# Oxford Research Encyclopedia of Linguistics

## Pragmatics and Language Evolution

Marieke Woensdregt and Kenny Smith

## Summary and Keywords

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

From the point of view of both speaker and listener, doing pragmatics requires reasoning about the minds of others. For instance, a speaker has to think about what knowledge they share with the listener to choose what information to explicitly encode in their utterance and what to leave implicit. A listener 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 developed in humans compared to other animals.

What we know about how pragmatics (and the underlying ability to make inferences about the minds of others) has evolved. Biological evolution and cultural evolution are the two main processes that can lead to the development of a complex behavior over generations, and we can explore to what extent they account for what we know about pragmatics.

In biological evolution, changes happen as a result of natural selection on genetically transmitted traits. In cultural evolution on the other hand, selection happens on skills that are transmitted 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 (for example, as a result of 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 that are so central to language use may themselves be partially shaped by social learning.

In light of empirical findings from comparative, developmental, and experimental research, we can consider the potential contributions of both biological and cultural evolutionary mechanisms to the evolution of pragmatics. The dynamics of types of evolutionary processes can also be explored using experiments and computational models.

# 1. Pragmatics and Mindreading

Being a competent language user does not just involve having access to a vocabulary and a grammar that are shared with others. It also involves knowing how to deploy those linguistic tools 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, that is, 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, Russell, and Hopkins (2010) found that if a human experimenter was facing a chimpanzee, the latter would use gestures to request a specific food item; if, however, the experimenter was facing away, they used vocalizations to attract the experimenter's attention first. Chimpanzees in the wild have also been observed to adapt their signalling behavior 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 that 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 & 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 at least with 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[1]. Several researchers have argued that even the simplest exchanges in human language require several levels of embedded reasoning about mental states (Scott-Phillips, 2015A; Sperber & Wilson, 1995), 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, 2016A, 2016C), 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, 2016C). This ability is often referred to as *theory of mind*, *mindreading*, *metapsychology*, or *mentalizing* (see e.g., Baron-Cohen, Leslie, & Frith, 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?

To answer these questions, we will start by providing an analysis of what human pragmatic competence consists of (section 2, "What Is Pragmatic Competence?"), followed by a breakdown of the psychological mechanisms involved (section 3, "Psychological Mechanisms Underlying Pragmatic Competence"). We will then go on to explore to what extent these psychological mechanisms are shared between humans and other primates (section 4, "Psychological Mechanisms Underlying Pragmatic Competence") 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 5, "The Biological Evolution of Human Pragmatic Skills"), followed by explanations drawing on cultural evolution (section 6, "The Cultural Evolution of Human Pragmatic Skills"). 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 7, "Have Language and Theory of Mind Co-Evolved?").

# 2. What Is Pragmatic Competence?

Pragmatic competence is what allows an individual to look beyond the literal meaning of an utterance 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 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, pp. 19–30). 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 analyze 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 signaler 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 signaler 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 analyzed 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 (1995) 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 behavior) or "Look, an eagle!" (an intention to attract attention, share an experience). To emphasise this point, Moore (2016C) reformulates the two intentions of the speaker as follows (conceded by Sperber & Wilson, 1995):

1. An intention to produce a particular response in the hearer/audience.
2. An intention that the hearer/audience recognizes intention 1. (Moore, 2016C)

Sperber and Wilson (1995) use the term *ostensive behavior* or simply *ostension* to describe communicative behavior that involves both these intentions; "behavior 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 that 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 behavior 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 signaler and receiver share sufficient background. Second, the content of an informative intention can be recovered by a hearer even without recognizing the encompassing communicative intention. This is especially the case in non-linguistic and non-conventionalized ostensive behavior, 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 realizing that the phone was moved there with an intention to signal something. However, the ability to recognize 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, recognizing that there is a communicative intention will motivate them to spend cognitive resources on inferring it (Csibra, 2010). This is what Sperber and Wilson (1995) 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 & Wilson, 2002). This makes it highly likely that human communication involves some level of mental state attribution and thus theory of mind (ToM).[2] To answer the question of how this pragmatic competence *evolved*, however, we need a theoretical framework for analyzing exactly what psychological processes are involved and what the precursors of these might be.

# 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 behavior in a way that increases the sender's fitness (Maynard Smith & Harper, 1995). An example of this kind of signal is *aposematism* (warning coloration), which we find, for instance, in poisonous frogs that have evolved a salient skin color that 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 signaler (Summers & 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 to accommodate the fact that, although signalers 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 signaling 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 behavior (Seyfarth, Cheney, & Marler, 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); that is, 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 (an example of *third-order* intentionality for instance is "Ella wants Steve to believe that she did not know about the surprise party"), and every order from second-order intentionality upwards involves the ability to entertain *metarepresentations*— to have representations *about* representations. This is something humans are remarkably good at—O'Grady, Kliesch, Smith, and Scott-Phillips (2015) showed that adults can keep track of mental state representations up to seven levels deep. How many levels of metarepresentation are minimally required to do ostensive-inferential communication is a question currently under debate, which is discussed in section 3.1, "Minimal Requirement for Ostensive-Inferential Communication."

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 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 is the ability to have beliefs about beliefs, etc. (e.g., Baron-Cohen, Jolliffe, Mortimore, & Robertson, 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 that 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, Cross, & Watson, 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, for example, 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, Strickland, & Jacob, 2014). Human children only start succeeding at these explicit tasks around the age of four (Wellman et al., 2001).[3] In contrast, *implicit* false belief understanding is measured using gaze direction or looking times, in tasks that 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 the character's false belief, or the amount of time the child spends looking at the character when they search for their toy in the location that was *un*expected 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 Barrett et al., 2013; Southgate, Senju, & Csibra, 2007, for the anticipatory looking

paradigm; and see Kovács, Téglás, & Endress, 2010; Onishi & Baillargeon, 2005; Surian, Caldi, & Sperber, 2007, 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. First, 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, Friedman, and German (2004) and Leslie (2005) argue that children have an innate mechanism for representing the mental states of others, but that they have learned as a default option that others' beliefs about the world are the same as their own (also known as a *reality bias; see also Birch & Bloom, 2004*). 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.[4] This first account is thus compatible with the view that *explicit* false belief tasks do not accurately reflect the mindreading abilities of young children.

Second, 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 she believes that.". A representation of a subdoxastic state such as "The character 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 character 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 (a) they cannot be integrated with informational states from other areas of cognition, and (b) 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.

Third, 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 Butterill 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" vs. "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 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 as necessary: (a) the ability to entertain metarepresentations, and (b) the ability to represent beliefs (as opposed to subdoxastic states) (see Scott-Phillips, 2014; Sperber, 2000; Sperber & Wilson, 2002; and Tomasello, 2008, for the metarepresentations claim; and see Breheny, 2006, for the beliefs claim). These arguments have subsequently been

used to claim that this type of communication is unique to humans (Scott-Phillips, 2014; Sperber, 2000; Tomasello, 2008). 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.

## 3.1. Minimal Requirements 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 (1995) 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 *recognize* 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/recognizes* the presence of tubers requires at most an ability to represent a registration or awareness 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 & Butterfill, 2009; Martin & 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 *recognize* (i.e., factual) the presence of tubers, this will still produce the same behavioral response. This would also explain why infants seem to be able to recognize communicative intentions very early on (Csibra, 2010), without having to posit that they can already represent abstract mental states like beliefs.[5]

Aside from an ability to represent beliefs, Sperber (2000) posits that ostensive-inferential communication also requires the ability to entertain *fourth-order* metarepresentations as the one depicted below (where *S* stands for sender and *R* stands for receiver):

| | |
|---|---|
| fourth order: | *S* intends |
| third order: | That *R* believe |
| second order: | That *S* intends |
| first order: | That *R* believe |
| representation: | That there are tubers for which they 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 behavioral 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 recognizing 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). Given this separation, we can break down the schema above into two separate second-order metarepresentations. For the act of address, the sender would (maximally) need to entertain a representation like the following:

| | |
|---|---|
| second order: | *S* intends that |
| first order: | *R* see that |
| representation: | *S* is addressing to *R* an action *x*. |

And for the act of sign production, a representation like the one below would suffice:

| | |
|---|---|
| second order: | *S* intends that |
| first order: | *R* recognize that |
| representation: | There are tubers for which they could dig. |

This is already less cognitively demanding than the fourth-order metarepresentation analysis of Sperber (2000), but Moore (2016A) shows how in certain cases even lower-level metarepresentations would suffice. If for example the sign that the sender produces is a point to the ground, the second-order metarepresentations above could be reduced to first-order metarepresentations as follows:

Act of address:

| | |
|---|---|
| first order: | *S* intends that |
| representation: | *R* attend and respond to her gesture. |

Sign production:

| | |
|---|---|
| first order: | *S* intends that |
| representation: | *R* looks at the ground by *S*'s feet. |

Furthermore, from the perspective of the sender, it is not necessary to *explicitly* represent the first 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.

To summarize, sending a signal ostensively and intentionally thus *minimally* requires only first-order metareprsentations in the case of *declarative* communication (i.e., information-sharing) and no

metarepresentations at all in the case of *imperative* communication (i.e., requests or demands, such as the pointing example above). From the receiver's perspective, there is always one extra level of metarepresentation required compared to what the sender needs to represent; second-order metarepresentations in the case of declarative communication, and first-order metarepresentations in the case of imperative communication.

Subdoxastic states and first-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. 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.

# 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 theorizing 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.

## 4.1. Mental State Representations

Most studies of great ape ToM have been conducted with captive chimpanzees. For instance, Kaminski, Call, and Tomasello (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. (2008) 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 can predict their behavior accordingly.[6] 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 a shift event, where, after the initial baiting of the cups, the reward was either lifted 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 as very desirable by both participants and one regarded as 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, Carpenter, Call, and Tomasello (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 behavior 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 behavior-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.

More recently, Krupenye et al. (2016) looked specifically at great apes' *implicit* signs of false-belief understanding, using the eye-tracking method. In this study, the apes (chimpanzees, bonobos and orang-utans) watched videos of a human actor interacting with another actor in a King Kong costume. In one set of videos, the actor was looking for King Kong who was hiding in one of two haystacks (experiment 1); in another set of videos, the actor was looking for a stone that King Kong had hidden in one of two boxes (experiment 2). In both

experiments, King Kong rehid (himself or the stone) while the actor was in another room, in order to induce a false belief in the actor. The actor then returned and ambiguously approached both locations. During this ambiguous approach, the ape participants' first anticipatory look towards the two possible hiding locations was measured. Results showed that the apes made significantly more first looks towards the location where the actor falsely believed his target to be than to the 'true belief' location. This result, in accordance with the looking time results of Krachun et al. (2009), suggests that apes' abilities to understand beliefs may be similar to those of human infants.

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 that can be used to reason with and act upon.[7] 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 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. In addition, evidence has also been found that chimpanzees are able to entertain at least first-order metarepresentations (Call and Carpenter, 2000; Call 2010; Beran, Smith and Perdue 2013). These beginnings of belief understanding and metarepresentation may be just enough to fulfill the minimal requirements for ostensive-inferential communication as defined by Moore (2016A), described in the previous section.

## 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 of whether great apes produce their signals (be it gestures or vocalizations) intentionally, that is, exhibiting an informative intention (with first- or second-order intentionality, according to the analysis of sign production in section 3.1). This is different from the question of whether great ape communication is ostensive, because ostension also requires a communicative intention, or in other words, overt intentionality. Liebal, Waller, Burrows, and Slocombe (2014, pp. 169–193) give an extensive overview of the different indicators of intentionality that have been adopted in studies of primate communication, and categorize some of these as strong and some as weak. The four weak criteria are (a) social use; (b) visual-orienting behavior or gaze alternation; (c) response-waiting; and (d) flexibility. The three strong criteria are (e) the production of a signal selectively for certain individuals in an audience (a subclass of social use); (f) the production of a signal only when the intended receiver is already attending to the signaler, or actively manipulating the attention of the receiver; and (g) 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 [f]) can also be viewed as an indicator of ostension, because it serves to draw attention to the fact that there is an informative intention; that is, it serves to signal the signalhood (Scott-Phillips, 2015B). The same has been argued for eye contact (part of criterion [b]) (Gómez, 1994, 2007).

The most compelling evidence that great apes can have informative intentions when communicating comes from studies of chimpanzees' vocalizations. Elaborating on an experimental design by Crockford, Wittig, Mundry, and

Zuberbühler (2012) using a model of a viper snake (a predator much feared by wild chimpanzees), Schel, Townsend, Machanda, Zuberbühler, and Slocombe (2013ʙ) evoked alarm calls from chimps traveling 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 [e], social use; and [g], persistence), and one weaker criterion (visual-orienting behavior 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, Machanda, Townsend, Zuberbühler, and Slocombe (2013ᴀ) 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 vocalizing.

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 behavior. The chimpanzees of Schel et al. (2013ʙ) showed some of this in their persistence behavior—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 behavior.

Moore (2016c) specifically reviews the possibility and occurrence of ostension in the gestural communication of great apes, and uses strong criterion (f) (deliberately solicit[ing] the attention of others before gesturing) as the indicator, citing two findings of such behavior. First, 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. Second, 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., 2013ᴀ, 2013ʙ) and showing some signs of ostensive behavior—at least in the case of captive apes communicating with human experimenters (Liebal, Call, Tomasello, Pika, 2004; Povinelli, Theall, Reaux, & Dunphy-Lelii, 2003).

# 5. The Biological Evolution of Human Pragmatic Skills

So far, we have seen that human pragmatic competence involves sophisticated ToM skills that allow humans to represent the beliefs of others in a way that is decoupled from their own representation of the world (e.g., Liu, Sabbagh, Gehring, & Wellman, 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 & Tomasello, 2008; Kaminski et al., 2008; Krachun et al., 2009), perhaps some implicit awareness of beliefs (Krachun et al., 2009; Krupene et al., 2016), and

an ability to entertain at least first-order metarepresentations (Call & Carpenter, 2000; Call, 2010; Beran, Smith, & Perdue, 2013). Evidence has also been found that great apes put these abilities to use in their communication, both in captivity and in the wild (Liebal et al., 2004; Povinelli et al., 2003; Schel et al., 2013ᴀ, 2013ʙ). Discussion is ongoing about whether or not this qualifies as ostensive communication proper (Moore, 2016ᴄ; Scott-Phillips, 2015ʙ), 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 recognize 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, "Tʜᴇ Cᴜʟᴛᴜʀᴀʟ Eᴠᴏʟᴜᴛɪᴏɴ ᴏꜰ Hᴜᴍᴀɴ Pʀᴀɢᴍᴀᴛɪᴄ Sᴋɪʟʟꜱ," 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, Wu, & Su, 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 specialized 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, Caldi, and Sperber (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 (1–4 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 & 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 & Perez-Zapata, 2014, for a review). In addition, a twin study by Hughes, Jaffee, Happé, Taylor, Caspi, and Moffitt (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—that is, 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 (Csibra & Gergely, 2011; Scott-Phillips, 2014, 2015ʙ; Sperber, 2000; Tomasello, 2008).[8] 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 (2015ʙ), language "is made possible by mechanisms of metapsychology and is made powerful by mechanisms of association" (Scott-Phillips, 2015ʙ, p. 64) (where *mechanisms of association* refers to the ability to establish a conventional code where arbitrary vocalizations 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 & Vrba, 1982).

The question then becomes, why and how did 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., Burkart, Hrdy, & Van Schaik, 2009; Byrne, 1996; Sterelny, 2012; Tomasello, Melis, Tennie, Wyman, & Herrmann, 2012; Whiten & Erdal, 2012). The advantages that full-blown ToM brings to such lives are an increased ability to predict and manipulate each other's behavior and an increased ability for cooperation. The hypothesis that human social cognition has evolved for the purpose of cooperation has been put forward by such as Sterelny (2012), Tomasello et al. (2012), and 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, Brown, & Sullivan, 2013), there are certain conditions that have to be met for cooperation to become adaptive (i.e., to constitute a selective advantage) (Sterelny, 2012). First, cooperation should come with a relatively high benefit and low cost. Second, individuals need to interact repeatedly to build up relations of reciprocal helping, allowing individuals to build up social alliances. Third, 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.[9] 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.[10]

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 & Gergely, 2006, p. 8) that makes human infants particularly receptive to teaching. Specifically, this package includes the implicit ToM abilities that allow infants to recognize communicative intentions from very early on, through a special sensitivity to ostensive behavior (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 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; Wrangham & 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.

# 6. The Cultural Evolution of Human Pragmatic Skills

Many systems of human knowledge and behavior 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 & McElreath, 2003). Humans are, by far, the most pervasively cultural species on the planet, and language (one of our many socially learned behaviors) is one of our most striking cultural feats (Smith & Kirby, 2008; Thompson, Kirby, & Smith, 2016). Could our unusually developed capacity for reasoning about mental states in others also be a product of cultural evolution?

Heyes (2012в) 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. First, 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. Second, 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, Fernyhough, Wainwright, Das Gupta, Fradley, & Tuckey, 2002; Slaughter, Peterson, & Mackintosh, 2007; Taumoepeau & Ruffman, 2006; Taumoepeau & Ruffman, 2008). Third, the combined findings of Taumoepeau and Ruffman (2006) and Taumoepeau and Ruffman (2008) 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, Lyn, Schaeffer, and Hopkins (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, Call, Hernández-lloreda, Hare, and Tomasello (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, Russell, and Hopkins (2010) looking at the ability of great apes to understand declarative signals (pointing and vocalizations). In this study, enculturated chimpanzees and bonobos were found to significantly outperform their standard-reared counterparts in their comprehension of ostensive points and vocalizations 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 as it does in humans.

This suggest 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 refer 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, refer to that which allows humans to use their representations of the mental states of others *explicitly*, both in reasoning and behavior—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 & Butterfill, 2009; Kampis, Somogyi, Itakura, & Király, 2013; Rakoczy, 2012). Based on the evidence summarized 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, 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 (Hughes et al., 2005; Meins et al., 2002; Slaughter et al., 2007; Taumoepeau & Ruffman, 2006, 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*.

# 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 & Gergely, 2011; Moore, 2016ᴀ; Sterelny, 2012; Tomasello et al., 2012; Whiten & Erdal, 2012;) and has been fleshed out more elaborately by Malle (2002). 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, 2012ᴀ) 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. First, evidence that language-learning depends on ToM abilities is provided by Parish-Morris, Hennon, Hirsh-Pasek, Golinkoff, and Tager-Flusberg (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 requires inferring the speaker's intention, unlike language- and mental-age-matched typically developing children.

Second, 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, de Villiers, de Villiers, & Hoffmeister, 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 & Tomasello, 2003).

Third, 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 signs 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, and that these mental state terms in turn predicted the extent of their false belief understanding at 4.5-years, 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, Kirby, Cummins, and Smith (2016) has attempted to formalize 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 summarized 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, Tamariz, Cornish, & Smith, 2015).

# 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 recognize 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 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., & Wilson, D. (1995). *Relevance: Communication and cognition*. Oxford: Blackwell.
    Find this resource:

Scott-Phillips, T. C. (2015b). *Speaking our minds*. New York: Palgrave Macmillan.

Find this resource:

Wellman, H. M. (2014). *Making minds: How theory of mind develops*. Oxford: Oxford University Press.
    Find this resource:

Apperly, I. (2010). *Mindreaders: The cognitive basis of "theory of mind."* New York: Psychology Press.
    Find this resource:

Liebal, K., Waller, B. M., Burrows, A. M., & Slocombe, K. E. (2014). *Primate communication: A multimodal approach*. Cambridge, U.K.: Cambridge University Press.
    Find this resource:

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

## Journal Articles

Scott-Phillips, T. C. (2015). Nonhuman primate communication, pragmatics, and the origins of language. *Current Anthropology*, *56*(1), 56–80.
    Find this resource:

Moore, R. (2016c). Meaning and ostension in great ape gestural communication. *Animal Cognition*, *19*(1), 223–231.
    Find this resource:

Moore, R. (2016a). **Gricean communication and cognitive development**. *The Philosophical Quarterly. Advance online publication*. doi:10.1093/pq/pqw049
    Find this resource:

Tomasello, M., Melis, A. P., Tennie, C., Wyman, E., & Herrmann, E. (2012). Two key steps in the evolution of human cooperation: The interdependence hypothesis. *Current Anthropology*, *53*(6), 673–692.
    Find this resource:

Heyes, C. M., & Frith, C. D. (2014). The cultural evolution of mind reading [review]. *Science*, *344*(6190), 1243091.
    Find this resource:

## References

Ale, S. B., Brown, J. S., & 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.
    Find this resource:

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, *116*(4), 953–970.
    Find this resource:

Bar-On, D. (2013). Origins of meaning: Must we "go Gricean"? *Mind & Language*, *28*(3), 342–375.

Find this resource:

Bar-On, D. (2016). Sociality, expression, and this thing called language. *Inquiry*, *59*(1), 56–79.
    Find this resource:

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & 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.
    Find this resource:

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a theory of mind? *Cognition*, *21*, 37–46.
    Find this resource:

Barrett, H. C., Broesch, T., Scott, R. M., He, Z., Baillargeon, R., Wu, D., et al. (2013). Early false-belief understanding in traditional non-Western societies. *Proceedings of the Royal Society B*, *280*(1755), 20122654.
    Find this resource:

Beran, M. J., Smith, J. D., & Perdue, B. M. (2013). Language-trained chimpanzees (*Pan troglodytes*) name what they have seen, but look first at what they have not seen. *Psychological Science*, *24*(5), 660–666.
    Find this resource:

Birch, S. A. J., & Bloom, P. (2004). Understanding children's and adults' limitations in mental state reasoning. *Trends in Cognitive Sciences*, *8*(6), 255–260.
    Find this resource:

Breheny, R. (2006). Communication and folk psychology. *Mind & Language*, *21*(1), 74–107.
    Find this resource:

Brooks, R., & 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.
    Find this resource:

Brüne, M., & Brüne-Cohrs, U. (2006). Theory of mindevolution, ontogeny, brain mechanisms, and psychopathology. *Neuroscience & Biobehavioral Reviews*, *30*(4), 437–455.
    Find this resource:

Burkart, J. M., Hrdy, S. B., & Van Schaik, C. P. (2009). Cooperative breeding and human cognitive evolution. *Evolutionary Anthropology*, *18*(5), 175–186.
    Find this resource:

Byrne, D. (1996). Machiavellian intelligence II. *Evolutionary Anthropology*, *5*(5), 172–180.
    Find this resource:

Call, J. (2010). Do apes know that they could be wrong? *Animal Cognition*, *13*, 689–700.
    Find this resource:

Call, J., & Carpenter, M. (2000). Do apes and children know what they have seen? *Animal Cognition*, *4*, 207–220.
Find this resource:

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, *12*(5), 187–192.
Find this resource:

Carston, R. (2002). *Thoughts and utterances*. Malden, MA: Blackwell.
Find this resource:

Crockford, C., Wittig, R. M., Mundry, R., & Zuberbühler, K. (2012). Wild chimpanzees inform ignorant group members of danger. *Current Biology*, *22*, 142–146.
Find this resource:

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language*, *25*(2), 141–168.
Find this resource:

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In Y. Munakata & M. H. Johnson (Eds.), *Processes of change in brain and cognitive devolpment: Attention and performance, XXI* (pp. 249–274). New York: Oxford University Press.
Find this resource:

Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society B*, *366*(1567), 1149–1157.
Find this resource:

Dennett, D. C. (1983). Intentional systems in cognitive ethology: The Panglossian paradigm defended. *The Behavioral and Brain Sciences*, *6*, 343–390.
Find this resource:

Gärdenfors, P. (2003). *How homo became sapiens: On the evolution of thinking*. New York: Oxford University Press.
Find this resource:

Gómez, J.-C. (1994). Mutual awareness in primate communication: A Gricean approach. In S. Parker, M. Boccia, & R. Mitchell (Eds.), *Self-recognition and awareness in apes, monkeys and children* (pp. 61–80). Cambridge, U.K.: Cambridge University Press.
Find this resource:

Gómez, J. C. (2007). Pointing behaviors in apes and human infants: A balanced interpretation. *Child Development*, *78*(3), 729–734.
Find this resource:

Gould, S. J., & Vrba, E. S. (1982). Exaptation: A missing term in the science of form. *Paleobiology*, *8*(1), 4–15.
Find this resource:

Grice, H. P. (1957). Meaning. *The Philosophical Review*, *66*(3), 377–388.
Find this resource:

Grice, H. P. (1975). Logic and conversation. In H. P. Grice (Ed.), *Studies in the way of words* (pp. 305–315). Harvard University Press.

Find this resource:

Helming, K. A., Strickland, B., & Jacob, P. (2014). Making sense of early false-belief understanding. *Trends in Cognitive Sciences*, *18*(4), 167–170.

Find this resource:

Henrich, J., & McElreath, R. (2003). The evolution of cultural evolution. *Evolutionary Anthropology*, *12*(3), 123–135.

Find this resource:

Herrmann, E., Call, J., Hernández-lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science*, *317*(5843), 1360–1366.

Find this resource:

Heyes, C. (2012a). New thinking: The evolution of human cognition. *Philosophical Transactions of the Royal Society B*, *367*, 2091–2096.

Find this resource:

Heyes, C. (2012b). What's social about social learning? *Journal of Comparative Psychology*, *126*(2), 193–202.

Find this resource:

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

Find this resource:

Hughes, C., Jaffee, S. R., Happé, F., Taylor, A., Caspi, A., & Moffitt, T. E. (2005). Origins of individual differences in theory of mind: From nature to nurture? *Child Development*, *76*(2), 356–370.

Find this resource:

Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition*, *109*, 224–234.

Find this resource:

Kampis, D., Somogyi, E., Itakura, S., & Király, I. (2013). Do infants bind mental states to agents? *Cognition*, *129*, 232–240.

Find this resource:

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.

Find this resource:

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*(6012), 1830–1834.

Find this resource:

Krachun, C., Carpenter, M., Call, J., & Tomasello, M. (2009). A competitive nonverbal false belief task for children and apes. *Developmental Science*, *12*(4), 521–35.

Find this resource:

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science*, *354*(6308), 110–114.

    Find this resource:

Leavens, D. A., Russell, J. L., & Hopkins, W. D. (2010). Multimodal communication by captive chimpanzees (Pan troglodytes). *Animal Cognition*, *13*(1), 33–40.

    Find this resource:

Leslie, A. M. (2005). Developmental parallels in understanding minds and bodies. *Trends in Cognitive Sciences*, *9*(10), 459–462.

    Find this resource:

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in "theory of mind." *Trends in Cognitive Sciences*, *8*(12), 528–533.

    Find this resource:

Liebal, K., Call, J., Tomasello, M., & Pika, S. (2004). To move or not to move: How apes adjust to the attentional state of others. *Interaction Studies*, *5*(2), 199–219.

    Find this resource:

Liebal, K., Waller, B. M., Burrows, A. M., & Slocombe, K. E. (2014). *Primate communication: A multimodal approach*. New York: Cambridge University Press.

    Find this resource:

Liu, D., Sabbagh, M. A., Gehring, W. J., & Wellman, H. M. (2004). Decoupling beliefs from reality in the brain: An ERP study of theory of mind. *NeuroReport*, *15*(6), 991–995.

    Find this resource:

Lohmann, H., & Tomasello, M. (2003). The role of language in the development of false belief understanding: A training study. *Child Development*, *74*(4), 1130–1144.

    Find this resource:

Lyn, H., Russell, J. L., & Hopkins, W. D. (2010). The impact of environment on the comprehension of declarative communication in apes. *Psychological Science*, *21*(3), 360–365.

    Find this resource:

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

    Find this resource:

Martin, A., & Santos, L. R. (2016). What cognitive representations support primate theory of mind? *Trends in Cognitive Sciences*, *20*(5), 375–382.

    Find this resource:

Maynard Smith, J., & Harper, D. G. C. (1995). Animal signals: Models and terminology. *Journal of Theoretical Biology*, *177*(3), 305–311.

> Find this resource:

Meins, E., Fernyhough, C., Wainwright, R., Das Gupta, M., Fradley, E., & Tuckey, M. (2002). Maternal mind-mindedness and attachment security as predictors of theory of mind understanding. *Child Development*, *73*(6), 1715–1726.

> Find this resource:

Moore, R. (2014). Ontogenetic constraints on Grice's theory of communication. In D. Matthews (Ed.), *Pragmatic development in first language acquisition* (pp. 87–104). London: John Benjamins.

> Find this resource:

Moore, R. (2016a). **Gricean communication and cognitive development**. *The Philosophical Quarterly*. Advance online publication. doi:10.1093/pq/pqw049

> Find this resource:

Moore, R. (2016b). **Gricean communication, joint action, and the evolution of cooperation**. *Topoi*. Advance online publication. doi:10.1007/s11245-016-9372-5

> Find this resource:

Moore, R. (2016c). Meaning and ostension in great ape gestural communication. *Animal Cognition*, *19*(1), 223–231.

> Find this resource:

Nunn, C. L. (2011). *The comparative approach in evolutionary anthropology and biology*. Chicago: University of Chicago Press.

> Find this resource:

O'Grady, C., Kliesch, C., Smith, K., & Scott-Phillips, T. C. (2015). The ease and extent of recursive mindreading, across implicit and explicit tasks. *Evolution and Human Behavior*, *36*(4), 313–322.

> Find this resource:

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

> Find this resource:

Parish-Morris, J., Hennon, E. A., Hirsh-Pasek, K., Golinkoff, R. M., & Tager-Flusberg, H. (2007). Children with autism illuminate the role of social intention in word learning. *Child Development*, *78*(4), 1265–1287.

> Find this resource:

Penn, D. C., & 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.

> Find this resource:

Povinelli, D. J., Theall, L. A., Reaux, J. E., & Dunphy-Lelii, S. (2003). Chimpanzees spontaneously alter the location of their gestures to match the attentional orientation of others. *Animal Behaviour*, *66*, 71–79.

> Find this resource:

Pyers, J. E., & Senghas, A. (2009). Language promotes false-belief understanding: Evidence from learners of a new sign language. *Psychological Science*, *20*(7), 805–812.

> Find this resource:

Rakoczy, H. (2012). Do infants have a theory of mind? *The British Journal of Developmental Psychology*, *30*, 59–74.

> Find this resource:

Rubio-Fernández, P., & Geurts, B. (2012). How to pass the false-belief task before your fourth birthday. *Psychological Science*, *24*(1), 27–33.

> Find this resource:

Russell, J. L., Lyn, H., Schaeffer, J. A., & 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.

> Find this resource:

Scarantino, A. (2013). Rethinking functional reference. *Philosophy of Science*, *80*(5), 1006–1018.

> Find this resource:

Schel, A. M., Machanda, Z., Townsend, S. W., Zuberbühler, K., & Slocombe, K. E. (2013a). Chimpanzee food calls are directed at specific individuals. *Animal Behaviour*, *86*(5), 955–965.

> Find this resource:

Schel, A. M., Townsend, S. W., Machanda, Z., Zuberbühler, K., & Slocombe, K. E. (2013b). Chimpanzee alarm call production meets key criteria for intentionality. *PLoS ONE*, *8*(10), 1–11.

> Find this resource:

Schick, B., de Villiers, P., de Villiers, J., & Hoffmeister, R. (2007). Language and theory of mind: A study of deaf children. *Child Development*, *78*(2), 376–396.

> Find this resource:

Scott-Phillips, T. C. (2014). *Speaking our minds*. New York, NY: Palgrave Macmillan.

> Find this resource:

Scott-Phillips, T. C. (2015a). Meaning in animal and human communication. *Animal Cognition*, *18*(3), 801–805.

> Find this resource:

Scott-Phillips, T. C. (2015b). Nonhuman primate communication, pragmatics, and the origins of language. *Current Anthropology*, *56*(1), 56–80.

> Find this resource:

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science*, *210*(4471), 801–803.

> Find this resource:

Shannon, C. E. (1948). **A mathematical theory of communication**. *The Bell System Technical Journal*, *27*(3), 379–423.

Find this resource:

Skyrms, B. (2004). *The stag hunt and the evolution of social structure*. Cambridge, U.K.: Cambridge University Press.

Find this resource:

Slaughter, V., & Perez-Zapata, D. (2014). Cultural variations in the development of mind reading. *Child Development Perspectives*, *8*(4), 237–241.

Find this resource:

Slaughter, V., Peterson, C. C., & 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.

Find this resource:

Slocombe, K. E., & Zuberbühler, K. (2007). Chimpanzees modify recruitment screams as a function of audience composition. *Proceedings of the National Academy of Sciences*, *104(*43), 17228–17233.

Find this resource:

Smith, K., & Kirby, S. (2008). Cultural evolution, implications for understanding the human language faculty and its evolution. *Philosophical Transactions of the Royal Society B*, *363*(1509), 3591–3603.

Find this resource:

Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science*, *13*(6), 907–12.

Find this resource:

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through of false belief by attribution. *Psychological Science*, *18*(7), 587–592.

Find this resource:

Sperber, D. (2000). Metarepresentations in an evolutionary perspective. In D. Sperber (Ed.), *Metarepresentations: A multidisciplinary perspective*. New York: Oxford University Press.

Find this resource:

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2d ed.). Oxford: Basil Blackwell.

Find this resource:

Sperber, D. A. N., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind & Language*, *17*, 3–23.

Find this resource:

Sterelny, K. (2012). Language, gesture, skill: The co-evolutionary foundations of language. *Philosophical Transactions of the Royal Society B*, *367*, 2141–2151.

Find this resource:

Stout, D. (2011). Stone toolmaking and the evolution of human culture and cognition. *Philosophical Transactions of the Royal Society B*, *366*(1567), 1050–1059.

Find this resource:

Summers, K., & Clough, M. E. (2001). The evolution of coloration and toxicity in the poison frog family (Dendrobatidae). *Proceedings of the National Academy of Sciences*, *98*(11), 6227–6232.
    Find this resource:

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science*, *18*(7), 580–586.
    Find this resource:

Taumoepeau, M., & Ruffman, T. (2006). Mother and infant talk about mental states relates to desire language and emotion understanding. *Child Development*, *77*(2), 465–481.
    Find this resource:

Taumoepeau, M., & Ruffman, T. (2008). Stepping stones to others' minds: maternal talk relates to child mental language and emotion understanding. *Child Development*, *79*(2), 284–302.
    Find this resource:

Thompson, B., Kirby, S., & Smith, K. (2016). Culture shapes the evolution of cognition. *Proceedings of the National Academy of Sciences*, *113*(16), 4530–4535.
    Find this resource:

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

Tomasello, M., Melis, A. P., Tennie, C., Wyman, E., & Herrmann, E. (2012). Two key steps in the evolution of human cooperation: The interdependence hypothesis. *Current Anthropology*, *53*(6), 673–692.
    Find this resource:

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief, *Child Development*, *72*(3), 655–684.
    Find this resource:

Wharton, T. (2003). Natural pragmatics and natural codes. *Mind & Language*, *18*(5), 447–477.
    Find this resource:

Whiten, A., & Erdal, D. (2012). The human socio-cognitive niche and its evolutionary origins. *Philosophical Transactions of the Royal Society B*, *367*, 2119–2129.
    Find this resource:

Woensdregt, M. S., Kirby, S., Cummins, C., & 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*.
    Find this resource:

Wrangham, R., & Carmody, R. (2010). Human adaptation to the control of fire. *Evolutionary Anthropology*, *19*(5), 187–199.
    Find this resource:

Xia, H., Wu, N., & 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).

Find this resource:

Zuberbühler, K. (2013). Acquired mirroring and intentional communication in primates. *Language and Cognition*, *5*(2–3), 133–143.

Find this resource:

## Notes:

(1.) Although see *Consciousness and Cognition*, 2015, vol. 36 for a special issue on the extent to which certain mental states *can* be perceived directly.

(2.) Note however that (Sperber & Wilson, 2002) propose that humans have evolved a "comprehension module" that is dedicated directly to inferring informative intentions once a communicative intention is recognized, which may reduce the amount of mindreading required.

(3.) 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.

(4.) See Helming, Strickland, and Jacob (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).

(5.) Although see Southgate et al. (2010) for evidence that 17-month-old infants *are* able to use belief representations to infer referential intentions.

(6.) However see Penn and Povinelli (2007) for a discussion of alternative interpretations of such findings in terms of behavior rules.

(7.) See Martin and Santos (2016) for another classification in terms of "awareness relations" versus "representational relations."

(8.) Although see Bar-On (2016) for a different account, in which language and pragmatic ability evolved more in lockstep.

(9.) What sets this type of foraging apart from the group hunting we see in lions and orcas, for example, is that collaborative foraging refers to a situation where (a) individuals *have to* collaborate in order to benefit; (b) the yield of a collaboration has to be greater than any solo foraging alternative; and (c) 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).

(10.) Although Moore (2016b) and others argue that ostensive-inferential communication does not require cooperation.

**Marieke Woensdregt**

University of Edinburgh

**Kenny Smith**

School of Philosophy, Psychology and Language Sciences, University of Edinburgh