



The cultural evolution of mind-modelling

Richard Moore^{1,2,3} 

Received: 19 November 2019 / Accepted: 31 August 2020
© The Author(s) 2020

Abstract

I argue that uniquely human forms of ‘Theory of Mind’ (or ‘ToM’) are a product of cultural evolution. Specifically, propositional attitude psychology is a linguistically constructed folk model of the human mind, invented by our ancestors for a range of tasks and refined over successive generations of users. The construction of these folk models gave humans new tools for thinking and reasoning about mental states—and so imbued us with abilities not shared by non-linguistic species. I also argue that uniquely human forms of ToM are not required for language development, such that an account of the cultural origins of ToM does not jeopardise the explanation of language development. Finally, I sketch a historical model of the cultural evolution of mental state talk.

Keywords Cultural evolution · Theory of mind · Propositional attitudes · Language development

1 Introduction

Mindreading (involving the use of a ‘theory of mind’, or ‘ToM’) is the ability to attribute intentions, beliefs, and desires (and related mental states). This practice of ‘folk psychology’ characteristically takes the form of attributing propositional attitudes. These are combinations of representations (e.g., *it is raining*) and metarepresentational attitudes towards those representations (e.g., *Felix believes that it is raining*). Propositional attitudes can include beliefs and desires which, when combined, generate intentions, which lead to action. The ability to craft folk psychological explanations is a central component of a human ToM and is valuable for a number of reasons—including the explanation of behaviour. For example, if Felix wants to go out but believes that it

✉ Richard Moore
r.t.moore@gmail.com

¹ Department of Philosophy, University of Warwick, Coventry CV4 7AL, UK

² Fachbereich Sozial- und Bildungswissenschaften, Fachhochschule Potsdam, Potsdam, Germany

³ School of Philosophy, Australian National University, Canberra, Australia

is raining and prefers to stay dry, he may decide to take an umbrella. Someone who can attribute these mental states can explain why Felix took an umbrella when he left home.

While adult humans are better at mindreading than other species, the origins of uniquely human ToM are disputed. Nativist accounts (e.g. Sperber 2000; Scott-Phillips 2014) argue that humans are born with better ToM than other species—and that this is why only humans acquire language. In contrast, constructivists argue that ToM is something that children learn (e.g. Garfield et al. 2001; Van Cleave and Gauker 2010; Jary 2010)—such that there are fundamental differences in the mindreading abilities of children and adults. Heyes and Frith (2014) further emphasise that ToM is a product of cultural evolution—a set of tools for thinking about minds that was invented by our ancestors and developed by subsequent generations. On this development of the constructivist view, different cultures might attribute mental states in different ways. Our early human ancestors likely did not attribute propositional attitudes at all.

Developing previous constructivist accounts, here I propose an account of the cultural origins of human ToM. I argue that the cultural evolution of natural language has enabled us to model propositional attitudes for the first time, and so to employ new tools for thinking about minds. Building on more basic abilities for tracking agents' perceptual and goal-directed behavioural states—abilities that we share with other species—folk psychological models have given humans unique ToM abilities. I suggest three forms of ToM that have been enabled thus: (1) the comparison of propositional attitudes, (2) the stacking of propositional attitudes (higher order ToM), and (3) level-2 perspective taking—the ability to reason about how things look to others. I also argue that an account of language acquisition is not threatened by the developmental dependence of human ToM upon language; and I sketch an account of the possible historical emergence of ToM language.

In the next section I argue that uniquely human ToM is developmentally dependent upon language. This is to motivate the claim that uniquely human forms of ToM are products of linguistic and cultural evolution; and to motivate the development of a conceptual story about the cultural origins of human mindreading. Readers familiar with evidence that human ToM is learned and language dependent may now prefer to skip ahead to Sect. 2.3.

2 Mindreading and language development: interpreting the data

Empirical data suggest that elements of mindreading are developmentally dependent upon both communicative interaction and language. However, the interpretation of the data are complicated by apparently conflicting findings. These stem from the two different—'implicit' and 'explicit'—paradigms with which the ontogenetic development of ToM has been tested.

2.1 Verbal ('explicit') false belief tasks

Historically, the definitive test of ToM has been the 'explicit' false belief task (Wimmer and Perner 1983), which tests subjects' ability to attribute false beliefs. Understanding that beliefs can be false is a prerequisite of grasping that others occupy an epistemic perspective on the world (Dennett 1978; Bennett 1978). In the original task subjects watched a boy, Maxi, hide his chocolate in a cupboard before leaving the room. After Maxi left, subjects watched his mother move his chocolate to a different location. When Maxi returned, subjects were asked where he would look for his chocolate. Children younger than four reported that he would look where the chocolate was hidden; not where Maxi had last seen it. This was interpreted as showing that children younger than 4 years do not understand that beliefs can be false—and so lack a ToM. Since Wimmer and Perner's study, many have sought to determine why false belief understanding is late developing. A consistent finding has been a correlation between ToM and language development. The ability to pass explicit false belief tasks is predicted by semantic, syntactic, and pragmatic competence.

With respect to semantics, young children's performance on explicit false belief tasks is predicted by the frequency of their mothers' use of mental state verbs (Ruffman et al. 2002; Adrian et al. 2005)—suggesting that exposure to mental state talk is key to understanding minds. However, since mental state terms are mostly embedded in syntactically distinctive sentences, it may be syntax and not semantics that drives explicit ToM success (Pyers 2006).

A correlation between explicit ToM success and children's mastery of sentential complement syntax is now well established (de Villiers and Pyers 2002; Milligan et al. 2007; Low 2010; Grosse-Wiesmann et al. 2017a, b). Sentential complements are clauses embedded under propositional attitude verbs within a sentence—e.g., *Felix believes that Zoë is in the office*. Such forms help to represent false beliefs because their form emphasises the contrast between a proposition and an attitude towards that proposition. They can therefore be used to model thoughts in which the main clause is false (e.g., because Zoë is at home) but the whole is true (because Felix's belief was false). Since this is the linguistic structure that we use to express false beliefs, de Villiers and de Villiers (2000) argued that mastery of sentential complements is necessary for false belief understanding. This might be especially true for high order representations like *Kofi believes that Felix believes that Zoë is in the office*. Children grasp second order metarepresentations (*A believes that B believes that p*) only around 6 years (Perner and Wimmer 1985; Grueneisen et al. 2015).

Training studies support the idea that false belief understanding is facilitated by sentential complement mastery. For example, Lohmann and Tomasello (2003) trained children in different forms of discourse before testing them in an explicit false belief task. During the training an experimenter talked to 3-year-olds about 'deceptive' objects (e.g., an eraser that looked like a car) using language that was varied across conditions. Where the experimenter described the deceptive appearance of the object using complement clauses, children performed better. However, since their performance improved similarly where experimenters expressed conflicting attitudes without using sentential complements, these cannot be necessary for improved performance.

Other studies also show that sentential complements are not sufficient for an explicit understanding of false beliefs. For example, in German ‘that’-clauses are used to discuss both belief and desire, yet German children still understand desire talk earlier than belief talk (Perner et al. 2003).

One possibility is that it is not complement sentence mastery per se that drives the false belief understanding required for explicit ToM tasks, but an understanding of the ways in which individuals can have different attitudes towards the same states of affairs (including epistemic attitudes that can be correct or incorrect). Since speakers can communicate divergent (and false) perspectives without using mental state vocabulary or sentential complements, it may be exposure to dialogue that is critical for children’s ToM development (Harris et al. 2005), rather than complement mastery alone.¹ This is consistent with the possibility that specific forms of syntax facilitate false belief understanding, even if they are not necessary for it.

Evidence of the ToM abilities of users of Nicaraguan Sign Language (NSL)—invented by children at a school in Nicaragua after it opened in 1977—suggests that language contributes something that conversation alone does not. Deaf children born to hearing parents pass explicit false belief paradigms significantly later than hearing children born to hearing parents (and also Deaf children born to Deaf parents) because they are deprived of communicative input early in life (Peterson and Siegal 2000). Since children at the NSL school were mostly born to hearing parents, their ToM was typically underdeveloped upon arrival. To study the effects of language acquisition on ToM reasoning, Pyers and Senghas (2009) tested two generations of adults who had attended the school. Those who joined the school later and learned a more sophisticated version of NSL performed better in explicit false belief tasks than those who had learned a more rudimentary version of NSL (Pyers and Senghas 2009). Since both cohorts were experienced communicators, the improved ToM of the later cohort seems best explained by appeal to differences in the language they had learned.

A final source of data for the development of ToM comes from cross-cultural studies. Children from some parts of the world pass explicit false belief tests later than in others. For example, in Samoa minds are considered to be opaque, and talking about others’ mental states is taboo (Ochs 1988). Samoan children are therefore less exposed to mental state talk than elsewhere. Perhaps as a result, most do not pass explicit false belief tasks until they are 8 years old, with a third of 10–12-year-olds still performing poorly (Mayer and Träuble 2013).

These findings support the hypothesis that explicit mindreading is developmentally correlated with both the mastery of sentential complements and experience of communicating with others. Some have therefore argued that complement clause syntax and conversational experience are individually necessary and together sufficient for the development of uniquely human ToM (Garfield et al. 2001). I will (roughly) endorse

¹ de Villiers (2005) defended the syntax-only explanation of ToM development by refining her account of the sentential complements needed for ToM. To explain why German children’s use of ‘that’-clauses did not suffice for ToM, she differentiated between *realis* and *irrealis* complements, with the former (used for describing observable reality) deemed necessary for false belief understanding, and the latter (used for describing possibilities) used in desire talk. As Van Cleave and Gauker (2010) observe, this distinction maps neither to clear differences between belief and desire talk, nor to salient syntactic features that could be learned by children—making the syntax-only hypothesis less plausible than one that takes conversational experience of clashing perspectives to be necessary for mastery of sentential complements.

this view. Following others (e.g., Garfield et al. 2001; Rakoczy 2017; O'Madagain and Tomasello 2019), I argue that complement clause syntax gives humans new representational tools with which to model propositional attitudes. Communicative interaction is both the background against which linguistic tools for talking about minds are learned and the historical foundation for their invention. While conversation gives children experience of how perspectives on the world can differ, the acquisition of language gives them a way in which to represent and reason about these differences—for example, by facilitating their representation of contrasting attitudes to the same proposition.

Nonetheless, claims about the necessity of language for ToM must be qualified. Evidence shows that some stroke victims have retained their false belief reasoning abilities despite losing the ability to process the grammatical structures that enable false belief reasoning (Apperly et al. 2006). In that case, if the role of language is necessary, it must be in a developmental sense. Language is needed for acquiring representational abilities that can persist even after an agent's ability to use the relevant linguistic forms is lost. Further, I do not claim that language is the only way to acquire a human-like ToM. It may be that ToM-like representations could be acquired non-linguistically (Berio 2020). In that case even a developmental necessity claim is not a metaphysical claim about human possibility, so much as a claim about our normal developmental trajectory. Uniquely human ToM is learned and language and communication are the standard routes through which we learn it.

Before developing a positive account of ToM development, I say something about nativist alternatives. Since the development of non-verbal false belief paradigms, new nativist accounts of mindreading have complicated the interpretation of the ToM data.

2.2 Non-verbal ('implicit') false belief tasks

Following the findings of Wimmer and Perner (1983), many accepted that young children cannot understand mental states. This presented a number of problems for developmental accounts of human cognition. Not least, accounts of language development have often held that language acquisition requires developed ToM (see Breheny 2006; Moore 2017a, 2018b). This led to what Astington (2006, p. 196) described as a “paradox at the heart” of cognitive development research: language acquisition requires a developed ToM—which is seemingly language dependent. If this paradox is real, language development may be explicable only by assuming that, current data aside, human ToM is innate or early developing (Sperber 2000; Scott-Phillips 2014).

In 2005 two studies showed a way out of the paradox by suggesting that infants might possess a ToM after all. In contrast to the ‘explicit’ verbal measures used in the earlier paradigms, these studies used ‘implicit’ non-verbal looking time measures (Onishi and Baillargeon 2005; Surian et al. 2007). Whereas explicit tasks track children's ability to report on how agents with false beliefs will act, the latter use children's gaze behaviour to determine whether they anticipate that agents with false beliefs will act as if they had true beliefs. When shown scenarios like the original paradigm, children of 15-months looked longer when the Maxi character looked for his object in the correct location—suggesting surprise. This was interpreted as showing that infants can track others' false beliefs, and make predictions about their behaviour on this

basis—supporting the conclusion that older children’s failure in explicit tasks was unrelated to ToM development.

Since 2005, a large literature has tried to make sense of these apparently conflicting data. If infants can pass implicit false belief tasks by the end of their first year, then why do they fail explicit tasks before they are four? Some argue that belief understanding is innate, but that younger children are unable to recruit that understanding in explicit mindreading tasks (Carruthers 2013). This may be because 3-year-olds lack familiarity with certain types of mental state discourse and are confused by the experimenter’s questions (Helming et al. 2014, 2016; Westra 2017). This argument gains support from an explicit ToM task that simplified the pragmatics of the questions posed, and which children passed at three (Rubio-Fernández and Geurts 2012, 2016).

While some still hold that pre-verbal infants understand false belief, evidence for this claim looks increasingly underwhelming. In recent years, the failure to replicate a number of implicit tasks has mired this area of research in controversy. Rakoczy and Behne have described the current findings regarding implicit studies as “complex, confusing and puzzling” (Rakoczy and Behne 2019, p. 94), and concluded that infants’ ToM abilities are unknown. Others conclude that at least strong interpretations of infant ToM abilities (attributing to them an understanding of propositional attitudes) are unsupported by current data (Poulin-Dubois et al. 2018).

Despite this confusion, several studies now suggest that an early developing or innate capacity for tracking (but not fully representing) belief-like states is enhanced through language development. Studies of ToM development in 3–4-year-old children show that the abilities recruited in implicit ToM are not the same as those recruited in explicit ToM tasks. For example, Low (2010) showed that while performance in explicit mindreading tasks is correlated with the development of sentential complement syntax, success in implicit tasks is not. Additionally Grosse Wiesmann and colleagues (2020) have replicated their own finding that in 3- and 4-year-olds different areas of the brain are recruited in implicit and explicit ToM tasks (Grosse Wiesmann et al. 2017a, b). Explicit ToM reasoning takes place in the precuneus and temporoparietal junction, which is implicated in adult ToM. Implicit ToM reasoning, however, is supported by an independent neural network including the supramarginal gyrus, which is implicated in visual perspective-taking and action observation. Explicit ToM is also supported by white matter maturation in brain regions associated with adult ToM—and which are under-developed in toddlers (Grosse Wiesmann et al. 2017a, b, 2020).

The possibility that an early-developing or innate belief tracking ability is enriched by language gains further support from recent studies showing that chimpanzees, bonobos and orang-utans all succeed in an implicit ToM task (Krupenye et al. 2016). This finding suggests that the cognitive mechanisms needed for implicit ToM tasks may be common to all great ape species (and so present in our last common ancestor too). While the finding has yet to be replicated, it gains support from evidence that chimpanzees can track the knowledge states of their peers. They both avoid food that dominant individuals have seen (Hare et al. 2001) and call to warn naïve peers of the presence of sleeping snakes (Crockford et al. 2012, 2017; Moore 2019).

2.3 'Lean' interpretations of implicit ToM tasks

If these data support a developmental account of ToM, the precise nature of non-verbal mindreading is still not understood. Many have argued that data from implicit false belief tasks (and other non-verbal perspective-taking tasks—e.g. Hare et al. 2001; Crockford et al. 2012) can be explained without a propositional attitude psychology (e.g., Apperly and Butterfill 2009; Apperly 2010; Rakoczy 2017; Heyes 2018). Two of these approaches have sought to provide an account of the cognitive foundations of ToM by specifying the mechanisms that support implicit ToM. Heyes (2018) has argued that implicit mindreading can be explained by appeal to domain- (and species-) general cognitive resources, including memory, attention, and associative learning. In contrast, Apperly and Butterfill (Apperly and Butterfill 2009; Apperly 2010; Butterfill and Apperly 2013; Low et al. 2016) have argued for a 'two-systems' view, where implicit ToM is explained by an evolved cognitive module for 'minimal mindreading' that enables agents to track relationships between objects and other agents' lines of sight. When supplemented with additional evolved heuristics—e.g., the knowledge that agents act only in light of what they have seen—these 'registrations' permit the tracking of belief-like states (albeit ones that lack propositional contents). Tracking belief-like states enables minimal mindreaders to make fairly reliable predictions about how agents with true and false beliefs will act.

An account of human ToM should ultimately have something to say about the cognition that supports implicit ToM. Here, though, I remain open-minded. It is parsimonious to assume that infants and great apes share the cognitive mechanisms that support the tracking of belief-like states (Sober 2005; Moore 2017d). Since my own research is primarily concerned with explaining the differences between these species, and with providing an account of what enabled ToM development in the former but not the latter, this makes it legitimate to bracket questions about the common mechanisms that support implicit ToM.

While the Heyes and two-systems accounts both endorse a language-dependence view of human ToM, they say little about how creatures lacking a developed ToM might invent and acquire one. Since my goal is to fill in that gap, my interest in pre-verbal ToM is motivated by considerations of what is needed for language to develop. My concern is to sketch a theoretical account of how ToM could be invented and learned; and so to free us from the assumption that uniquely human ToM must be innate because it is needed for language development (Sperber 2000; Scott-Phillips 2014).

3 The cultural evolution of mind-modelling

If human ToM is learned, how should we think about its historical development, and the relationship between ToM and other cognitive abilities? Here I spell out three commitments of the view I defend.

First, if human ToM is developmentally dependent on the mastery of certain *natural language* forms (like sentential complements), then in key respects it is a human invention. Even if the foundations of language use are innate (e.g. Berwick and Chom-

sky 2016), natural languages like English and German were created, and the earliest languages likely contained little syntactic complexity. More complex forms of syntax emerged only as young languages were developed and refined over successive generations of use and under cultural selection for greater expressive power (see Sect. 5 for discussion). As better tools were invented, language users preferentially adopted them, while discarding older, less powerful types of linguistic construction (Christiansen and Chater 2016). This makes it highly likely that the earliest languages lacked the syntactic complexity needed for ToM talk, and that human ToM is therefore a product of cultural evolution.

Second, and building on other constructivist accounts (de Villiers and de Villiers 2000; Garfield et al. 2001; Maibom 2003; Van Cleave and Gauker 2010; O'Madagain and Tomasello 2019), we can think of the explanatory framework of propositional attitude psychology as a language-based folk model of the mind. It was likely invented by our ancestors for a range of tasks connected to human sociality, including but not limited to predicting and explaining behaviour. Conceived in this way, models are theory-like knowledge structures designed by their users for describing hypothesised states (Godfrey-Smith 2005; see also Maibom 2003). Models differ from theories by being theoretically less developed. For example, the simplest folk models might contain inconsistent propositions; or clusters of wisdom that have yet to be rationally integrated. Because models can describe the states they model more or less accurately, they can increase in complexity over time. Those that start off as a loose patchwork of informally sketched ideas may become more systematic as they are refined over generations. Models are therefore useful theoretical tools for characterising processes of knowledge formation that started off in some of the earliest human communities.

Third, building on work on the cultural invention of cognitive tools by Dennett (1995) and Heyes (Heyes and Frith 2014; Heyes 2018), I suggest that our folk psychological models have provided humans with a new format for representing intentions, beliefs and desires. This representational format does not replace the more basic interactions that ground our earliest social interactions and enable our language development. However, it has given us new abilities for talking and thinking about mental states, allowing us to theorise about minds in new ways. Human ToM has thus extended beyond that of other species. Just as the expressive power of our natural languages have increased over time, so our cognitive powers also extended, as new linguistic forms facilitated the development of expressively more powerful folk psychological models.

3.1 Folk psychology as a model

The idea that folk psychology can be conceived as a model of the mind stems from Dennett (1987, p. 43ff.; see also Gopnik et al. 1997; Maibom 2003; Godfrey-Smith 2005; Jara-Ettinger 2019). As Dennett argued, folk psychology (in his term, 'intentional systems analysis') is a powerful tool for rationalising behaviour. Idealising somewhat, if an agent's beliefs and desires are treated as those that it ought to have given its situation (e.g., those that would ensure its survival), then its actions can be predicted before they happen and underlying rationales explained.

Dennett treats propositional attitude psychology as an interpretative framework that underwent cultural selection for the prediction and explanation of behaviour. This framework gives its users a powerful tool for social interaction while allowing them to remain agnostic about which sorts of states support propositional attitude ascriptions. Terms like ‘belief’ and ‘desire’ may be applied to a cognitive system without assuming that it has conscious, first-personal states. By treating the agent *as if* it is an intentional system, its behaviour can nonetheless be explained. Scientific models help their users to understand the hypothesised states they are modelling by in some sense resembling them (Godfrey-Smith 2005). However it is a virtue of the Dennettian approach (and other similar ones, e.g. Maibom 2003) that this resemblance can take a number of forms. This is consistent with the possibility that the users of folk psychological models might conceive of minds in a number of ways. Early human may not have had deep insights into the nature of mental life prior to developing folk psychological models, or they might have conceived of these states in very different ways from us. (Even now philosophers are divided about how the states described by folk psychology should be construed. While some hold that propositional attitudes are literal descriptions of content bearing states (e.g. Fodor 2008), others hold that “talk of mental states is a useful pretence for describing people and their behaviour” (Toon 2016).²)

On an approach that treats folk psychology as the cultural development of models, we should think of its development as a gradual process in which communities of language-users modified their languages to create and refine their tools for talking about one another. Such developments would have taken place during everyday interactions like hunting trips or campfire meetings. Prediction and explanation need not have been the only motives for developing folk models. Thus early models might have been enriched through the development of language for coordinating behaviour (Van Cleave and Gauker 2010; Tomasello 2014), holding others accountable for breaking commitments (Jary 2010; Geurts 2019a), and through cultural practices of storytelling and the sharing of oral histories (Hutto 2007). We need not suppose that any single societal function was the primary driver of invention and innovation. Once linguistic tools were invented, they would be put to use for whatever range of tasks benefitted their users. Folk models innovated in one domain might be refined across a range of tasks, with these tasks also varying across communities.³

If folk psychology is the cultural construction of theory-like models of behaviour, we can also see why ToM might deliver only crude generalisations of how agents behave (Maibom 2003), and why folk models might incorporate the cultural prejudices of their users (Eickers 2019). Explanatory models might reflect culturally grounded assumptions about how individuals do and should behave. As a result, some elements of models would not be common to all communities. While relatively little is currently known about cultural differences in mindreading, evidence suggests substantial vari-

² My view is realist. Humans possess first personal states that correspond broadly to elements in our folk psychology. Nonetheless, some of the distinctions we make in language may be motivated by explanatory concerns rather than salient experiential differences. Additionally, following Hurley (1998), I do not assume that first personal states will correspond neatly to sub-personal states.

³ Within domains of use ToM may still follow predictable developmental trajectories. See Van Cleave and Gauker (2010) for detailed (albeit speculative) discussion.

ation. Lillard (1998; via Wierzbicka 1992) observes that while all known languages have words that correspond roughly to ‘want’, ‘think’, ‘know’, and ‘feel’, these words may not represent the same concepts. Moreover, while the Cartesian conception of the self as a seat of thoughts, feelings, and desires that cause behaviour is central to western thinking, not all cultures attach the same significance to attributions of mental states. Hindu Indians tend to emphasise situations rather than character traits as causes of behaviour (Miller 1984), and the Ifaluk Pacific Islanders emphasise the role of peers as causes (Lutz 1985). Philosophical conceptions of ToM should accommodate this variety. Culturally grounded folk models can reflect these differences.

This description doesn’t yet show how folk psychological modelling projects could extend human cognition. But once we think of propositional attitude psychology as a construction project whose value lies in its social utility we can envisage how the demand for better models might have led to the refinement and development of more rudimentary ones. There are numerous contexts in which more precise tools for describing behaviour (and the mental states underlying behaviour) would be valuable, leading to pressure for a greater range of linguistic tools, and permitting—for example—the expression of more fine-grained distinctions between epistemic and conative states. In turn these would support inferences that would not previously have been possible.

The idea that culturally evolved cognitive tools extend human cognition is not new. It is well established that the natural number system has changed the cognition of which we are capable. An illustration of how this happens will give us a point of comparison for thinking about the cultural evolution of mental state talk and the new forms of cognition that it enables.⁴

3.2 The cultural evolution of number cognition

We now know that the cultural evolution of counting systems has changed human number cognition. Recent empirical research suggests that humans are born with two cognitive systems relevant to enumeration (Everett 2017; also Xu 2003). One system, present in infants (Feigenson and Carey 2003), is used to make precise judgements about small numbers of objects (three in infancy, increasing to four in adulthood). A second system is used for tracking approximate quantities larger than four and is also present in 6-month old infants (Xu 2003). This Approximate Number System enables our judgements about relative quantities. Together these systems enable us to make precise judgements about small quantities of objects and approximate judgements about larger quantities. Since similar abilities seem to be present in primates (Brannon and Terrace 1998; Hauser and Carey 2003), the mechanisms that support such judgements are likely to be phylogenetically old.

Evidence that precise calculations about the relationships between large numbers became possible only with the invention of natural number systems comes from anumeric communities. The Pirahã people of the Amazon, whose language contains no

⁴ Apperly and Butterfill (2009) also note an analogy between ToM and number cognition. While they present evidence that number cognition relies on two distinct cognitive systems, my concern is to show how developed number cognition is a product of cultural evolution.

precise number terms, have been shown to struggle to make precise numerical judgements for quantities greater than three (Frank et al. 2008). While the ability to perform complex calculations is now shared by most cultures, the development of number representations emerged only relatively slowly in human history. The oldest known written number system, found in Sumeria, is only around 5300 years ago (ibid.). In all known counting systems, representational tools for higher numerosities emerged through the extension of the object individuation system, via the innovation of number words that mapped to exact values in a tally system. In many languages number words are related to the words for hands and feet (e.g., the word for ‘five’ may be derived from the word for ‘hand’), suggesting that these were the tallies onto which number words were first mapped. This is why many counting systems are base 5 (single hand), 10 (both hands), or 20 (hands and feet) (Everett 2017).

The invention of numbers for larger integers allowed for the precise enumeration of larger quantities. However they were still of only limited use for complex calculations. These became tractable only with the historical invention of ways of representing the value zero—first seen in the Sumerian culture around 5kya, but independently reinvented by both Mayan and Indian mathematicians (Kaplan 1999). The invention of a tool for representing zero dramatically facilitated the performance of long multiplication and division.

As adults who have mastered natural number systems, it is easy to forget that our ancestors could not calculate like we do. While many humans can now mentally compute that $907/3 = 302.3$ recurring, the first of our ancestors to do this lived tens of thousands of years after the great exodus of early humans from Africa 60 kya. Prior to the development of natural number systems with which to systematise and precisify our calculations, ordinary individuals could not make precise comparisons of quantity, or multiply one large number by another. In that respect, our current mathematical abilities are a product of cultural tools developed and refined by our ancestors.

4 Natural language and folk psychology

Just as the development of a natural number system enabled agents to draw fine-grained contrasts between numerosities, the development of folk psychological models enabled new forms of mental state cognition. Perhaps three abilities were enabled thus: (1) the comparison of propositional attitudes; (2) the stacking of propositional attitudes; and (3) ‘level 2’ perspective taking—that ability to track not just *what* others perceive but *how* they represent it. These abilities became possible because developments in natural languages helped our ancestors to track relationships that would otherwise have exceeded their representational capacities.⁵

⁵ Language dependent models could be natural language descriptions of elements of an interaction, or scenarios that are modelled visually and then interpreted using propositional attitude sentences. I remain open minded about the exact form that models take; both varieties may be in use.

4.1 Contrasting mental states

With respect to point (1) consider how the formulation of different propositional attitude verbs can help us to clarify our different attitudes towards the same proposition:

Richard *believes* that the train will leave on time.

Marie *doubts* that the train will leave on time.

If we knew that Richard and Marie were aiming to catch the same train, these contrasting attitudes would help us to understand why Richard rushed to leave while Marie stopped to get lunch.

Cases like this one show how language can extend thought. In a compelling recent account of how children acquire adult-like false belief understanding, O'Madagain and Tomasello (2019) argue that children learn that others' attitudes towards the truth and falsity of the same proposition can differ because with adults they can engage in joint attention towards the propositions that speakers assert. The existence of conflicting epistemic attitudes comes into view through the comparison of individuals' inconsistent responses. Suppose that Richard and Marie are together told that the train will leave at 1 pm. While Richard picks up his bag and makes to leave, Marie laughs and puts the kettle on. Their non-verbal responses towards the same proposition can be used both as a starting point for learning about disagreement (credulity versus doubt, e.g.), and for learning the language (e.g. propositional attitude verbs) that helps us to keep track of disagreements.

When children start to pass traditional false belief tasks, they may be using language to represent both that *The chocolate is **not** in the cupboard* and, simultaneously, that *Maxi believes that the chocolate **is** in the cupboard*. Here language helps us to track the conflict between how the world is and epistemic attitudes towards it, and so to predict that Maxi will look in the wrong location. Simply being able to formulate and entertain contrasting epistemic attitudes—by placing them in a common linguistic format—facilitates the construction of better explanatory and predictive models.

There is evidence that when adults are prevented from thinking linguistically in explicit false belief tasks, their performance falters. However, the evidence is imperfect. Newton and de Villiers (2007) tested adults in a simple non-verbal false belief task. They found that when participants had to repeat heard sentences while watching a video in which an agent acquired and subsequently acted upon a false belief, more than half failed (18/31 unsuccessful). In a second condition in which participants had to tap out a heard rhythm instead, almost all subjects passed (29/35 successful). This finding suggests that even in a non-verbal false belief task, interrupting participants' language abilities undermines their performance. This would be predicted if they were using language to keep track of what they were watching. Nonetheless, this interpretation of the data has been complicated by a recent study. Dungan and Saxe (2012) found that, at high tempos, adult performance in matched paradigms was also inhibited when participants had to tap out heard rhythms. While this subsequent finding is consistent with the possibility that adult ToM is both language-based and apt to be interrupted by demanding non-linguistic tasks, it also opens up the possibility that the Newton

and de Villiers (2007) findings are better explained by the greater working memory demands in the language condition. Further research is therefore needed.⁶

4.2 Stacking mental states

Linguistically constructed mental state models could also play a causal role in the ability to represent higher orders of mental states, like this third order metarepresentation:

Richard knows that Kofi doubts that Marie believes the train will leave on time.

Some argue that such complex representations are foundational to many aspects of human life—not least communication (Grice 1989; Sperber 2000; Scott-Phillips 2014)—and so potentially something that humans have evolved to represent (e.g. Sperber 2000; O’Grady et al. 2015).

Despite the relative ease with which adults understand high order metarepresentations (O’Grady et al. 2015), developmental data show that the ability to track them is both difficult for children and slow developing. Liddle and Nettle (2006) found that 10 and 11-year-old children track the contents of third-order metarepresentations slightly above chance, but that even 12-year-olds struggled to track fourth-order metarepresentations. Additionally 6-year-olds but not 5-year-olds have been shown able to reason about second-order beliefs (Perner and Wimmer 1985, Grueneisen et al. 2015). This is consistent with the possibility that higher order metarepresentations are language dependent, and develop contingently upon the ability to embed a proposition within multiple propositional attitude phrases. Mastery of higher order metarepresentations would then be acquired as children acquire fluency in the use of the longer sentences needed to model them. This could be tested with new studies of the developmental relationship between sentential complement syntax and higher order of ToM.

4.3 Level 2 perspective taking

A third sense in which language may extend human social cognition is through ‘level 2’ perspective-taking (e.g. Flavell et al. 1981).

Level 1 perspective taking involves tracking what different agents have and have not seen. Both infants and great apes do this. For example, chimpanzees vocalise the presence of snakes to peers who have not seen them (Crockford et al. 2012, 2017). Similarly, 12-month-olds point out the location of an object to an experimenter who has lost it (Liszkowski et al. 2006). If both young infants and great apes are capable of level 1 perspective-taking, it may be innate in the *hominin* lineage (Apperly and Butterfill 2009; although see Heyes 2018).

Level 2 perspective-taking is the ability to grasp how things appear to others when perceived from different perspectives. It is later developing in children. Moll and Meltzoff (2011) ran a study in which young children sat opposite an experimenter with two identical blue objects placed between them. A yellow filter was placed between

⁶ Claims about language-based thought and working memory are not inconsistent. One possibility is that language helps humans to think better because it organises working memory more efficiently..

the experimenter and one object so that that object looked green to her but blue to the subject. When the experimenter requested either “the blue one” or “the green one” (without letting her gaze fall upon the desired object), 3-year-old children correctly selected the object (*ibid.*). However, in a follow up study (Moll et al. 2013), 3-year-old children in the same setup could not answer “How do you see it from over there? ... How do I see it from over here?” questions, where this required contrasting how things looked different to themselves and the experimenter. Only 4.5-year-olds could do this. Moll and colleagues explain this finding on the basis that, while 3-year-olds can track how things look to others (‘taking perspectives’) they are unable to represent that the same thing can simultaneously look different to individuals seeing it from different perspectives (‘confronting perspectives’).

Despite several attempts to elicit level 2 perspective-taking in chimpanzees, no evidence for it has been found. In one recent study a chimpanzee competed with a conspecific over two breadsticks (Karg et al. 2016). While the subject could see that the sticks were the same size, one appeared larger to the competitor. The competitor was able to choose a breadstick first, but her choice was hidden from the subject. If subjects could track that one stick looked bigger to their competitor they could get food for themselves in every trial by choosing the stick that looked smaller to their competitor. Otherwise they would get rewarded only at chance (50%). In contrast to 6-year-old children, chimpanzees did not perform above chance.

Level 2 perspective-taking has been hypothesized to be developmentally dependent upon language (Apperly and Butterfill 2009). One possibility is that older children can do it because they use language to reconstruct the visual perspectives of others. This might take the form of a linguistically framed contrast between different perspectives on the same object—e.g., *X sees that the object is blue* and *Y sees that the object is green*. A possible explanation for why three-year-olds cannot confront perspectives is that, even if they can pick up on visual and verbal cues to make judgements about how things look to others, they cannot construct for themselves the linguistic models that facilitate understanding inconsistent appearances. Evidence for this hypothesis could be sought via studies of the developmental relationship between language and level 2 perspective-taking, and the possibility of impairing level 2 perspective-taking with verbal shadowing tasks (or similar methods).

4.4 Linguistic models of other minds

While natural language plays a fundamental role in the development of human ToM reasoning, this does not entail that mind-modelling is accompanied by an inner monologue formulating syllogisms out of propositional attitudes. In challenging or unfamiliar situations this may sometimes happen, but for users practised in the manipulation of models, elements of ToM reasoning may become automated. In these cases the deliberate reconstruction of propositional attitudes may become no more necessary for mindreading than is calculating to know the square root of 256. Through a process of downward modularisation (Apperly 2010), our perception may become theory-laden. This is presumably why a stroke patient, PH, was found to retain first and

second order ToM abilities even after substantial impairment of their comprehension of the syntactic forms associated with ToM success (Apperly et al. 2006).

For related reasons, this account should also avoid objections that it intellectualises our interactions with others. It is in some respects a version of “Theory Theory”—the view that we come to understand other minds by learning a theoretical body of knowledge. Some argue that Theory Theory misconstrues our understanding of others by presenting “our initial stance with respect to others” as “essentially estranged” (Hutto 2004, p. 549; see also McGeer 2007, p. 146 and Zawidzki 2013, 2019). On this objection, a barrier of (pseudo-)scientific theorising alienates us from our peers, and intellectualises our dealings with them. The mistake behind this complaint is the thought that, because ToM can involve elements of theory, our social cognition is fundamentally a reflective, intellectual process. I do not claim that. Because models are constructed in language, our invention and recruitment of them depends on a more basic repertoire of affective and embodied socio-cognitive skills—including but not limited to gaze tracking, an understanding of goal directed activity, and a host of other empathic relations and cooperative motivations (Tomasello 2014; Rubio-Fernandez 2020). These abilities constitute the socio-cognitive foundation of our language use. Mental state models supplement our foundational ways of interacting with others without replacing them. Similarly, just as we do not understand others solely on the basis of learned theories, we may also sometimes know other minds by simulation (Goldman 2006).

5 Inventing and acquiring a ToM

I previously mentioned a serious objection for the claim that human ToM is enabled by language: the possibility that language is itself dependent on human ToM. This concern has been left largely unaddressed by proponents of the language-first view, who have had little to say about the development of the languages whose existence they presuppose (e.g., de Villiers and de Villiers 2000; Garfield et al. 2001; Heyes and Frith 2014). If an account of the cultural origins of human ToM is to be credible, more needs to be said.

5.1 ToM and language development

Gricean communication is the name given to communication that involves agents who act with and attribute communicative intentions (Grice 1957, 1989). It is thought to require various demanding ToM abilities, including an understanding of belief and of fourth order metarepresentations (see Moore 2017a, 2018b). Since many argue that Gricean communication is also necessary for language development (e.g. Sperber 2000; Tomasello 2008; Scott-Phillips 2014), this is taken to be evidence that high order metarepresentational abilities must be innate (Sperber and Wilson 2002; Sperber 2000; Scott-Phillips 2014; see Breheny 2006 and Moore 2017a for discussion).⁷ This is a

⁷ Tomasello (1999, 2008, 2019) thinks children undergo a conceptual revolution in their ToM between 3 and 4 years, and that this is connected to the development of language and cooperation (Tomasello 2018,

reason to take nativism about ToM seriously, even if empirical data underdetermine its plausibility.

Before considering whether uniquely human ToM must precede language development I want to start by agreeing with the neo-Gricean view: Language development must be grounded in pragmatic interpretation—that is, in speakers who can act with and attribute communicative intentions.⁸ This is for a number of reasons (Moore 2018b). First, acting with and attributing communicative intent is necessary for the invention of natural languages because pragmatic interpretation is the foundation against which the meanings of semantic and syntactic elements can be introduced and calibrated. Moreover, where natural languages with only a limited vocabulary and syntax exist, and can be used only to formulate ambiguous utterances (see Sect. 5), pragmatic interpretation will be necessary for the interpretation of messages. Second, the best way to explain language development in ontogeny is by accepting that it is pre-verbal infants' pragmatic interpretation abilities that enable them to figure out the meanings of words and sentences. This makes the existence of pre-verbal Gricean communicative interaction the best prospect for explaining language development (*ibid.*). Where my account departs from other pragmatics-first accounts is by denying that the demands of Gricean communication are a reason for thinking that uniquely human ToM must be early developing or innate.

I have argued that ('minimally') Gricean communication is, contrary to the consensus view, socio-cognitively undemanding (Moore 2016, 2017a, b, c, 2018a, b). Acting with communicating intent requires only knowing how to produce signs (e.g. words or gestures) in order to express one's communicative goals, and knowing how to address these signs to the attention of interlocutors in ways that elicit an appropriate response (Moore 2017a). Attributing communicative intent requires only grasping when one is being addressed by another, and knowing how to interpret the goals with which utterances are produced. This requires neither complex metarepresentation, nor a developed propositional attitude psychology, nor even concepts of mental states like belief. It does not even require mastery of an extensive repertoire of signs (Moore 2016, 2017a, b). Consequently, the ToM abilities needed for language development are not the same ones that develop only with language and communication; and they are present in both young infants and great apes (Moore 2016, 2017c). An account of language development grounded in Gricean communication is therefore consistent with a story that takes the uniquely human ToM be developmentally dependent upon language. The paradox of language development can be avoided.

5.2 Pragmatic interpretation and the Grammaticalisation of language

Against a background of non-linguistic agents who act with and attribute communicative intentions, both the acquisition and cultural evolution of natural languages can

Footnote 7 continued

2019; O'Madagain & Tomasello, 2019). Nonetheless, he holds that infants possess a form of ToM that is not shared by great apes, which enables humans to acquire language (Tomasello, 2008). I argue against this view (Moore 2017d).

⁸ The word 'pragmatics' is used in a number of ways, with unhelpful consequences for explaining language development (Bar-On and Moore 2017; Moore 2017d). I use the term only to describe acts of interpreting communicative intent. Some (e.g. Geurts 2019a, b) would think this restriction parochial.

be explained. With respect to cultural evolution, processes of semantic and syntactic innovation would have been key.

When our ancestors started to develop the first natural languages, their proto-languages would have contained an initially small number of words and little grammatical structure. As speakers became more fluent in their sign use, they innovated new words and grammatical constructions in order to better express themselves. Innovation was possible precisely because speakers were already Gricean communicators. As Grice (1989) supposed, new words and grammatical constructions would have entered on the back of speakers' innovative uses. Once a speaker had used a new construction to communicate a message, and an interlocutor had successfully interpreted their intended message, similar messages could be communicated using the same combinations of signs (Moore 2013). As particular uses of words became more strongly associated with particular communicative functions, semantic and syntactic conventions emerged.

Recent work by Progovac (2015) indicates the likely form of proto-syntactic languages. Progovac argues that the propositional structures of contemporary natural languages were preceded by a stage of grammaticalisation that consisted of single verb-like and noun-like elements bound together non-hierarchically.⁹ In these proto-sentences, the verb like structure took only a single argument that specified neither a subject nor an object. During this stage, the most complex available utterances consisted of simple combinations like:

[1] Eat chicken.

Such phrases would be highly ambiguous. For example, [1] does not distinguish between:

- [2] The chicken is eating.
- [3] I eat the chicken.
- [4] You will eat the chicken.
- [5] You ate the chicken.

Utterances like [1] would have provided users with a range of tools for communicating and coordinating with others. Because of their ambiguity, speakers using such constructions would have depended upon their interlocutors' pragmatic interpretation skills for communication to succeed. Nonetheless, these constructions would have served as a foundation against which semantic innovation and grammaticalisation could develop.¹⁰ On Progovac's account, small clause grammars like the one exemplified in [1] served as a foundation for the emergence of verb and tensed phrases—permitting both clear distinctions between the subjects and objects of a verb (as in [2], [3] and [4]) and the introduction of tense markers for reporting the temporal struc-

⁹ I am sympathetic to two further claims about such utterances, although I won't defend them here. First, Progovac's proto-syntax may capture the utterances produced by enculturated chimpanzees and bonobos (Savage-Rumbaugh et al. 1998; Rivas 2005). Second, this proto-syntax may resemble pre-verbal thought in both human and other great ape species.

¹⁰ Some may regard the transitions from [1] to [2]–[5] as transitions from non-propositional to propositional content. I do not. On the measurement theoretic account of propositions I prefer (Matthews 2011), there is no hard distinction between propositional and non-propositional contents.

ture of an agent's actions ([4] and [5]).¹¹ With the development of sentential structures that facilitated the unambiguous expression of relatively complex propositions, a new platform would emerge against which attitude expressions could be developed.

5.3 From propositions to propositional attitude psychology

If explicit ToM is grounded in the ability to model propositional attitudes in language then it is also necessary to say something about the origin of propositional attitude reports. Evidence from comparative linguistics supports the hypothesis that the first propositional attitude verbs described perceptual relationships that could be tracked using non-verbal perceptual mechanisms—e.g.:

Kofi saw that the food is there.

There are several reasons for thinking that such reports were the foundation for more abstract epistemic state reports. First, since perceptual states can reliably be tracked using only behavioural cues—e.g., whether or not something has crossed an agent's line of sight (Butterfill & Apperly 2013)—perceptual state descriptions could emerge in the absence of a developed ToM. The underlying mechanisms could be explained in terms of general-purpose learning (Heyes 2018) or a phylogenetically old cognitive mechanism for tracking agents' object registrations (Apperly and Butterfill 2009). Second, more abstract epistemic states, i.e. those that co-vary less reliably with behavioural cues, can be conceived of as superordinate categories for combining elements of perceptual deliverances. For example, knowledge might initially have been construed as an epistemic state that combined the deliverances of the senses without specifying the modality by which some proposition was known. Third, in many languages, verbs related to knowledge are etymologically derived from perception verbs. Based on comparative studies of language, Sweetser (1990) held that the extension of perception verbs to cognition verbs was likely a feature of all languages:

The objective, intellectual side of our mental life seems to be mainly linked with the sense of vision, although other senses ... occasionally take on intellectual meanings as well. There are major similarities in our general linguistic treatment of vision and intellection. (*ibid.*, p.37)

Sweetser shows that knowledge terms in Indo-European languages originated from the metaphorical extension of verbs related to seeing (*ibid.*). Nonetheless, contrary to her prediction, the foundational status of vision with respect to knowledge talk is not universal. In aboriginal languages like Dalabon verbs related to knowledge and thinking derive from hearing verbs. In one language, Warluwarra, a single verb *-rlari* means both 'to hear' and 'to think' (Evans and Wilkins 2000). While the Australian languages falsify Sweetser's prediction, they are consistent with the weaker but related hypothesis that verbs related to knowing derive from perception terms.

¹¹ Progovac (2015) also argues the ability to combine constructions hierarchically required natural selection for language-specific syntax. This may be but her view of the stages of syntax development is consistent with alternative accounts. For example, syntactic complexity in natural languages might have emerged under selection pressure for general-purpose cognitive abilities [e.g., sequence learning (Chater and Christiansen 2016)] and the cultural evolution of natural language grammars.

While perceptual state reports are one possible source of epistemic state verbs (and their accompanying concepts), they are not the only one. Mental state terms related to knowing and believing can also derive from direct speech reports (Jary 2010; Geurts in press, *Mind and Language*). Very often people's beliefs can be inferred from what they say. Thus we can imagine a metaphorical extension of the use of speech report terms like 'said that p ' to cover cases in which a speaker had not said that p , but where commitment to the truth of p could be inferred from her actions. Such constructions are used in another Aboriginal language, Ungarinyin, spoken in North-Western Australia. In that language the sentence "gunin nya-nga-yi-minda a-ma jirri" can be translated both as *He says: 'I will cover her'*, and *He thinks: 'I will cover her'* (Geurts in press, *Mind and Language*; via Spronck 2016, p. 259).¹²

These potential roots for the development of mental state talk are not in competition. It may be that knowledge talk in all languages has derived from perception talk, and belief talk always from speech talk, but there could also be cultural variation in these patterns. Future empirical research may provide us with clear answers. In the meantime I propose a number of conclusions. (1) There are multiple possible routes to the development of the mental state verbs via the metaphorical extension of verbs used for reporting perception and speech. (2) These cases show that epistemic terms can be conceived of as abstractions from verbs grounded in descriptions correlated with behaviour—without presupposing a developed understanding of mental states. (3) In light of the potential for variation in the construction of mental state verbs, there is reason to expect empirical variation in how different groups of language users think and talk about minds. Each of these claims is consistent with the claim that our uniquely human propositional attitude psychology is enabled by the cultural evolution of models of human psychology formulated in natural language.

5.4 Non-Gricean accounts of language development

Before concluding, more must be said about the developmental relationship between ToM and language. Some sympathetic to language-first accounts of ToM now argue that natural language development can be explained without assuming that pre-verbal speakers can act with and attribute communicative intent (Gauker 2002; Jary 2010; Bar-On 2013; Millikan 2017; Geurts 2019a, b; for influential older works see also Sellars 1956; Dennett 1996).

To illustrate with one example, Geurts (2019) has argued that we can think of the contents of utterances not as ToM-involving expressions of communicative intent, but as overt normative commitments to certain courses of action.¹³ Since these give rise to publicly observable behavioural outcomes they can be grasped without the need for inferences about speakers' mental states. Since any successful account of language development must be consistent with explanations of both how infants come to grasp

¹² The example I use here is adapted from Spronck's original. Thanks to Alan Rumsey (private correspondence) for the verb substitution he suggested..

¹³ For discussion of Dennett's account of infant language acquisition see Moore (2010). Discussion of Bar-On's view can be found in Moore (2018b)—although Bar-On has since incorporated the notion of 'expression pragmatics' into her view (Arnold & Bar-On 2020)..

the contents of others' utterances, and of how our ancestors could develop the earliest natural languages, this raises the question of how creatures lacking an adult-like ToM could know which commitment a speaker has made—i.e. how non-verbal creatures grasp the contents of others' utterances. Geurts argues that where a speaker's words are ambiguous, which commitment a speaker has undertaken can be inferred from factors that are independent of the speaker's mental states. These include considerations of the coherence of rival interpretations (Geurts 2019b), and statistical correlations between utterances and behavioural outcomes (Geurts 2019a). As a result, he argues, there is "no reason to suppose that mindreading is the driving force of pragmatics" (Geurts 2019b, p. 6).

While I agree with Geurts that appeals to statistical correlations play an important role in language development, and that developed ToM will not, I doubt that the range of children's utterance interpretation will be explicable without appeal to an understanding of communicative goals. For infants who lack knowledge of a language, considerations of *sentential* coherence (e.g., of which words frequently co-occur) can play only a minor role in early utterance comprehension. Nonetheless, young children succeed in learning the meanings of words even where they hear the name of a new object only once (making statistical explanations unlikely), and where they do not see the object at the time they hear it named (suggesting that they are not relying only on crude behavioural cues). In a study by Akhtar and Tomasello (1996), an experimenter told 24-month-olds "Let's find the gazzer", before trying and failing to open the door of the barn in which the toy was hidden. Unable to find the toy, the experimenter turned her attention to other things. Nonetheless children later demonstrated that they had learned the gazzer's name. Seemingly they inferred the speaker's referential intention in light of an understanding of her ongoing goal-directed activity.

Cases like this remind us that central to our understanding of communicative behaviour is a more general understanding of goal-directed behaviour. This enables us to interpret ambiguous utterances in light of a prior grasp of what agents are trying to achieve. It is this that provides us with much of the sense of coherence that drives utterance interpretation—reminding us that communicative intentions are just a subset of agents' purposive activity. From this it follows that the best way to make sense of language development will not be to give up on the idea that communicative intentions are foundational, but to develop a more nuanced understanding of the varieties of purposive activity that are at work in ascriptions of communicative intent, and particularly the pragmatic inferences central to language development. Statistical learning will be important here, but statistical inferences will be best understood as part of the evidence that children use to make sense of others' goal directed behaviour, including their goal directed communicative behaviour.

6 Closing remarks

In this paper I have sketched a framework for making sense of the possibility that human ToM emerged on the back of communal language development. I hope also to have shown how new avenues of research might provide further empirical evidence for the hypothesis presented here. The developmental dependence of higher-order min-

reading on language would be illustrated if the relevant ToM abilities are correlated with the mastery of higher order sentential complements and also impaired by tasks that interfere with language cognition (e.g., Newton and De Villiers 2007). If language-based tasks interfere with level-2 perspective taking, this would also be evidence that we use language to construct representations of how things look to others.

The account developed here predicts that there may be undiscovered variation in how historical communities have thought about minds—and in whether they thought about mental states at all. While there is evidence of cultural variation in ToM reasoning, we know relatively little about these differences and even less about how mental states have been conceived historically, and whether cultural differences in mental state talk correlate with differences in ToM. This is something that will be understood only with more systematic empirical research. This could be pursued through comparative linguistic analyses of mental state talk across cultures. If it turned out that some natural languages—including historical languages—lack some or all terms for mental states, this would be evidence for the cultural evolution hypothesis. The evidence would be even stronger if speakers of these languages turned out to perform poorly on language-correlated ToM tasks.

If this evidence is forthcoming, a new project will await us: reconsidering the behaviours of ancestral humans in light of the possibility that they lacked the ToM that we now possess. It will then fall to us to work out when and where in human history uniquely human forms of ToM arose.

Acknowledgements For discussion and written feedback on various incarnations of this paper I would like to thank Dorit Bar-On, Rachael Brown, Steve Butterfill, Nick Evans, Bart Geurts, Celia Heyes, John Michael, Ron Planer, Paula Rubio Fernandez, Alan Rumsey, Kim Sterelny, Tad Zawidzki, and an anonymous reviewer for Synthese. Additional thanks are owed to the Department of Philosophy at the Australian National University. The talk on which this paper was based was written while I was an RSSH Fellow there in March and April of 2018 and was presented at the ANU's 2018 Interdisciplinary Workshop on Social Cognition. The paper was developed in a subsequent trip in August 2019, and presented at the ANU's 2019 workshop on Kinds of Intelligences. This work was supported by an ANU RSSH Visiting Fellowship awarded to Richard Moore in 2018, and by the UKRI Future Leaders Fellowship awarded in 2020 to support Moore's project on The Communicative Mind.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adrian, J., Clemente, R., Villanueva, L., & Rieffe, C. (2005). Parent–child picture-book reading, mothers' mental state language and children's theory of mind. *Journal of Child Language*, 32(03), 673–686.
- Akhtar, N., & Tomasello, M. (1996). Two-year-olds learn words for absent objects and actions. *British Journal of Developmental Psychology*, 14(1), 79–93.
- Appery, I. (2010). *Mindreaders: The cognitive basis of "theory of mind"*. Hove: Psychology Press.

- Apperly, I., & Butterfill, S. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, 116(4), 953.
- Apperly, I., Samson, D., Carroll, N., Hussain, S., & Humphreys, G. (2006). Intact first-and second-order false belief reasoning in a patient with severely impaired grammar. *Social Neuroscience*, 1(3–4), 334–348.
- Arnold, K., & Bar-On, D. (2020). Primate pragmatics, expressive behavior, and the evolution of language. *Animal Behavior and Cognition*, 7(2), 117–130.
- Astington, J. (2006). The developmental interdependence of theory of mind and language. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of Human Sociality* (pp. 179–206). New York: Berg.
- Bar-On, D. (2013). Origins of meaning: Must we ‘go Gricean’? *Mind and Language*, 28(3), 342–375.
- Bar-On, D., & Moore, R. (2017). Pragmatics and the signaller-receiver asymmetry in language evolution research. In K. Andrews & J. Beck (Eds.), *Routledge handbook of philosophy of animal minds* (pp. 291–300). Milton Park: Taylor & Francis.
- Bennett, J. (1978). Some remarks about concepts. *Behavioral and Brain Sciences*, 1(04), 557–560.
- Berio, L. (2020). Culturally embedded schemata for false belief reasoning. *Synthese*. <https://doi.org/10.1007/s11229-020-02655-7>.
- Berwick, R., & Chomsky, N. (2016). *Why only us: Language and evolution*. Cambridge: MIT press.
- Brannon, E. M., & Terrace, H. S. (1998). Ordering of the numerosities 1 to 9 by monkeys. *Science*, 282(5389), 746–749.
- Breheny, R. (2006). Communication and folk psychology. *Mind and Language*, 21(1), 74–107.
- Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind & Language*, 28(5), 606–637.
- Carruthers, P. (2013). Mindreading in infancy. *Mind and Language*, 28(2), 141–172.
- Christiansen, M. H., & Chater, N. (2016). *Creating language: Integrating evolution, acquisition, and processing*. MIT Press.
- Crockford, C., Wittig, R., Mundry, R., & Zuberbühler, K. (2012). Wild chimpanzees inform ignorant group members of danger. *Current Biology*, 22(2), 142–146.
- Crockford, C., Wittig, R. M., & Zuberbühler, K. (2017). Vocalizing in chimpanzees is influenced by social-cognitive processes. *Science Advances*, 3(11), e1701742.
- de Villiers, J. (2005). Can language acquisition give children a point of view? In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind*. Oxford: OUP.
- de Villiers, J., & de Villiers, P. (2000). Linguistic determinism and the understanding of false belief. In P. Mitchell & K. J. Riggs (Eds.), *Children’s reasoning and the mind* (pp. 191–228). Hove: Psychology Press.
- de Villiers, J., & Pyers, J. (2002). Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cognitive Development*, 17(1), 1037–1060.
- Dennett, D. C. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences*, 1, 568–570.
- Dennett, D. C. (1987). *The intentional stance*. Cambridge: MIT press.
- Dennett, D. C. (1995). *Darwin’s dangerous idea*. New York: Simon & Schuster.
- Dennett, D. (1996). *Kinds of minds*. New York: Basic Books.
- Dungan, J., & Saxe, R. (2012). Matched false-belief performance during verbal and nonverbal interference. *Cognitive Science*, 36(6), 1148–1156.
- Eickers, G. (2019). *Scripted alignment: A theory of social interaction*. Ph.D. Dissertation, FU Berlin.
- Evans, N., & Wilkins, D. (2000). In the mind’s ear: The semantic extensions of perception verbs in Australian languages. *Language*, 76, 546–592.
- Everett, C. (2017). *Numbers and the making of us: Counting and the course of human cultures*. Cambridge: Harvard UP.
- Feigenson, L., & Carey, S. (2003). Tracking individuals via object-files: Evidence from infants’ manual search. *Developmental Science*, 6(5), 568–584.
- Flavell, J., Everett, B., Croft, K., & Flavell, E. (1981). Young children’s knowledge about visual perception: Further evidence for the Level 1–Level 2 distinction. *Developmental Psychology*, 17, 99–103.
- Fodor, J. A. (2008). *LOT 2: The language of thought revisited*. Oxford: OUP.
- Frank, M. C., Everett, D. L., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition*, 108(3), 819–824.
- Garfield, J. L., Peterson, C. C., & Perry, T. (2001). Social cognition, language acquisition and the development of the theory of mind. *Mind and Language*, 16(5), 494–541.
- Gauker, C. (2002). *Words without meaning*. MIT Press.

- Geurts, B. (2019a). Communication as commitment sharing: Speech acts, implicatures, common ground. *Theoretical Linguistics*, 45(1–2), 1–30.
- Geurts, B. (2019b). What's wrong with Gricean pragmatics. In A. Botinis (ed.) *ExLing 2019: Proceedings of the 10th international conference of Experimental Linguistics* (pp. 1–9).
- Geurts, B. (in press). First saying, then believing: The pragmatic roots of folk psychology.
- Godfrey-Smith, P. (2005). Folk psychology as a model. *Philosophers' Imprint*, 5(6), 1–16.
- Goldman, A. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford: OUP.
- Gopnik, A., Meltzoff, A., & Bryant, P. (1997). *Words, thoughts, and theories*. Cambridge: MIT Press.
- Grice, H. P. (1989). *Studies in the way of words*. Cambridge: Harvard UP.
- Grice, P. (1957). Meaning. *The Philosophical Review*, 66, s377–s388.
- Grosse Wiesmann, C., Friederici, A., Singer, T., & Steinbeis, N. (2017a). Implicit and explicit false belief development in preschool children. *Developmental Science*, 20(5), e12445.
- Grosse Wiesmann, C., Friederici, A. D., Singer, T., & Steinbeis, N. (2020). Two systems for thinking about others' thoughts in the developing brain. *Proceedings of the National Academy of Sciences*, 117(12), 6928–6935.
- Grosse Wiesmann, C., Schreiber, J., Singer, T., Steinbeis, N., & Friederici, A. D. (2017b). White matter maturation is associated with the emergence of theory of mind in early childhood. *Nature Communications*, 8, 14692.
- Grüneisen, S., Wyman, E., & Tomasello, M. (2015). “I know you don't know I know...” Children use second-order false-belief reasoning for peer coordination. *Child Development*, 86(1), 287–293.
- Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behaviour*, 61(1), 139–151.
- Harris, P., de Rosnay, M., & Pons, F. (2005). Language and children's understanding of mental states. *Current Directions in Psychological Science*, 14(2), 69–73.
- Hauser, M. D., & Carey, S. (2003). Spontaneous representations of small numbers of objects by rhesus macaques: Examinations of content and format. *Cognitive Psychology*, 47(4), 367–401.
- Helming, K., Strickland, B., & Jacob, P. (2014). Making sense of early false-belief understanding. *Trends in Cognitive Sciences*, 18(4), 167–170.
- Helming, K., Strickland, B., & Jacob, P. (2016). Solving the puzzle about early belief-ascription. *Mind and Language*, 31(4), 438–469.
- Heyes, C. (2018). *Cognitive gadgets: The cultural evolution of thinking*. Cambridge: Harvard UP.
- Heyes, C. M., & Frith, C. D. (2014). The cultural evolution of mind reading. *Science*, 344(6190), 1243091.
- Hurley, S. L. (1998). Vehicles, contents, conceptual structure, and externalism. *Analysis*, 58(1), 1–6.
- Hutto, D. D. (2004). The limits of spectatorial folk psychology. *Mind and Language*, 19(5), 548–573.
- Hutto, D. (2007). Folk psychological narratives: The sociocultural basis of understanding reasons. MIT press.
- Jara-Ettinger, J. (2019). Theory of mind as inverse reinforcement learning. *Current Opinion in Behavioral Sciences*, 29, 105–110.
- Jary, M. (2010). Assertion and false-belief attribution. *Pragmatics & Cognition*, 18(1), 17–39.
- Kaplan, R. (1999). *The nothing that is: A natural history of zero*. Oxford: OUP.
- Karg, K., Schmelz, M., Call, J., & Tomasello, M. (2016). Differing views: Can chimpanzees do level 2 perspective-taking? *Animal Cognition*, 19(3), 555–564.
- 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.
- Liddle, B., & Nettle, D. (2006). Higher-order theory of mind and social competence in school-age children. *Journal of Cultural and Evolutionary Psychology*, 4(3–4), 231–244.
- Lillard, A. (1998). Ethnopsychologies: cultural variations in theories of mind. *Psychological Bulletin*, 123(1), 3.
- Liszkowski, U., Carpenter, M., Striano, T., & Tomasello, M. (2006). 12- and 18-month-olds point to provide information for others. *Journal of Cognition and Development*, 7(2), 173–187.
- 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.
- Low, J. (2010). Preschoolers' implicit and explicit false-belief understanding: Relations with complex syntactical mastery. *Child Development*, 81(2), 597–615.

- Low, J., Apperly, I. A., Butterfill, S. A., & Rakoczy, H. (2016). Cognitive architecture of belief reasoning in children and adults: A primer on the two-systems account. *Child Development Perspectives*, 10(3), 184–189.
- Lutz, C. (1985). Ethnopsychology compared to what? Explaining behavior and consciousness among the Ifaluk. In G. M. White & J. Kirkpatrick (Eds.), *Person, self and experience* (pp. 35–79). Cambridge: CUP.
- Maibom, H. (2003). The mindreader and the scientist. *Mind and Language*, 18(3), 296–315.
- Matthews, R. (2011). Measurement-theoretic accounts of propositional attitudes. *Philosophy Compass*, 6(11), 828–841.
- Mayer, A., & Träuble, B. E. (2013). Synchrony in the onset of mental state understanding across cultures? A study among children in Samoa. *International Journal of Behavioral Development*, 37(1), 21–28.
- McGeer, V. (2007). The regulative dimension of folk psychology. In D. Hutto & M. Ratcliffe (Eds.), *Folk psychology reassessed*. Dordrecht: Springer.
- Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development*, 78(2), 622–646.
- Millikan, R. (2017). *Beyond concepts: Unicepts, language, and natural information*. Oxford University Press.
- Moll, H., & Meltzoff, A. (2011). How does it look? Level 2 perspective-taking at 36 months of age. *Child Development*, 82(2), 661–673.
- Moll, H., Meltzoff, A., Merzsch, K., & Tomasello, M. (2013). Taking versus confronting visual perspectives in preschool children. *Developmental Psychology*, 49(4), 646.
- Moore, R. (2010). *Learning to do things with words*. Ph.D. thesis, University of Warwick.
- Moore, R. (2013). Imitation and conventional communication. *Biology & Philosophy*, 28(3), 481–500.
- Moore, R. (2016). Meaning and ostension in great ape gestural communication. *Animal Cognition*, 19(1), 223–231.
- Moore, R. (2017a). Gricean communication and cognitive development. *Philosophical Quarterly*, 67(267), 303–326.
- Moore, R. (2017b). Convergent minds: Ostension, inference and Grice's third clause. *Interface Focus*, 7, 20160107.
- Moore, R. (2017c). Social cognition, stag hunts, and the evolution of language. *Biology and Philosophy*, 32(6), 797–818.
- Moore, R. (2017d). Pragmatics-first approaches to the evolution of language. *Psychological Inquiry*, 28(2–3), 206–210.
- Moore, R. (2018a). Gricean communication, joint action, and the evolution of cooperation. *Topoi*, 37(2), 329–341.
- Moore, R. (2018b). Gricean communication, language development, and animal minds. *Philosophy Compass*, 13(12), e12550.
- Moore, R. (2019). Utterances without force. *Grazer Philosophische Studien*, 96(3), 342–358.
- Newton, A. M., & de Villiers, J. G. (2007). Thinking while talking: Adults fail nonverbal false-belief reasoning. *Psychological Science*, 18(7), 574–579.
- 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.
- O'Madagain, C., & Tomasello, M. (2019). Joint attention to mental content and the social origin of reasoning. *Synthese*. <https://doi.org/10.1007/s11229-019-02327-1>.
- Ochs, E. (1988). *Culture and language development: Language acquisition and language socialization in a Samoan village*. Cambridge: CUP.
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308, 255–258.
- Perner, J., Sprung, M., Zauner, P., & Haider, H. (2003). Want that is understood well before say that, think that, and false belief: A test of de Villiers's linguistic determinism on German-speaking children. *Child Development*, 74(1), 179–188.
- Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that..." attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology*, 39(3), 437–471.
- Peterson, C., & Siegal, M. (2000). Insights into theory of mind from deafness and autism. *Mind and Language*, 15(1), 123–145.

- Poulin-Dubois, D., Rakoczy, H., Burnside, K., Crivello, C., Dörrenberg, S., & Edwards, K., et al. (2018). Do infants understand false beliefs? We don't know yet—A commentary on Baillargeon, Buttelmann and Southgate's commentary. *Cognitive Development*, 48, 302–315.
- Progovac, L. (2015). *Evolutionary syntax*. Oxford: Oxford UP.
- Pyers, J. E. (2006). Constructing the social mind: language and false-belief understanding. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of Human Sociality* (pp. 207–228). New York: Berg.
- 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.
- Rakoczy, H. (2017). In defense of a developmental dogma: Children acquire propositional attitude folk psychology around age 4. *Synthese*, 194(3), 689–707.
- Rakoczy, H., & Behne, T. (2019). Commitment sharing as crucial step toward a developmentally plausible speech act theory? *Theoretical Linguistics*, 45(1–2), 93–97.
- Rivas, E. (2005). Recent use of signs by chimpanzees (*Pan troglodytes*) in interactions with humans. *Journal of Comparative Psychology*, 119(4), 404.
- Rubio-Fernandez, P. (2020). Pragmatic markers: The missing link between language and theory of mind. *Synthese*. <https://doi.org/10.1007/s11229-020-02768-z>.
- Rubio-Fernández, P. & Geurts, B. (2012). How to pass the false-belief task before your fourth birthday. *Psychological Science*, 0956797612447819.
- Rubio-Fernández, P., & Geurts, B. (2016). Don't mention the marble! The role of attentional processes in false-belief tasks. *Review of Philosophy and Psychology*, 7(4), 835–850.
- Ruffman, T., Slade, L., & Crowe, E. (2002). The relation between children's and mothers' mental state language and theory-of-mind understanding. *Child Development*, 73(3), 734–751.
- Savage-Rumbaugh, S., Shanker, S., & Taylor, T. (1998). *Apes, language, and the human mind*. Oxford: OUP.
- Scott-Phillips, T. (2014). *Speaking our minds: Why human communication is different, and how language evolved to make it special*. London: Palgrave MacMillan.
- Sellars, W. (1956). Empiricism and the philosophy of mind. *Minnesota Studies in the Philosophy of Science*, 1(19), 253–329.
- Sober, E. (2005). Comparative psychology meets evolutionary biology: Morgan's Canon and cladistic parsimony. In L. Daston & G. Mitman (Eds.), *Thinking with animals: New perspectives on anthropomorphism* (pp. 85–99). New York: Columbia UP.
- Sperber, D. (2000). Metarepresentations in an evolutionary perspective. In D. Sperber (Ed.), *Metarepresentations: A Multidisciplinary Perspective* (pp. 117–137). Oxford: OUP.
- Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind and Language*, 17(1–2), 3–23.
- Spronck, S. (2016). Evidential fictive interaction (in Ungarinyin and Russian). In E. Pascual & S. Sandler (Eds.), *The conversation frame: Forms and functions of fictive interaction* (pp. 255–275). Amsterdam: John Benjamins.
- Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science*, 18(7), 580–586.
- Sweetser, E. (1990). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge UP.
- Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge: Harvard UP.
- Tomasello, M. (2008). *Origins of human communication*. Cambridge: MIT.
- Tomasello, M. (2014). *A natural history of human thinking*. Cambridge: Harvard UP.
- Tomasello, M. (2018). How children come to understand false beliefs: A shared intentionality account. *Proceedings of the National Academy of Sciences*, 115(34), 8491–8498.
- Tomasello, M. (2019). *Becoming human: A theory of ontogeny*. Cambridge: Belknap Press.
- Toon, A. (2016). Fictionalism and the folk. *The Monist*, 99(3), 280–295.
- Van Cleave, M., & Gauker, C. (2010). Linguistic practice and false-belief tasks. *Mind and Language*, 25(3), 298–328.
- Westra, E. (2017). Pragmatic development and the false belief task. *Review of Philosophy and Psychology*, 8(2), 235–257.
- Wierzbicka, A. (1992). *Semantics, culture, and cognition: Universal human concepts in culture-specific configurations*. Oxford: OUP.
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1), 103–128.

- Xu, F. (2003). Numerosity discrimination in infants: Evidence for two systems of representations. *Cognition*, 89(1), B15–B25.
- Zawidzki, T. (2019). A new perspective on the relationship between metacognition and social cognition: Metacognitive concepts as socio-cognitive tools. *Synthese*. <https://doi.org/10.1007/s11229-019-02477-2>.
- Zawidzki, T. W. (2013). *Mindshaping: A new framework for understanding human social cognition*. Cambridge: MIT Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.