another’s mental state in a way that is independent from their own representation of reality. As such, false belief understanding is often considered a hallmark of full-blown ToM capacity. In empirical studies of false belief understanding, a distinction is often made between explicit and implicit measures.

*Explicit* false belief understanding is measured in tasks where the participant has to give an explicit response based on their understanding of the false belief of another agent, e.g. by pointing to or saying in which location a story character will look for a toy according to their false belief. This requires a capacity to overtly reason about others’ mental states from a detached, third-person perspective (Helming et al., 2014). Human children only start succeeding at these explicit tasks around the age of four (Wellman et al., 2001).<sup>2</sup> In contrast, *implicit* false belief understanding is measured using gaze direction or looking times, in tasks which don’t require any explicit response or decision on the part of the participant. These tasks involve either measuring children’s anticipatory looks to a location where they expect a story character will search based on their false belief, or the amount of time the child spends looking when the story character searches for their toy in the location that was *unexpected* based on their false belief (with longer looking times indicating surprisal). This type of experiment has provided evidence that children are able to represent false belief-like states much earlier on, from as young as 7 months old (see Southgate et al. (2007) and Barrett et al. (2013) for the anticipatory looking paradigm and Onishi and Baillargeon (2005); Surian et al. (2007) and Kovács et al. (2010) for the violation-of-expectation paradigm).

Explanations of this discrepancy between when implicit and explicit false belief understanding become available can be divided into three kinds. Firstly, there is the account that human infants are able to represent false beliefs from very early on (perhaps even from birth) but that the ability to produce the correct explicit response requires inhibition and selection mechanisms that take several years to mature. For instance, Leslie et al. (2004) and Leslie (2005) argue that children have an innate mechanism for representing the mental states of others, but have learned as a default option that others’ beliefs about the world are the same as their own (also known as a ‘reality bias’). The development from an implicit to an explicit ToM ability then involves the development or maturation of a ‘selection process’ that allows children to select among the different belief states they have represented; until this selection process is fully developed, children fail to suppress their reality bias, leading them to give the wrong answer in a false belief task.<sup>3</sup> This first account is thus compatible with the view that *explicit* false belief tasks do not accurately reflect children’s abilities.

Secondly, there is the account that argues that it is the *representational* mechanism that has to mature, rather than the capacity to select between possible representations. For example, Rakoczy (2012) distinguishes between beliefs proper and *subdoxastic states*, which can be states like ‘has an inclination to think that’ or ‘will be likely to behave as if

<sup>2</sup>Although see Rubio-Fernández and Geurts (2012) for evidence that a different phrasing of the task allows children to pass it at three years old.

<sup>3</sup>See Helming et al. (2014) for a discussion of two other biases that may cause children to give the wrong response in explicit tasks (cooperative bias and referential bias).

she believes that’. To take the looking time study of Onishi and Baillargeon (2005) as an example, a representation of the form ‘The experimenter will have an inclination to think that the toy is in the yellow box’ would produce the same results as a representation of the form ‘The experimenter believes that the toy is in the yellow box’, and the same is true for experiments using a gaze direction or active helping paradigm. According to this account, subdoxastic states are different from proper beliefs because (i) they cannot be integrated with informational states from other areas of cognition and (ii) they are not accessible to conscious introspection, meaning that a child holding such representations would fail to produce the correct response in explicit (but not implicit) tests of false belief understanding. This second account is thus compatible with the view that *implicit* false belief tasks are not testing full-blown ToM ability.

Thirdly and finally, there is the two-systems account, which argues that implicit and explicit false belief tasks measure two separate systems that are both part of the full-blown human ToM capacity but that develop in different ways and at different ages. For instance, Apperly and Butterfill (2009) argue that later-developing, explicit false-belief understanding is a result of flexible cognitive processes that depend in their development on language and executive functions, whereas early, implicit false-belief understanding is a result of a set of less flexible, cognitively efficient processes that are available before language and executive functions develop. Given this hypothesis, Apperly and Butterfill predict that early, implicit ToM is likely to be limited in rather arbitrary ways, both in terms of the type of ‘content’ that can be represented (e.g. ‘that the toy is in the yellow box’ versus ‘that Ella doesn’t know that Steve was not really ill’) and the type of ‘psychological roles’ that can be attributed (e.g. ‘x *believes* y’ vs. ‘x *thinks* y’ vs. ‘x *desires* y’ etc.). This third account is thus compatible with the view that both implicit and explicit false belief tasks accurately measure some part of children’s ToM, but that they tap into two different underlying systems.

When it comes to ostensive-inferential communication, there are two ToM abilities that have been argued to be necessary: i) the ability to entertain metarepresentations and ii) the ability to represent beliefs (as opposed to subdoxastic states) (see Sperber, 2000; Sperber and Wilson, 2002; Tomasello, 2008; and Scott-Phillips, 2014 for the metarepresentations claim, and Breheny, 2006 for the beliefs claim). These arguments have subsequently been used to claim that this type of communication is unique to humans (Sperber, 2000; Tomasello, 2008; Scott-Phillips, 2014). However, in recent years there have been moves to re-examine whether human communication necessarily involves such sophisticated mental operations, or whether the *minimal* cognitive requirements for doing pragmatics might be less demanding.

### 1.3.1 Minimal requirement for ostensive-inferential communication

For instance, Moore (2016a) argues that to understand informative communicative intentions it is often sufficient to distinguish knowing from not knowing, and it is not necessary to have an understanding of false beliefs. To use an example of Moore (in turn adopted from Tomasello (2008)): if a sender makes a digging motion towards the ground to signal that there are likely to be tubers to dig for, this motion would be communicative in the original definition of Sperber and Wilson (1986/95) only when the sender has the intention to make the receiver *believe* that there are tubers in that patch of ground. However, Moore argues that for the sender to have the intention that the receiver should *attend to*, *see* or *recognise* the presence of tubers would have the same effect and would make the signal no less communicative or intentional. Holding the intention that someone *attends to/sees/recognises* the presence of tubers requires at most an ability to represent a ‘registration’ or ‘aware-



ness’ relation between that individual and a piece of information, which is less cognitively demanding than representing a belief (i.e. a ‘propositional attitude’ or ‘representational relation’ that can be false) (Apperly and Butterfill, 2009; Martin and Santos, 2016). The same argument holds from the point of view of the receiver. Say if a fully ostensive sender has the intention to make the receiver *believe* (i.e. non-factual) that there are tubers in this particular patch of ground, and the receiver understands this rather as the sender having the intention to make her *recognise* (i.e. factual) the presence of tubers, this will still produce the same behavioural response. This would also explain why infants seem to be able to recognise communicative intentions (Csibra, 2010), without having to posit that they can already represent abstract mental states like beliefs.

Secondly, the view of ostensive-inferential communication put forward by Sperber (2000) posits that communication requires the ability to entertain *fourth-order* metarepresentations as the one depicted below:

fourth order:	You intend
third order:	That I believe
second order:	That you intend
first order:	That I believe
representation:	That there are tubers for which we could dig.

However, Moore (2016a) argues that ostensive-inferential communication consists of two *functionally distinct* components. The first component is the act of sending a signal (with the intention of invoking a certain behavioural response in the receiver), and the second component is the act of attracting attention towards the ‘signalhood’ of that signal (with the intention of the receiver recognising the first intention). Moore calls this first component the *sign production* and the second component the *act of address* (similar to the aforementioned act of ostension). Based on this separation, Moore claims that the act of sign production and act of address by themselves only require (at most) second-order metarepresentations. For the act of sign production, a representation like the one below would suffice:

second order:	<i>S</i> intends that
first order:	<i>H</i> recognise that
representation:	There are tubers for which they could dig.

And for the act of address, the sender would (maximally) need to entertain a representation like the following:

second order:	<i>S</i> intends that
first order:	<i>H</i> see that
representation:	<i>S</i> is addressing to him an action <i>x</i> .

Moore goes on to argue that in most cases, even lower-order representations will suffice. For example, from the perspective of the sender it is not necessary to *explicitly* represent the second order of either of the above two metarepresentations. The sender only needs to *have* these intentions, she does not need to be aware of them. From the receiver’s perspective, second-order metarepresentations are necessary to understand the act of sign production in cases of *declarative* communication (i.e. information-sharing), but to understand *imperative* communication (e.g. requests) first-order metarepresentations would suffice.

Subdoxastic states and second-order metarepresentations could thus be potential precursors of the full-blown pragmatic competence we find in humans, and could even turn

out to be sufficient for doing some of our everyday linguistic communication. However, it seems likely that human language use *can* involve representations of proper belief states and fourth-order metarepresentations, at least sometimes (Frank and Goodman, 2012; Franke and Jäger, 2014). Before moving on to the question of how these representational skills have evolved, we will first review to which extent their precursors are present in other primates.

## 1.4 Pragmatic competence in great apes

Comparative research is a good place to start when studying the evolution of a species-specific trait, because it offers valuable insights into the starting point from which the trait of interest evolved (Nunn, 2011). If precursors of the trait are present in related species, it is likely that those were already present in their last common ancestor with the species under investigation, and thus do not require a species-specific evolutionary account. In the case of human pragmatic competence therefore, the question we need to ask before theorising about its evolution is what parts of this trait we share with other primates.

Here we will limit our discussion to the nonhuman great apes (i.e. *Hominidae* — orangutans, gorillas, chimpanzees and bonobos), because they are our closest living relatives and because most research on intentionality in nonhuman communication has focused on these species. We will first discuss the findings regarding ToM abilities in great apes, followed by the evidence that they employ these in their communication.

### 1.4.1 Mental state representations

Most studies of great ape ToM have been conducted with captive chimpanzees. For instance, Kaminski et al. (2008) explored ToM in a task in which two chimpanzees compete over food rewards. The chimpanzees were positioned opposite each other in separate enclosures, with a table with three cups placed in between them. In each trial one of the chimpanzees (the ‘subject’) observed an experimenter placing food rewards in two of the three cups. The other chimpanzee (the ‘competitor’) either also witnessed the baiting of all cups, or of only one of them (in which case their view was occluded by an opaque panel during the baiting of one of the cups). Subsequently, both chimpanzees were allowed to choose one of the cups and receive its reward: either the subject got to choose first, or the competitor chose first (and the subject’s sight was occluded while the competitor made their choice).

Kaminski et al. found that when the competitor only saw the baiting of one of the cups, and the subject got to choose second, the subjects more often chose the ‘unknown’ reward (the one not witnessed by the competitor) than the ‘known’ reward. In contrast, when the subject was allowed to choose first, they were equally likely to go for the known and the unknown reward. Kaminski et al. concluded that chimpanzees can represent what others know based on what they have seen, and predict their behaviour accordingly.<sup>4</sup> The chimps behaved as if they knew that if the competitor only knows the location of one of the rewards they are likely to pick that one, which means that when choosing second the subject would be better to go for the reward that was unknown to the competitor. Kaminski et al. also conducted the same experiment with human children (mean age 6) and adults, and found a similar pattern of results.

In a second false belief task, Kaminski et al. used the same set-up, but added a ‘lift’ and ‘shift’ event, where after the initial baiting of the cups the reward was either lifted

<sup>4</sup>However see Penn and Povinelli (2007) for a discussion of alternative interpretations of such findings in terms of behaviour-rules.

and replaced in the same cup (lift condition), or lifted and replaced in a different cup (shift condition). This lift or shift event was either witnessed by both participants, or by the subject only. In addition, Kaminski et al. now made the two rewards different in quality: one regarded very desirable by both participants and one regarded less desirable.

When running this experiment with six-year-old children, Kaminski et al. found that in the condition where the subject got to choose second, they picked the high-quality reward more often than the low-quality reward in the ‘unknown shift’ condition (where the shift had not been witnessed by the competitor) but not in the ‘unknown lift’ condition. This shows that the children were able to distinguish between the condition where the competitor’s belief about the high-quality reward was still accurate (‘unknown lift’) and the condition where the competitor’s belief had been rendered false (‘unknown shift’). Chimpanzees on the other hand did not act differently in these two different conditions: in both cases they went for the cup containing the high-quality reward slightly more often than the cup with the low-quality reward.

Krachun et al. (2009) elaborated on this study, using a similar competitive set-up but testing both chimpanzees and bonobos, and measuring looking times in addition to explicit choice responses to see if apes do show implicit signs of false belief understanding. In this study there were only two cups and one reward, and the competitor was a human experimenter who either had a true or false belief about the location of the reward. The human competitor got to choose first in each condition, but in the crucial trials they intentionally did not manage to reach the cup in time before the table was moved over for the ape subject to make their choice. If the subjects were able to represent the competitor’s false belief and predict her behaviour accordingly, they could use the competitor’s unsuccessful reach as an indicator of the reward’s location (the ‘reached-for’ cup in the true-belief case; the other cup in the false-belief case). As one would expect based on the results of Kaminski et al. (2008), the apes’ explicit choice responses in these two conditions were not significantly different: in both cases they selected the ‘reached-for’ cup (resulting in a reward in the true belief condition, and no reward in the false belief condition). Looking times however revealed a different pattern: subjects did *look* longer at the unchosen cup before making their choice in the false belief condition than in the true belief condition. This may indicate some awareness of the competitor’s false belief, even if the subject was not able to use this for deciding which cup to choose. This could be either because these apes lack the necessary inhibition to suppress the tendency to go for the ‘reached-for’ cup (an explanation in line with the failure-to-inhibit account of Leslie et al., 2004), or because their false belief representations are too ‘subdoxastic’ to be integrated with the rest of their behaviour-prediction procedures (following Rakoczy, 2012). When testing 4.5- to 5-year-old children on the same task, Krachun et al. (2009) found that they *did* respond as if they understood that the experimenter had a false belief: choosing the reached-for cup in the true belief condition and the other cup in the false belief condition.

Although caution should always be exercised in drawing conclusions from the relatively small number of studies that have been conducted on the ToM abilities of great apes, and absence of evidence cannot be taken as evidence of absence (especially not in primatology experiments, which are methodologically extremely challenging), we can tentatively conclude that great ape cognition includes the ability to represent mental states, but that these representations may fall short of proper beliefs.<sup>5</sup> As far as we are aware however, a study in the same vein as Kaminski et al. (2008) and Krachun et al. (2009) has not yet been run with

---

<sup>5</sup>See Martin and Santos (2016) for another classification in terms of ‘awareness relations’ versus ‘representational relations’.

human infants. Therefore, it is as yet unclear to what extent the difference in performance on these experiments between great apes and human children is due to a difference in biology and to what extent it is due to a difference in cultural input. Based on the current evidence, we can conclude that great apes have the beginnings of some of the cognitive capacities putatively involved in ostensive-inferential communication, but probably not at the same level of sophistication as seen in humans above the age of five. These beginnings however may be just enough to fulfil the minimal requirements for ostensive-inferential communication as defined by Moore (2016a), described in the previous section.

#### 1.4.2 Intentional and ostensive communication in great apes

A second, related question is to what extent great apes employ these ToM-like capacities in their communication. Most studies of ‘primate pragmatics’ have focused on the question whether great apes produce their signals (be it gestures or vocalisations) intentionally, i.e. exhibiting second-order intentionality, with an informative intention. This is different from the question whether great ape communication is ostensive, because ostension also requires a communicative intention, or in other words, ‘overt intentionality’. Liebal et al. (2014, sections 8.3-8.6) give an extensive overview of the different indicators of intentionality that have been adopted in studies of primate communication, and categorise some of these as strong and some as weak. The four weak criteria are (i) social use; (ii) visual-orienting behaviour or gaze alternation; (iii) response-waiting; and (iv) flexibility. The three strong criteria are (v) the production of a signal selectively for certain individuals in an audience (a subclass of ‘social use’); (vi) the production of a signal only when the intended receiver is already attending to the signaller, or actively manipulating the attention of the receiver; and (vii) persistence and elaboration of the signal when the communicative goal is not or only partially met. Active manipulation of the receiver’s attention (part of criterion (vi)) can also be viewed as an indicator of ostension, because it serves to draw attention to the fact that there is an informative intention; i.e. it serves to ‘signal the signalhood’ (Scott-Phillips, 2015b). The same has been argued for eye contact (part of criterion (ii)) (Gómez, 1994, 2007).

The most compelling evidence that great apes can have informative intentions when communicating comes from studies of chimpanzees’ vocalisations. Elaborating on an experimental design by Crockford et al. (2012) using a model of a viper snake (a predator much feared by wild chimpanzees), Schel et al. (2013b) evoked alarm calls from chimps travelling in groups through the forest. They found that at least some types of alarm calls that the chimps produced in these episodes satisfy strong criteria for intentionality (criteria (v) social use and (vii) persistence), and one weaker criterion (visual-orienting behaviour or gaze alternation, in that the alarm-calling chimp will alternate looking between the snake model and their audience). In a second study focusing on chimpanzees’ food calls, Schel et al. (2013a) investigated whether these calls are directed at specific other individuals or not. The results of this study showed that feeding chimps were significantly more likely to produce ‘rough grunts’ (a food call interpreted as a generic ‘invitation to come and eat’) for higher-ranking individuals and good friends than for others, and looked in the direction from which they expected the intended audience to appear while vocalising.

These two studies provide the strongest evidence to date that non-human primates have something that looks like informative intentions in their natural communication. Informative intentions are of course only part of what it means to do ostensive-inferential communication, and the presence of informative intentions do not imply the presence of communicative

intentions (see e.g. Bar-On, 2013). Instead, the best indicator for a communicative intention is ostensive behaviour. The chimpanzees of Schel et al. (2013b) showed some of this in their persistence behaviour — an alarm-calling chimp would persist in alarm calling until their audience was safe — but to our knowledge no studies of primate communication have been conducted focusing specifically on ostensive behaviour.

Moore (2015) specifically reviews the possibility and occurrence of ostension in the gestural communication of great apes, and uses strong criterion (vi) (‘deliberately solicit[ing] the attention of others before gesturing’) as the indicator, citing two findings of such behaviour. Firstly, Povinelli et al. (2003) found that chimpanzees change the location of their gestures to make sure they are in the line of sight of a human experimenter. Secondly, Liebal et al. (2004) found that all four species of great apes moved into the line of sight of a human experimenter before gesturing to request food — chimpanzees and bonobos doing so even when they had to move away from the food in order to get in front of the experimenter. If moving oneself and one’s gestures deliberately into the line of sight of an interlocutor is taken as an act of intentionally drawing the receiver’s attention to the sign, these findings can be interpreted as acts of ostension.

Overall we can conclude that great apes do indeed use their limited understanding of mental states in their communication; producing signals with an informative intention (Schel et al., 2013b,a) and showing some signs of ostensive behaviour — at least in the case of captive apes communicating with human experimenters (Povinelli et al., 2003; Liebal et al., 2004).

## 1.5 The biological evolution of human pragmatic skills

So far we have seen that human pragmatic competence involves sophisticated ToM skills which allow humans to represent others’ beliefs in a way that is decoupled from their own representation of the world (e.g. Liu et al., 2004) and to entertain such representations up to several levels of embedding (i.e. metarepresentations) (O’Grady et al., 2015). Our closest primate relatives (great apes) share some precursors of these skills, including the ability to represent what others know (Call and Tomasello, 2008; Kaminski et al., 2008; Krachun et al., 2009) and perhaps some implicit awareness of beliefs (Krachun et al., 2009). Evidence has also been found that great apes put these abilities to use in their communication, both in captivity and in the wild (Schel et al., 2013b,a; Povinelli et al., 2003; Liebal et al., 2004). Discussion is ongoing about whether or not this qualifies as ostensive communication proper (Scott-Phillips, 2015b; Moore, 2015), but we will now turn to theories of how the *Homo* lineage got from this rather limited pragmatic competence to the pragmatic competence we find in humans today — specifically, the flexible usage of the ability to hold and recognise informative and communicative intentions, which allows for the use of highly ambiguous utterances and improvised ostensive signals. In the current section we focus on explanations involving biological evolution, and in section 1.6 we review explanations involving cultural evolution.

Biological evolution works with naturally occurring variation in traits that are transmitted genetically from generation to generation. The genes underpinning a particular trait are selected for if that trait increases the fitness (i.e. number of offspring) of an individual bearing that trait, relative to other competing traits. The best evidence that a trait has evolved by this route is of course to find the genes that code for the trait in question, and to identify the signals of selection in their distribution within and across populations. However, complex cognitive skills like those involved in ToM are probably reliant on many different

genes interacting with each other and the environment, making it hard to identify the genes involved (although see Xia et al., 2012, for a first attempt). As a result, other indicators are often used to try to work out if a given trait is genetically encoded and therefore potentially a target of natural selection, including: whether or not the trait in question comes online early on in infancy (indicating relatively little role for learning and therefore increasing the likelihood that the trait is largely determined genetically); whether it develops similarly in different individuals and different environments (again indicating a limited role for learning from experience); and whether there is a specialised neural substrate for the trait that can be selectively impaired (suggesting that the trait has relatively direct genetic underpinnings).

For ToM, the looking time studies of Onishi and Baillargeon (2005); Surian et al. (2007) and Kovács et al. (2010) suggest that infants are able to represent false belief-like states from as young as 7 months old, and a gaze-direction study of Barrett et al. (2013) suggests that implicit false-belief understanding in young children (1 to 4 years old) is similar across many different cultures. Together, these studies suggest that these capacities might be relatively experience-independent and therefore strongly constrained by genetics. In addition, neuroimaging studies of both typically-developing adults and individuals with autism and other psychopathology suggest that humans have a brain network dedicated to ToM, which can be selectively impaired either from birth (as is the case in autism) or through brain injury later in life (see Brüne and Brüne-Cohrs, 2006, for a review). These neurological findings suggest that ToM has a relatively clear biological and genetic basis without which it cannot develop normally. However, cross-cultural studies of the developmental stages of mental state understanding (from 3 to 9 years old) show that cultural environment does have an influence, at least on the order in which different aspects of ToM are acquired (see Slaughter and Perez-Zapata, 2014, for a review). In addition, a twin study by Hughes et al. (2005) suggests that the majority of variance in ToM skills among individuals is explained by environmental rather than genetic factors. Thus, environment and learning contribute to ToM development as well.

Taken together, these observations suggest that at least some components of ToM are genetically transmitted and thus biologically evolved. Since these capacities seem uniquely well-developed in humans, this prompts the question of what selective pressures drove the elaboration of ToM and/or pragmatic capacities in our lineage — i.e. what selective advantages would come from the ability to reason about the mental states of others?

Most accounts of how the biological underpinnings of pragmatic competence evolved in humans agree on the point that these evolved *before* language itself (i.e. the conventional code with vocabulary and grammar) existed (Sperber, 2000; Tomasello, 2008; Csibra and Gergely, 2011; Scott-Phillips, 2014, 2015b).<sup>6</sup> In this ‘pragmatics-first’ view of language evolution, the ToM abilities that make up pragmatic competence initially evolved not for the purpose of language, but to serve some other function. Once this other pressure led to the improvement of ToM and/or metarepresentational abilities, these skills were then re-appropriated by language. Or, in the words of Scott-Phillips (2015b), language “is made possible by mechanisms of metapsychology and is made powerful by mechanisms of association” (Scott-Phillips, 2015b, p. 64) (where ‘mechanisms of association’ refers to the ability to establish a conventional code where arbitrary vocalisations or gestures are associated with particular meanings, i.e. a vocabulary). This pragmatics-first account is reminiscent of the evolutionary process known as *exaptation*, where a particular trait gets co-opted for a use that is different from the one it was originally selected for (Gould and Vrba, 1982).

<sup>6</sup>Although see Bar-On (2016) for a different account in which language and pragmatic ability evolved more in lockstep.

The question then becomes why and how the ToM abilities underlying pragmatic competence evolved, if it was not for language. Most theories that try to explain the remarkable social intelligence we find in primates, and humans especially, place its source in our increasingly complex social lives (e.g. Byrne, 1996; Burkart et al., 2009; Sterelny, 2012; Tomasello et al., 2012; Whiten and Erdal, 2012). The advantages that full-blown ToM brings to such lives are an increased ability to predict and manipulate each other’s behaviour, and an increased ability for cooperation. The hypothesis that human social cognition has evolved for the purpose of cooperation has been put forward by e.g. Sterelny (2012); Tomasello et al. (2012); Whiten and Erdal (2012). The essential idea that these theories have in common is that there is something special about the hunter-gatherer lifestyle that hominins adopted during the Pleistocene which made cooperation and honest information sharing beneficial enough to be selected for by biological evolution.

Because cooperating and sharing information are acts of trust that come at the risk of being exploited (e.g. Ale et al., 2013), there are certain conditions that have to be met in order for cooperation to become adaptive (i.e. to constitute a selective advantage) (Sterelny, 2012). Firstly, cooperation should come with a relatively high benefit and low cost. Secondly, individuals need to interact repeatedly in order to build up relations of reciprocal helping, allowing individuals to build up social alliances. Thirdly, there should be a mechanism for detecting so-called ‘free-riders’ (individuals who benefit without contributing). And finally, there should be a way of punishing these free-riders that is not too costly when compared to the benefits of cooperation. Sterelny (2012) and Tomasello et al. (2012) argue that these conditions were met when due to a change in ecology hominins in the Pleistocene started foraging collaboratively.

Collaborative foraging (such as big-game hunting) can only work if a group of individuals works together towards a joint goal and shares the spoils fairly.<sup>7</sup> Sterelny (2012), Tomasello et al. (2012), and Whiten and Erdal (2012) argue that this requires a ToM ability that is more sophisticated than what we find in great apes today, and that the selective advantage for (groups of) individuals who possessed such ability would have been strong enough for this trait to lead to more offspring. Aside from working together towards a joint goal, such improved ToM abilities would allow these early hominins to communicate more effectively; enabling them to work together on perfecting skills and tool use, and passing these on from generation to generation.<sup>8</sup>

This ability to pass on knowledge and skills from generation to generation by itself has also been argued to be the main selective pressure that has led to the sophisticated ToM ability and communication we find in humans. This idea is outlined in Csibra and Gergely’s (2011) ‘Natural Pedagogy’ hypothesis, which states that humans are born with a “well-organised package of biases, tendencies and skills” (Csibra and Gergely, 2006, p. 8) that makes human infants particularly receptive to teaching. Specifically, this package includes the implicit ToM abilities that allow infants to recognise communicative intentions from very early on, through a special sensitivity to ostensive behaviour (such as eye contact, infant-directed speech and contingent reactivity) (Csibra, 2010). Csibra and Gergely (2011) argue that this ‘natural pedagogy package’ is transmitted genetically and that it evolved

<sup>7</sup>What sets this type of foraging apart from the group hunting we see in e.g. lions and orcas is that collaborative foraging refers to a situation where (i) individuals *have to* collaborate in order to benefit; (ii) the yield of a collaboration has to be greater than any solo foraging alternative; and (iii) any alternative solo foraging has to be abandoned (risked) in order to collaborate. These three criteria are also what make up the ‘Stage Hunt’ game in game theory (Skyrms, 2004).

<sup>8</sup>Although Moore (2016b) and others argue that ostensive-inferential communication does not require cooperation.

as a biological adaptation for teaching and cultural transmission. The argument here is that as hominins developed skills and artefacts that became increasingly sophisticated and increasingly opaque in terms of their means-end relation, teaching became more and more important to enable reliable transmission of these skills and cultural practices. Such cultural transmission was important for evolving tool use and cooking practices, which both had a clear selective advantage for humans (see respectively Stout (2011) and Wrangham and Carmody (2010)).

To conclude, there may be certain ToM skills that have evolved specifically in humans because they formed biological adaptations to the hunter-gatherer lifestyle that our ancestors adopted during the Pleistocene. Two possible sources that gave rise to a selection pressure that resulted in abilities needed for ostensive-inferential communication are cooperation and cultural transmission, both of which benefit from an increased ability to represent intentions (both individual and shared) and to engage in ostensive communication. Interestingly, the second of these two adaptations — cultural transmission — in turn unlocks a much more rapid and flexible mechanism for adaptation: cultural evolution.

## 1.6 The cultural evolution of human pragmatic skills

Many systems of human knowledge and behaviour are culturally transmitted — passed on from generation to generation through social learning, rather than via genes. Cultural transmission leads to cultural evolution, where knowledge and skills accumulate over time, and adapt rapidly to the demands of both the environment and the minds through which they are transmitted (Henrich and McElreath, 2003). Humans are by far the most pervasively cultural species on the planet, and language (one of our many socially-learned behaviours) is one of our most striking cultural feats (Smith and Kirby, 2008; Thompson et al., 2016). Could our unusually developed capacity for reasoning about mental states in others also be a product of cultural evolution?

Heyes (2012b) and Heyes and Frith (2014) review evidence from experimental, developmental and neurocognitive studies showing that social learning plays a role in the development of ToM, suggesting that ToM is (at least in part) a product of cultural evolution. Firstly, as mentioned in the previous section, Hughes et al. (2005) found in a longitudinal twin-study that individual differences in mental state understanding are strongly correlated with verbal ability, and that this correlation is for the most part explained by environmental (rather than genetic) influences. In addition, Hughes et al. (2005) present indirect evidence that these environmental factors are composed largely of discourse with parents and siblings. Secondly, Heyes and Frith review several studies showing that children’s ToM development is predicted by their parents’ use of mental state terms and causal-explanatory statements about the mind (e.g. ‘She is smiling because she is happy’) (Meins et al., 2002; Taumoepeau and Ruffman, 2006; Slaughter et al., 2007; Taumoepeau and Ruffman, 2008). Thirdly, the combined findings of Taumoepeau and Ruffman (2006) and Taumoepeau and Ruffman (2008) also provide tentative evidence that parents (consciously or unconsciously) control their mental state discourse in such a way that they *tailor* it to the ToM abilities of their children. Taken together, these findings show a tight coupling between discourse about mental states and a child’s ToM development.

In addition, Russell et al. (2011) compared great apes (chimpanzees and bonobos) who were reared in standard captivity environments (zoos and laboratories) to great apes reared in rich socio-communicative environments (ape language projects), to see how much influence socio-communicative training by humans could have on great apes’ social cognition. The



standard-reared apes in this study received only the necessary human interactions involved in feeding and other animal husbandry. The ‘enculturated’ apes on the other hand had received extensive socio-communicative input from humans in the form of language training (training the comprehension of spoken language using specially designed ‘lexigrams’), although not all apes included in the study had been equally successful at this task. The results of this study showed that where the standard-reared apes performed worse on social cognition tasks (assessing communicative skills and understanding of attentional state and eye-gaze) than on physical cognition tasks, this difference was not present in the enculturated apes. Moreover, when compared to the performance of 2.5-year-old children on the same task, tested in a study by Herrmann et al. (2007), the enculturated ape group performed similarly to the children on the social cognition tasks, and even outperformed them on a task assessing understanding of the attentional state of an experimenter. Although the results of the standard-reared apes were not hugely different, they performed worse than the children on the task assessing the production of communicative signals and did not outperform the children in any of the other social cognition tasks. Similar results were found in a study by Lyn et al. (2010) looking at great apes’ ability to understand declarative signals (pointing and vocalisations). In this study enculturated chimpanzees and bonobos were found to significantly outperform their standard-reared counterparts in their comprehension of ostensive points and vocalisations produced by human experimenters. The studies by Russell et al. (2011) and Lyn et al. (2010) thus show that environment can make a difference in the development of social cognition in great apes just like it does in humans.

This suggests a role for cumulative culture in the evolution of ToM. Although there might be a biological basis for ToM development that all humans share, the more sophisticated ToM abilities — such as higher-order metarepresentations and proper ‘representational’/‘propositional’ representations of mental states — may depend on cultural transmission. Heyes and Frith (2014) refer to these two parts of ToM as ‘implicit’ and ‘explicit’ ToM (echoing the conclusions of e.g. Kaminski et al., 2008; Krachun et al., 2009; Rakoczy, 2012). Implicit ToM skills in this framework refers to the abilities responsible for the tracking of belief-like states found in infants by Onishi and Baillargeon (2005); Surian et al. (2007) and Kovács et al. (2010). These include gaze-following and joint attention, which develop early on in infancy and are shared with other great apes (and thus presumably part of our genetic endowment). Explicit ToM abilities on the other hand refers to that which allows humans to use their representations of others’ mental states *explicitly*, both in reasoning and behaviour — this requires mental state representations that are independent from the individual’s own representation of reality (i.e. so-called ‘representational’ or ‘propositional’ representations) (e.g. Apperly and Butterfill, 2009; Rakoczy, 2012; Kamps et al., 2013). Based on the evidence summarised above, Heyes and Frith (2014) argue that these explicit ToM abilities develop through social learning rather than the maturation of innate cognitive modules.

As briefly mentioned above, the power of cultural evolution is that it enables rapid accumulation of skills — where each generation can add some sophistication to the cognitive constructs that they get handed from the previous generation. In the case of explicit ToM abilities, this could take the form of increasingly elaborate, socially-transmitted practices of discussing and reasoning about the mental states in others — also known as *folk psychology*. However, it is hard to imagine how such discussion and teaching about mental states would happen without language; especially considering the fact that all studies reviewed above as evidence for social learning of ToM place emphasis on the role of discourse with parents and siblings (Meins et al., 2002; Hughes et al., 2005; Taumoepeau and Ruffman, 2006; Slaughter

et al., 2007; Taumoepeau and Ruffman, 2008). This leads to an interesting final hypothesis about the evolution of pragmatic competence: that ToM and language (in the sense of the conventional code with vocabulary and grammar) have *co-evolved*.

## 1.7 Have language and theory of mind co-evolved?

The hypothesis that ToM and linguistic communication have co-evolved played at least some role in all theories of the evolution of human social cognition described in section 1.5 (Csibra and Gergely, 2011; Sterelny, 2012; Tomasello et al., 2012; Whiten and Erdal, 2012), and has been fleshed out more elaborately by Malle (2002) and Moore (2016a). However, it is hard to find evidence for such scenarios of how cognitive skills evolved, since our ancestors in the *Homo* lineage have gone extinct and minds do not leave fossils. There are several types of indirect evidence that can be collected to test hypotheses like these (see e.g. Heyes, 2012a) however, one of which is evidence for co-development; if the development of one skill (e.g. explicit mindreading) is dependent on the development of another (e.g. language), the former could not have developed to the same extent when the latter had not yet evolved.

There is persuasive evidence consistent with the hypothesis that language and ToM co-develop. Firstly, evidence that language-learning depends on ToM abilities is provided by Parish-Morris et al. (2007). In a study comparing children with autism to typically-developing children, they showed that although 5-year-old autistic children have some ability to use social cues (pointing and eye gaze) to direct their attention in word learning, they perform at chance when learning new words required inferring the speaker’s intention, unlike language- and mental-age-matched typically-developing children.

Secondly, the reverse phenomenon has also been observed, namely that the development of ToM depends in part on having access to language. Deaf children of hearing parents, who lack consistent linguistic input during the first years of their life, were shown to have delayed ToM development relative to deaf children of deaf parents, who receive sign language input from birth (Schick et al., 2007). Similarly, a study with typically-developing children showed that simply training children on the use of mental state verbs with sentential complements accelerated their false belief understanding (Lohmann and Tomasello, 2003).

Thirdly, in a study comparing different age-groups of signers of the recently emerged Nicaraguan Sign Language, Pyers and Senghas (2009) showed that the bootstrap effect of language on ToM development continues on into adulthood. Pyers and Senghas found that the first cohort of signers (mean age 27), whose language had very limited mental state vocabulary, were worse at understanding false belief than the second cohort (mean age 17) who had more words for mental states. Moreover, a follow-up study two years later revealed that the first-cohort signers had improved in their false belief understanding and that this either followed or co-occurred with, but never preceded, an expansion of mental state vocabulary.

Finally, a recent longitudinal study by Brooks and Meltzoff (2015) provides direct evidence that language and ToM co-develop. They showed that gaze-following in 10.5-month-old infants predicted their production of mental state terms at 2.5-years-old, and that these mental state terms in turn predicted the extent of their false belief understanding at 4.5-years-old, even though gaze-following did not directly predict false belief understanding. Thus, this shows evidence of an indirect relation between early sensitivity to social cues and later ToM ability, mediated by language.

Recent work by Woensdregt et al. (2016) has attempted to formalise this co-development hypothesis in a computational model in which Bayesian agents learn both a language and

a way of inferring other agents’ perspectives — replicating several of the co-development findings summarised above. How these co-developmental dynamics play out over the course of (cultural and biological) evolution is an interesting question for future research that could be addressed with such a computational model, using the iterated learning framework (Kirby et al., 2015).

## 1.8 Biological, cultural, and co-evolution of pragmatic competence

To conclude, pragmatics is a part of human language use that requires an evolutionary account of its own; separate from an account of how the linguistic code evolved. Pragmatic competence involves the ability to recognise and entertain informative and communicative intentions, which in turn requires an ability to represent mental states — often referred to as theory of mind (ToM). Although there are some ToM abilities that humans share with nonhuman primates — and that were thus already present before the linguistic code evolved — these abilities are limited in crucial ways when compared to the ToM abilities of adult humans. Specifically, nonhuman primates seem incapable of entertaining fully ‘representational’/‘propositional’ representations of mental states, and are presumably also limited in their ability of entertaining higher-order metarepresentations.

One possibility is that these more sophisticated ToM abilities evolved in humans for the purpose of either cooperation or cultural transmission (or both), as a result of biological adaptation. Such biological evolution may have led to an increased sensitivity to acts of ostension and/or an increased motivation to engage in shared intentionality. However, another intriguing possibility is that (part of) these more sophisticated, explicit ToM abilities evolved through cultural evolution — where cognitive skills are transmitted from generation to generation through social learning. This second possibility may have been unlocked by an initial biological adaptation that allowed for more reliable cultural transmission. Cultural evolution of ToM would have allowed for an accumulation of cultural practices for discussing and reasoning about the minds of others; which may have been key to the evolution of the sophisticated explicit ToM skills we find in humans today.

Such cultural accumulation of mental state reasoning may not have been possible without language however, which leads to the hypothesis that language (in the sense of the linguistic code) and pragmatic competence have co-evolved. This possibility deserves exploration in future research.

## Further Reading

Ordered according to the structure of the chapter:

- **Books:**

- Sperber, D. and Wilson, D. (1995). *Relevance: Communication and Cognition*. Basil Blackwell, Oxford.
- Scott-Phillips, T. C. (2015b). *Speaking Our Minds*. Palgrave Macmillan.
- Wellman, H. M. (2014). *Making minds: How theory of mind develops*. Oxford University Press.
- Apperly, I. (2010). *Mindreaders: the cognitive basis of "theory of mind"*. Psychology Press.

- Liebal, K., Waller, B. M., Burrows, A. M., and Slocombe, K. E. (2014). *Primate Communication: A Multimodal Approach*. Cambridge University Press.
- Tomasello, M. (2008). *Origins of Human Communication*. Cambridge, MA: MIT.

#### • Journal Articles

- Scott-Phillips, T. C. (2015). Nonhuman Primate Communication, Pragmatics, and the Origins of Language. *Current Anthropology*, 56(1):56-80.
- Moore, R. (2015). Meaning and ostension in great ape gestural communication. *Animal Cognition*.
- Moore, R. (2016a). Gricean Communication and Cognitive Development. *The Philosophical Quarterly*.
- Tomasello, M., Melis, A. P., Tennie, C., Wyman, E., and Herrmann, E. (2012). Two Key Steps in the Evolution of Human Cooperation: The Interdependence Hypothesis. *Current Anthropology*, 53(6):673-692.
- Heyes, C. M. and Frith, C. D. (2014). The cultural evolution of mind reading. *Science*, 344(6190):1243091.

## References

- Ale, S. B., Brown, J. S., and Sullivan, A. T. (2013). Evolution of Cooperation: Combining Kin Selection and Reciprocal Altruism into Matrix Games with Social Dilemmas. *PLOS ONE*, 8(5):1-9.
- Apperly, I. A. and Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, 116(4):953-70.
- Bar-On, D. (2013). Origins of Meaning: Must We ‘Go Gricean’? *Mind & Language*, 28(3):342-375.
- Bar-On, D. (2016). Sociality, Expression, and This Thing called Language. *Inquiry*, 59(1):56-79.
- Baron-Cohen, S., Jolliffe, T., Mortimore, C., and Robertson, M. (1997). Another Advanced Test of Theory of Mind: Evidence from Very High Functioning Adults with Autism or Asperger Syndrome. *Journal of Child Psychology and Psychiatry*, 38(7):813-822.
- Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the autistic child have a theory of mind? *Cognition*, 21:37-46.
- Barrett, H. C., Broesch, T., Scott, R. M., He, Z., Baillargeon, R., Wu, D., Bolz, M., Henrich, J., Setoh, P., Wang, J., and Laurence, S. (2013). Early false-belief understanding in traditional non-Western societies. *Proceedings of the Royal Society B*, 280(1755):20122654.
- Breheny, R. (2006). Communication and Folk Psychology. *Mind & Language*, 21(1):74-107.
- Brooks, R. and Meltzoff, A. N. (2015). Connecting the dots from infancy to childhood: A longitudinal study connecting gaze following, language, and explicit theory of mind. *Journal of Experimental Child Psychology*, 130:67-78.
- Brüne, M. and Brüne-Cohrs, U. (2006). Theory of mind evolution, ontogeny, brain mechanisms and psychopathology. *Neuroscience & Biobehavioral Reviews*, 30(4):437-455.
- Burkart, J. M., Hrdy, S. B., and Van Schaik, C. P. (2009). Cooperative breeding and human cognitive evolution. *Evolutionary Anthropology*, 18(5):175-186.
- Byrne, D. (1996). Machiavellian Intelligence II. *Evolutionary Anthropology*, 5(5):172-180.

- Call, J. and Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, 12(5):187–192.
- Carston, R. (2002). *Thoughts and Utterances*. Blackwell Publishing Ltd.
- Crockford, C., Wittig, R. M., Mundry, R., and Zuberbühler, K. (2012). Wild Chimpanzees Inform Ignorant Group Members of Danger. *Current Biology*, 22:142–146.
- Csibra, G. (2010). Recognizing Communicative Intentions in Infancy. *Mind & Language*, 25(2):141–168.
- Csibra, G. and Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In Munakata, Y. and Johnson, M. H., editors, *Processes of Change in Brain and Cognitive Development. Attention and Performance*, XXI, pages 249–274. Oxford University Press.
- Csibra, G. and Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society B*, 366(1567):1149–1157.
- Dennett, D. C. (1983). Intentional systems in cognitive ethology: The Panglossian paradigm defended. *The Behavioral and Brain Sciences*, 6:343–390.
- Frank, M. C. and Goodman, N. D. (2012). Predicting Pragmatic Reasoning in Language Games. *Science*, 336:998.
- Franke, M. and Jäger, G. (2014). Pragmatic Back-and-Forth Reasoning. In Reda, S. P., editor, *Pragmatics, Semantics and the Case of Scalar Implicatures*, pages 170–200. Palgrave MacMillan, New York.
- Gärdenfors, P. (2003). *How Homo Became Sapiens: On the Evolution of Thinking*. Oxford University Press.
- Gómez, J.-C. (1994). Mutual awareness in primate communication: A Gricean approach. In Parker, S., Boccia, M., and Mitchell, R., editors, *Self-recognition and awareness in apes, monkeys and children*, pages 61–80. Cambridge, UK: Cambridge University Press.
- Gómez, J. C. (2007). Pointing behaviors in apes and human infants: A balanced interpretation. *Child Development*, 78(3):729–734.
- Gould, S. J. and Vrba, E. S. (1982). Exaptation - A Missing Term in the Science of Form. *Paleobiology*, 8(1):4–15.
- Grice, H. P. (1957). Meaning. *The Philosophical Review*, 66(3):377–388.
- Grice, H. P. (1975). Logic and Conversation. In Grice, H. P., editor, *Studies in the Way of Words*, pages 305–315. Harvard University Press.
- Helming, K. A., Strickland, B., and Jacob, P. (2014). Making sense of early false-belief understanding. *Trends in Cognitive Sciences*, 18(4):167–170.
- Henrich, J. and McElreath, R. (2003). The Evolution of Cultural Evolution. *Evolutionary Anthropology*, 12(3):123–135.
- Herrmann, E., Call, J., Hernández-lloreda, M. V., Hare, B., and Tomasello, M. (2007). Humans Have Evolved Specialized Skills of Social Cognition: The Cultural Intelligence Hypothesis. *Science*, 1360(317):1360–1366.
- Heyes, C. (2012a). New thinking: the evolution of human cognition. *Philosophical Transactions of the Royal Society B*, 367:2091–2096.
- Heyes, C. (2012b). What’s social about social learning? *Journal of Comparative Psychology*, 126(2):193–202.
- Heyes, C. M. and Frith, C. D. (2014). The cultural evolution of mind reading. *Science*, 344(6190):1243091–1–6.
- Hughes, C., Jaffee, S. R., Happé, F., Taylor, A., Caspi, A., and Moffitt, T. E. (2005). Origins of Individual Differences in Theory of Mind: From Nature to Nurture? *Child Development*, 76(2):356–370.

- Kaminski, J., Call, J., and Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition*, 109:224–234.
- Kampis, D., Somogyi, E., Itakura, S., and Király, I. (2013). Do infants bind mental states to agents? *Cognition*, 129:232–240.
- Kirby, S., Tamariz, M., Cornish, H., and Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141:87–102.
- Kovács, Á. M., Téglás, E., and Endress, A. D. (2010). The social sense: susceptibility to others’ beliefs in human infants and adults. *Science*, 330(6012):1830–4.
- Krachun, C., Carpenter, M., Call, J., and Tomasello, M. (2009). A competitive nonverbal false belief task for children and apes. *Developmental Science*, 12(4):521–35.
- Leavens, D. a., Russell, J. L., and Hopkins, W. D. (2010). Multimodal communication by captive chimpanzees (Pan troglodytes). *Animal Cognition*, 13(1):33–40.
- Leslie, A. M. (2005). Developmental parallels in understanding minds and bodies. *Trends in Cognitive Sciences*, 9(10):459–62.
- Leslie, A. M., Friedman, O., and German, T. P. (2004). Core mechanisms in "theory of mind". *Trends in Cognitive Sciences*, 8(12):528–33.
- Liebal, K., Call, J., Tomasello, M., Pika, S., Call, J., and Tomasello, M. (2004). To move or not to move: How apes adjust to the attentional state of others. *Interaction Studies*, 5(2):199–219.
- Liebal, K., Waller, B. M., Burrows, A. M., and Slocombe, K. E. (2014). *Primate Communication: A Multimodal Approach*. Cambridge University Press.
- Liu, D., Sabbagh, M. a., Gehring, W. J., and Wellman, H. M. (2004). Decoupling beliefs from reality in the brain: an ERP study of theory of mind. *NeuroReport*, 15(6):991–995.
- Lohmann, H. and Tomasello, M. (2003). The Role of Language in the Development of False Belief Understanding: A Training Study. *Child Development*, 74(4):1130–1144.
- Lyn, H., Russell, J. L., and Hopkins, W. D. (2010). The impact of environment on the comprehension of declarative communication in apes. *Psychological Science*, 21(3):360–365.
- Malle, B. F. (2002). The relation between language and theory of mind in development and evolution. In Givón, T. and Malle, B. F., editors, *The evolution of language out of pre-language*, pages 265–284. Amsterdam: John Benjamins.
- Martin, A. and Santos, L. R. (2016). What Cognitive Representations Support Primate Theory of Mind? *Trends in Cognitive Sciences*, 20(5):375–382.
- Maynard Smith, J. and Harper, D. G. C. (1995). Animal signals: Models and terminology. *Journal of Theoretical Biology*, 177(3):305–311.
- Meins, E., Fernyhough, C., Wainwright, R., Das Gupta, M., Fradley, E., and Tuckey, M. (2002). Maternal Mind-Mindedness and Attachment Security as Predictors of Theory of Mind Understanding. *Child Development*, 73(6):1715–1726.
- Moore, R. (2014). Ontogenetic Constraints on Grice’s Theory of Communication. In Matthews, D., editor, *Pragmatic Development in First Language Acquisition*, pages 87–104. London: John Benjamins Publishing.
- Moore, R. (2015). Meaning and ostension in great ape gestural communication. *Animal Cognition*.
- Moore, R. (2016a). Gricean Communication and Cognitive Development. *The Philosophical Quarterly*.
- Moore, R. (2016b). Gricean Communication, Joint Action, and the Evolution of Cooperation. *Topoi*.
- Nunn, C. L. (2011). *The comparative approach in evolutionary anthropology and biology*. University of Chicago Press.

- O'Grady, C., Kliesch, C., Smith, K., and Scott-Phillips, T. C. (2015). The ease and extent of recursive mindreading, across implicit and explicit tasks. *Evolution and Human Behavior*, pages 1–10.
- Onishi, K. H. and Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308(5719):255–8.
- Parish-Morris, J., Hennon, E. A., Hirsh-Pasek, K., Golinkoff, R. M., and Tager-Flusberg, H. (2007). Children with autism illuminate the role of social intention in word learning. *Child Development*, 78(4):1265–87.
- Penn, D. C. and Povinelli, D. J. (2007). On the lack of evidence that non-human animals possess anything remotely resembling a 'theory of mind'. *Philosophical transactions of the Royal Society B*, 362(1480):731–744.
- Povinelli, D. J., Theall, L. A., Reaux, J. E., and Dunphy-Lelii, S. (2003). Chimpanzees spontaneously alter the location of their gestures to match the attentional orientation of others. *Animal Behaviour*, 66:71–79.
- Pyers, J. E. and Senghas, A. (2009). Language Promotes False-Belief Understanding: Evidence From Learners of a New Sign Language. *Psychological Science*, 20(7):805–812.
- Rakoczy, H. (2012). Do infants have a theory of mind? *The British Journal of Developmental Psychology*, 30:59–74.
- Rubio-Fernández, P. and Geurts, B. (2012). How to Pass the False-Belief Task Before Your Fourth Birthday. *Psychological Science*, 24(1):27–33.
- Russell, J. L., Lyn, H., Schaeffer, J. A., and Hopkins, W. D. (2011). The role of socio-communicative rearing environments in the development of social and physical cognition in apes. *Developmental Science*, 14(6):1459–1470.
- Scarantino, A. (2013). Rethinking functional reference. *Philosophy of Science*, 80(5):1006–1018.
- Schel, A. M., Machanda, Z., Townsend, S. W., Zuberbühler, K., and Slocombe, K. E. (2013a). Chimpanzee food calls are directed at specific individuals. *Animal Behaviour*, pages 1–11.
- Schel, A. M., Townsend, S. W., Machanda, Z., Zuberbühler, K., and Slocombe, K. E. (2013b). Chimpanzee Alarm Call Production Meets Key Criteria for Intentionality. *PLOS ONE*, 8(10):1–11.
- Schick, B., de Villiers, P., de Villiers, J., and Hoffmeister, R. (2007). Language and Theory of Mind: A Study of Deaf Children. *Child Development*, 78(2):376–96.
- Scott-Phillips, T. C. (2014). *Speaking Our Minds*. Palgrave Macmillan.
- Scott-Phillips, T. C. (2015a). Meaning in animal and human communication. *Animal Cognition*, pages 801–805.
- Scott-Phillips, T. C. (2015b). Nonhuman Primate Communication, Pragmatics, and the Origins of Language. *Current Anthropology*, 56(1):56–80.
- Seyfarth, R. M., Cheney, D. L., and Marler, P. (1980). Monkey Responses to Three Different Alarm Calls: Evidence of Predator Classification and Semantic Communication. *Science*, 210(4471):801–803.
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*.
- Skyrms, B. (2004). *The Stag Hunt and the Evolution of Social Structure*. Cambridge University Press.
- Slaughter, V. and Perez-Zapata, D. (2014). Cultural Variations in the Development of Mind Reading. *Child Development Perspectives*, 8(4):237–241.
- Slaughter, V., Peterson, C. C., and Mackintosh, E. (2007). Mind What Mother Says: Narrative Input and Theory of Mind in Typical Children and Those on the Autism Spectrum. *Child Development*, 78(3):839–858.
- Slocombe, K. E. and Zuberbühler, K. (2007). Chimpanzees modify recruitment screams as a function of audience composition. *PNAS*, 104(43):17228–17233.

- Smith, K. and Kirby, S. (2008). Cultural evolution: implications for understanding the human language faculty and its evolution. *Philosophical Transactions of the Royal Society B*, 363(1509):3591–603.
- Southgate, V., Senju, A., and Csibra, G. (2007). Action Anticipation Through of False Belief by Attribution. *Psychological Science*, 18(7):587–592.
- Sperber, D. (2000). Metarepresentations in an Evolutionary Perspective. In Sperber, D., editor, *Metarepresentations: A Multidisciplinary Perspective*. Oxford University Press.
- Sperber, D. and Wilson, D. (1995). *Relevance: Communication and Cognition, Second Edition*. Basil Blackwell, Oxford.
- Sperber, D. A. N. and Wilson, D. (2002). Pragmatics, Modularity and Mind-reading. *Mind & Language*, 17:3–23.
- Sterelny, K. (2012). Language, gesture, skill: the co-evolutionary foundations of language. *Philosophical Transactions of the Royal Society B*, 367:2141–2151.
- Stout, D. (2011). Stone toolmaking and the evolution of human culture and cognition. *Philosophical Transactions of the Royal Society B*, 366(1567):1050–1059.
- Summers, K. and Clough, M. E. (2001). The evolution of coloration and toxicity in the poison frog family (Dendrobatidae). *PNAS*, 98(11):6227–32.
- Surian, L., Caldi, S., and Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science*, 18(7):580–6.
- Taoumpeau, M. and Ruffman, T. (2006). Mother and Infant Talk about Mental States Relates to Desire Language and Emotion Understanding. *Child Development*, 77(2):465–481.
- Taoumpeau, M. and Ruffman, T. (2008). Stepping Stones to Others’ Minds: Maternal Talk Relates to Child Mental Language and Emotion Understanding. *Child Development*, 79(2):284–302.
- Thompson, B., Kirby, S., and Smith, K. (2016). Culture shapes the evolution of cognition. *PNAS*, 113(16):201523631.
- Tomasello, M. (2008). *Origins of human communication*. MIT Press.
- Tomasello, M., Melis, A. P., Tennie, C., Wyman, E., and Herrmann, E. (2012). Two Key Steps in the Evolution of Human Cooperation: The Interdependence Hypothesis. *Current Anthropology*, 53(6):673–692.
- Wellman, H. M., Cross, D., and Watson, J. (2001). Meta-Analysis of Theory-of-Mind Development: The Truth about False Belief. 72(3):655–684.
- Wharton, T. (2003). Natural Pragmatics and Natural Codes. *Mind & Language*, 18(5):447–477.
- Whiten, A. and Erdal, D. (2012). The human socio-cognitive niche and its evolutionary origins. *Philosophical Transactions of the Royal Society B*, 367:2119–2129.
- Woensdregt, M. S., Kirby, S., Cummins, C., and Smith, K. (2016). Modelling the co-development of word learning and perspective-taking. In *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*.
- Wrangham, R. and Carmody, R. (2010). Human adaptation to the control of fire. *Evolutionary Anthropology*, 19(5):187–199.
- Xia, H., Wu, N., and Su, Y. (2012). Investigating the Genetic Basis of Theory of Mind (ToM): The Role of Catechol-O-Methyltransferase (COMT) Gene Polymorphisms. *PLOS ONE*, 7(11).
- Zuberbühler, K. (2013). Acquired mirroring and intentional communication in primates. *Language and Cognition*, 5(2-3):133–143.