Modularity in language and theory of mind:

What is the evidence?

Michael Siegal         Luca Surian
University of Sheffield     University of Trieste

Correspondence:
Michael Siegal, Department of Psychology, University of Sheffield, Western Bank, Sheffield S10 2TP, UK; e-mail: M.Siegal@Sheffield.ac.uk, telephone: +44 (0)114 222 6506, fax: +44 (0)114 276 6515
Modularity in language and theory of mind: What is the evidence?

One essential characteristic of the human species that permits effective communication is possession of the grammar of language. Grammar is a powerful system that is critical for reducing mistakes in communicating information about potentially threatening events that are remote in time and space. A second essential human characteristic is possession of a theory of mind (ToM) involving the ability to reason about the mental states – the beliefs, desires, and intentions of others and how these differ from one’s own. ToM reasoning is vital for the appreciation and transmission of culture in the form of novels, theatre, and song, and more generally for the maintenance of family and social life.

In this chapter, based on evidence from cognitive developmental psychology, cognitive neuroscience and behavioral genetics, we discuss parallels in the development of grammar and ToM reasoning. Both grammar and ToM are acquired spontaneously and employed effortlessly by all typically developing children. We propose that these processes are the product of specialized, modular systems whose acquisition is characterized by the poverty of the environmental stimulus and whose developmental trajectory is impaired when the requisite input is not received within an early critical period. We then consider proposals that language, particularly some aspects of grammar, serves to support ToM reasoning. We conclude that, to some considerable extent, dissociations between ToM reasoning and grammar are present in both childhood and adulthood but that in typically developing persons these systems interact to support word learning and the emergence of specific cultural beliefs.

1 Parallels between the expression of grammar and ToM reasoning

1.1 Evidence for the poverty of the stimulus argument in the acquisition of grammar

Language is acquired spontaneously without formal instruction. Indeed newborns learn sounds for speech even while they are sleeping (Cheour et al., 2002) and, in the first few years of human development, virtually all children acquire the grammar of their native language. In fact, the grammar of language is mandatory in that we cannot stop ourselves from acquiring it.

One influential view on language acquisition is that the manifestation of structure in children’s language is triggered only by exposure to a linguistic input that is highly limited and fragmented – an indication of the fundamental innateness of grammar. According to the “poverty of the stimulus” account
(Chomsky, 1980; Crain & Pietroski, 2001; Laurence & Margolis, 2001; Newport, 1990; Stromwold, 2000), the acquisition of grammar proceeds automatically in a modular fashion, largely independently of nonverbal intelligence. Despite wide variations in their language environment, children acquire aspects of grammar in a fixed order at about the same time in their development. They make sense of a language input that is compatible not only with the grammar of their own native language but with the grammar of many others. The errors that they do make are highly predictable and often reflect what would be grammatical in another language (Crain & Pietroski, 2001). Further, it has long been established that children are not corrected for the grammaticality of their language but for its truth value (Brown & Hanlon, 1970). That children’s grammar unfolds largely in the absence of feedback on the grammaticality of their utterances is further testament to its biological foundations.

Evidence from deaf children who are cut off from speech input corroborates this account. Pettito and Marentette (1991) found that profoundly deaf infants of deaf parents display manual babbling using a reduced set of the phonetic units in American Sign Language. These results support the view that babbling is tied to the abstract structure of language rather than to input from the speech modality. Further analysis reveals that hearing babies exposed to the sign language of their deaf parents produce low frequency hand movements inside a tightly restricted space in front of the body that corresponds to the signing space in signed languages (Pettito et al., 2001). Moreover, deaf children display similarities in the linguistic structure of their gestural communication despite wide variations in the spoken language of their hearing mothers. The spontaneous gestural communication of American deaf children resembles that of Chinese deaf children rather than that of their own mothers (Goldin-Meadow, 2003; Goldin-Meadow & Mylander, 1998). It involves “language-making” skills that do not require a language model to be activated: segmenting words into morphemes and sentences into words, setting up a system of contrasts in morphology, and building syntactic structures. As shown in a recent study of sequential cohorts of Nicaraguan deaf children exposed initially to a highly degraded language environment, children spontaneously create structures for language acting largely independently of adults (Senghas & Coppola, 2001). Moreover, there appears to be a critical period for language acquisition as, in rare cases when children are not exposed to any language whatsoever, they do appear irreparably impaired

Of course the ‘poverty of the stimulus’ account of language acquisition is not one to which researchers universally subscribe. There are those who emphasize that the input plays a role more important than the role assigned to it by nativist theories (e.g., Cowie, 1999; Tomasello, 2003). Still, additional recent evidence now exists to counteract this opinion.

In a recent ingenious series of experiments, Lidz, Waxman, and Freedman (2003) demonstrated that infants aged 16-18 months comprehend the syntax of the pronoun *one* even though the language environment does not contain sufficient information to guide unaided learning. The infants in their investigation were shown images of single objects on a television monitor such as a yellow bottle. Then in a test phase, the infants saw two objects from the same category such as a yellow bottle and a blue bottle. In the ‘anaphoric’ condition, they heard the phase, “Now look. Do you see another one?” whereas in the control condition, they heard, “Now look. What do you see now?” As predicted infants looked significantly longer at the familiar item in the anaphoric than in the control condition in which they preferred to look at the novel item. Lidz et al. (2003, p. B72; see also Lidz, Gleitman, & Gleitman, 2003) point out that such results demonstrate that “innate linguistic structure guides language acquisition” since the linguistic input available to the infants cannot unambiguously support anaphoric representations.

Consistent with the view that linguistic structure within the language learner is the main source of grammatical knowledge, Gelman (in press, Volume I of this series) documents how young children have an understanding that nouns can be used to refer to generic kinds rather than solely to specific instances and that this understanding is not guided by either perceptual or linguistic cues. Rather, on available data, the expression of a system of generic terms appears to be driven by the theoretical structure that a noun is a generic term unless the context dictates to the contrary. Even deaf children from hearing homes who are without a language model express generic kinds in their gestures (Goldin-Meadow, 2003).

1.2 Evidence for a poverty of the stimulus argument in the acquisition of a theory of mind

A poverty of the stimulus account extends to the course of ToM reasoning in young children. In this
sense, we consider ToM in terms of children’s responses on tasks designed to determine whether they identify how a person with a false belief can initially be deceived about the location of an object or the contents of a container. In particular, these tasks often consist of the “Sally-Anne” task involving unexpected locations for objects (Baron-Cohen, Leslie, & Frith, 1985) and the “Smarties” task (Wimmer & Perner, 1983) involving misleading contents of containers. In the Sally-Anne task, children are told about Sally, a story character with a false belief about the location of a marble. The character is described as having placed the marble in a box but, when she is away, another story character called Anne moves it into a different location. The test question concerns where Sally - who has not witnessed the deception and therefore has a false belief - will look for the marble. In the Smarties task, children are shown a Smarties tube (M&M candies in the US) that, when opened, is revealed to contain pencils. The test question concerns what another person – who again has not witnessed the deception and therefore has a false belief – will think is in the tube.

On such tasks, all typically developing children appear to develop ToM reasoning at about the same time by about 4 to 5 years of age despite wide cultural variations in the extent to which they are exposed to talk about mental states of others (Avis & Harris, 1991; Wellman, Cross, & Watson, 2001). In this sense, the expression of ToM reasoning is parallel to that of grammar that is largely independent of wide variations in the extent to which children are exposed to language input. As would be expected, within this narrow age range, performance on ToM tasks is facilitated when the actual verb used implies that the actor might have a false belief as is the case in languages such as Mandarin and Greek (Lee, Olson, & Torrance, 1999; Maridaki-Kassotaki, Freeman, & Lewis, 2003). However, children who speak English and are younger than 4 years also for the most part succeed if the tasks are made more pragmatically explicit. For example, asking “Will Sally look first for her marble?” enables most (though not all) 3-year-olds children to ‘inhibit’ the interpretation that the question refers to where Sally will have to look or must look for the desired object and instead to interpret the question as intended to refer to the consequences of Sally holding an initial false belief about the location of an object (Joseph, 1998;; Nelson et al., 2003; Siegal & Beattie, 1991; Surian & Leslie, 1999).

Just as children are not taught the concepts of noun, verb and grammatical subject, they are not
taught the concept of belief. Instead, they receive information on the truth value of belief-desire propositions in discourse involving mental state terms – a process that triggers the expression of a concept of belief in which others' beliefs may not correspond to reality. With regard to deaf children, Marschark et al. (2000) have shown that late signers aged 8-13 years have the ability to attribute mental states correctly in generating stories about others with whom they have interacted hypothetically in story situations. However, the late signing deaf children of hearing parents have difficulties on ToM reasoning tasks even in adolescence (Russell et al., 1998). These difficulties persist even on versions of the tasks in which the test questions are made pragmatically explicit (Peterson & Siegal, 1995). In contrast to late signing deaf children, both normal hearing and native signing deaf children appear to enjoy an early conversational access that triggers the expression of belief-desire reasoning.

This pattern of results supports a poverty of the stimulus account for the acquisition of the ToM reasoning in normal children. This acquisition hinges upon receiving the requisite environmental input within a critical period in early development – a process that parallels the acquisition of grammar. Just as children require some minimal access to language for grammar to develop, they require at least some minimal access to conversational opportunities to display normal ToM reasoning skills. Conversational experience alerts children to the concept that others' beliefs can differ from their own and be false. Through this experience, they come to recognize that speakers are epistemic subjects who store and seek to provide information about the world, allowing access to a world of referents and propositions about intangible objects and creating the potential for imagining the past and future (Harris, 1996).

Therefore as is the case for language, the developmental trajectory of ToM reasoning is affected when the requisite input is not received within an early critical period as is the case for late signing deaf children, and possibly as well for some non-vocal children with cerebral palsy who also cannot easily engage in conversation that necessarily involve mental states (Dahlgren, Sandberg, & Hjelmquist, 2003).

A relevant finding is that some children with autism also display protracted difficulties on ToM tasks (Baron-Cohen et al., 1985). As discussed elsewhere (Siegal & Blades, 2003), one interpretation of this result is that autism may reflect an impairment in processing or attending to auditory information
that prevents full engagement in conversational exchanges and contributes to the preferential interest of persons with autism in objects and physical causality rather than in people and psychological processes. In such circumstances, children can be persistently impaired in appreciating that the minds of others contain a store of epistemic mental states, including false beliefs and beliefs that differ from their own.

In this sense, it is instructive to contrast concrete objects, e.g. ‘bees’, with ‘beliefs’ and other mental states. Children can point to bees – and concrete referents in general – to communicate messages. However, they cannot rely on such ostensive acts to point and communicate about false beliefs. To be able to share the meaning of a false belief, they need to exercise their capacity to inhibit the prepotent response that arises from a very simple ToM – one that operates under the premise that beliefs and reality truly correspond as these very often do (Fodor, 1992; Leslie, 2000). This process takes place within conversational exchanges with others about the nature of the inner worlds of mental states. Children in conversation are regularly faced with situations in which speakers may hold different beliefs or perspectives. Indeed to participate appropriately in conversation children have to keep these differences in mind (Clark, 1997). Their full development as shown on ToM reasoning tasks may require extensive exercise, and it is just the daily involvement in conversation that may give children the opportunity to practice the inhibitory skills required in false belief tasks.

2 Language and theory of mind: What is the relationship?

2.1 Does grammar provide the representational template necessary for ToM reasoning?

It has been widely reported that there is a correlation between ToM and language as shown on measures of grammar as well as knowledge of vocabulary and semantic word usage, both in typically developing children (Astington & Jenkins, 1999) and in children with autism (Happé, 1995). However, despite this well-documented relationship, the nature of the language-ToM relationship, and the extent to which language influences ToM reasoning as shown in performance on false belief tasks remains controversial. Some have claimed that the grammar of language enables children to entertain propositions that involve the simultaneous representation of alternative state of affairs such as the consequences of behavior by individuals who hold true or false beliefs (Astington & Jenkins, 1999; Perner, 1991; Smith, Apperly, & White, 2003). More specifically, others have maintained that it is the
acquisition of sentence complementation in the grammar of language that enables children to reason out solutions to false beliefs (De Villiers & Pyers, 2001). By this account, ToM reasoning is dependent on the possession of syntactic structures such as those that permit the embedding of false propositions within true statements (‘Mary knows that John (falsely) thinks chocolates are in the cupboard’).

However, it is likely that neither of these hypotheses fully captures the contribution of language to theory of mind. It may be that a certain level of syntax and semantics is necessary for ToM performance but, nevertheless, many young children are adept at syntax and semantics but still do poorly on ToM tasks. Although Hale and Tager-Flusberg (2003) and Lohmann and Tomasello (2003) have reported success at training ToM performance with exposure to instruction on sentence complementation, Ruffman et al. (2003) report evidence that ToM reasoning is related to general language ability rather than to specific aspects of syntax or semantics. Moreover, as Lohmann and Tomasello (in press) recognize, training studies on sentence complementation may in fact involve exposure to discourse that may foster conversational understanding which in turn promote success on false belief tasks.

As has been previously noted (Astington & Jenkins, 1999; Custer, 1996; Woolfe, Want, & Siegal, 2002), 3-year-olds who fail ToM tasks spontaneously produce sentence complements in their speech. They correctly answer questions involving sentence complementation if those sentences take the structure [person]-[pretends]-[that x] (e.g., “He pretends that his puppy is outside”). By contrast, 3-year-olds do poorly when given sentences that take the form [person]-[thinks]-[that x] (e.g., “He thinks that his puppy is outside”). Both use the same complements yet children only pass when “pretend” is used. Given these considerations, the syntax of sentence complementation falls short of providing a complete account of ToM performance, at least on pictorial tasks.

Converging evidence comes from studies of adults following brain damage who have become aphasic and have lost grammar though retaining their ToM reasoning ability (Varley & Siegal, 2000; Varley, Siegal, & Want, 2001). Though such patients have a language-configured mind that could be seen to support ToM development, their performance is consistent with the dissociation between grammar and ToM in childhood. Finally, there are many instances of sign languages and spoken Aboriginal Australian languages in which there is no sentence complementation (M. A. Baker, personal
communication). Instead of clausal complements such as “John told everyone that Mary washed the car”, users of such languages instead employ “clausal adjunct” forms such as “Mary having washed the car, John told everyone (it).” If complementation were necessary to instantiate ToM reasoning, no ToM would be possible in these language groups.

Grammar may thus be seen as a co-opted system that can support the expression of ToM reasoning but the possession of grammar does not guarantee successful performance on ToM tasks (Siegal & Varley, 2002; Siegal, Varley, & Want, 2001). Rather, ToM reasoning in young children is triggered, tuned and speeded up by engagement in conversation about mental states contents, such as what speakers want, pretend and believe.

2.2 Is comprehension in conversation achieved by a specialized submodule of ToM?

Since access to opportunities to converse about mental states appears to be pivotal in the expression of ToM, it is important to examine links between ToM and comprehension in conversation. To this end, it is useful to distinguish between links at two levels: functional and ontogenetic. At the functional level we discuss how and to what extent the ToM module is involved in conversation and how it connects with other cognitive components to allow the successful interpretations of communicative acts and the production of context appropriate utterances. In relation to the ontogenetic level, we aim to sketch how the acquisition of ToM relates to the development of communication skills.

2.2.1 Considerations at the functional level

There is a consensus that beliefs about beliefs, termed “metarepresentations”, are necessary to human communication, particularly to inferential communication (Grice, 1989). If we posit that metarepresentations are the output of a specialized mechanism, the ToM module, then the necessary and central role of the ToM module in conversation is apparent and is well recognized in current theories of human communication (Sperber, 1996). The claim that metarepresentations are necessary for many aspects of human communication is in line with common sense intuitions about conversational processes, but there is also abundant empirical support for it that comes from normally developing children and persons with autism. For example, persons with autism frequently
show a deficit on ToM reasoning and this deficit is associated with their ability to detect violations of the Gricean conversational maxims (Surian, Baron-Cohen, & van der Lely, 1996) and understand figurative language such as metaphors and irony (Happé, 1993).

Therefore the ToM module is a necessary component of normal conversational competence in school age children and adults. Sperber and Wilson (2002) have recently proposed an interesting analysis of the relation between ToM and conversation that goes beyond this view. According to their analysis, pragmatic inferences involved in communication are computed by a specialized sub-module that belongs to the human mind-reading system, rather than by the same ToM mechanism that is used to make sense of, and to predict, actions in general. To support this proposal, they emphasize that such inferences are usually processed in a fast and unconscious way, are drawn even by preverbal infants, and concern a specific type of input. Moreover, the domain on which they operate is quite “special”. Communicative acts, compared to the other non communicative actions, exhibit this peculiar character: they can be about an infinite range of (informative) intentions. By contrast, the intentions underlying others’ actions as portrayed in ToM tasks seem comparatively limited, given that real world constraints apply to real actions, but not to the semantic content of communicative acts. There is so much more that you can say compare to what you can do, or even try to do. Also, inferences of this sort have certainly been part of the human social interaction for long enough to make it plausible for a specialized sub-module to have been selected given its adaptive value. On this view, there is a dedicated module to retrieve a speaker’s meaning that is part of a larger ToM module. The proposal that there is dedicated sub-module to retrieve a speaker’s meaning is a departure both from Grice and from Sperber and Wilson’s (1986) own previous work.

We wish to discuss two main claims by Sperber and Wilson (2002) model. The first is that human communication involves a great deal of metapsychological processing because it is, at least for a substantial part, inferential in nature, rather than code-like. The second claim is that such metapsychological processing is the job of a specialized modular subsystem dedicated to pragmatic inferences.

We agree with the idea that human communication is in large part inferential, but we do not think
that the evidence for this comes from the research showing the involvement of ToM in communication. Metarepresentations are a necessary requirement not only in inferential models of communication, but also in many code models of communication, such as the mutual knowledge model proposed by Clark and Marshall (1981; see Sperber and Wilson, 1986, pp. 15-21 for a contrary view). The critical difference concerns the presence of a shared coding-decoding system, which is required only by the coding view. On this view, verbal and non verbal communication are based on the shared knowledge of verbal and non verbal codes, respectively. By looking at me and then looking at the door the speaker sends me a coded message that I will decode as ‘she thinks it is time to go’ and this interpretation is achieved without many assumptions about the rationality of the speaker. By contrast, on the inferential view, no coding systems are necessarily involved on this exchange, and the correct interpretation is derived by drawing a series of deductions based on the assumption that the speaker is cooperative and rational and she treats the addressee as a rational agent. However, since both views require metarepresentation and the involvement of metarepresentation does not allow us to decide between the two alternatives.

Rational processes play an important role in the Sperber and Wilson model of utterance interpretation. Hearers choose the most accessible interpretation among those that are contextually plausible or available. They do this by using a rational procedure that follows a path in which the effort required in constructing an appropriate interpretation is minimized. The chosen interpretation may not necessarily be the correct one, but it is the most rational interpretation given an expectation of relevance. However, the claim that hearers ‘choose’ a specific interpretation is somewhat at odd with the claim that they often stop at the first interpretation they construct because this interpretation satisfies the expectation of optimal relevance. If a hearer does not even access or represent other contextually possible implications of the speaker’s utterance (Sperber & Wilson, 2002, p. 20), than her ‘choice’ of one interpretation is in the eye of the observer rather than the hearer’s head.

The evidence from the infant literature is also suggestive at best, but certainly not conclusive. While preverbal infants communicate rather efficiently in ways that seem to exploit metapsychological resources, children may achieve their communicative success without the need to represent others’
beliefs and desires (see Gergely et al., 1995; Gergely & Csibra, 2003). At present, there is no convincing demonstration that preverbal infants can use metarepresentations. Even if we insist that infants are indeed capable of inferential communication, then true inferential communication might be achieved without the need for metarepresentational skills. This of course does not rule out that infants can evaluate the rational grounds of some actions (Gergely et al., 1995; Csibra et al., 1999). But one thing is to evaluate the rationality of an action given some biomechanical constraints and another is to evaluate the rationality of a communicative act or to interpret it assuming the rationality of the speaker or its optimal relevance. In this later case, since both costs and benefits are mostly defined in cognitive terms, it is hard to point out how one could do it without metarepresentations.

Turning to Sperber and Wilson’s second claim, the idea that there is a sub-module of the mind-reading system that is dedicated to pragmatic inferences may sound very strange to anyone who is used to think of modules as mechanisms that are relatively context independent and of pragmatics as the part of conversational competence that deals with the relationship between utterances and the communicative context. Nevertheless, this idea is appealing to those who are seriously concerned with the speed and accuracy with which utterances are constructed and interpreted in real life conversation. It is speed, automaticity and domain specificity, rather than informational encapsulation, that underscores the modular nature of the mechanism envisaged in Sperber and Wilson’s proposal. On the one hand, speed may just be the result of good automatization reached by means of practice, rather than of a dedicated sub-module (Bloom, 2002). On the other hand, speakers appear to be very fast in odd situations that are very different from what they have encountered before. These achievements point neither to a practice effect nor to complex inferential chains like those envisaged by Grice as the basis for following the implications of conversations.

Consider, as a working example, the case of ‘scalar implicatures’ (Carston, 1998). If a speaker says “I have two brothers” the hearer would usually interpret this as implying “exactly two brothers”. Or, if one say ‘x might be y’ adults speakers tend to rule out that this can be interpreted also as ‘x must be y’, despite the logical compatibility of such interpretation. Noveck (2001) found that children did not show the same bias, suggesting that children have difficulties in drawing
conversational implicatures. An investigation about the sources of such difficulties showed, however, that when the goal of the experimenter is made clearer, via a short training on pragmatic evaluation of utterances, children’s peculiarities are drastically reduced (Papafragou & Musolino, 2003). In a Gricean perspective, the hearer assumes the speaker’s cooperativeness, and rules out that she means ‘at least two brothers’, or ‘x must be y’ even if logically such interpretation are indeed compatible with the speaker’s utterance. Such interpretations are very often explicitly expressed when subjects are asked to correct an infelicitous utterance. In Papafragou and Musolino’s experiments, 6-years-olds corrected the speaker that said ‘some horses jumped over the fence’ when indeed the case was that all horses jumped the fence. However, more evidence is needed to support the claim about the psychological reality of conversational implicatures.

Suppose that we present a 6-year-old child with a set of 6 clowns, 3 are happy and 3 are sad. In each subset, one has a blue flower, one has a red flower and one has no flowers. If we ask the child, “Give me the happy one,” he will choose the happy clown without flowers (Surian & Job, 1987). This choice could be the outcome of children’s ability to draw a conversational implicature, roughly summarized as “He did not say anything about flowers. He would have said something about it if he intended to point to a clown with flowers. I can assume that he is cooperative and he genuinely wants me to know which clown he intends to refer to. Therefore he must mean the clown with no flowers.” In these settings, the nonverbal context ensures that the child readily accesses potential alternatives and chooses between them. But is this really a choice based on the recognition of an implicature? An alternative, more likely, process may have simply involved the construction, in the child’s head, of a ‘happy clown face’, or the representation of the feature ‘happy’, and then a search for the best match in the contextually available objects. This match is found with the object that exhibits the mentioned feature, but no other salient features that were found in the other alternatives. A computation of relevance may have indeed guided the child towards a specific (‘correct’) referent with no need for complex counterfactual reasoning concerning what the speaker would have said had he wanted to refer to a different clown. The speed and accuracy with which even young children perform such a task does not support a long and reflexive process, but is instead
coherent with Sperber and Wilson proposal about a dedicated pragmatics sub-module. However, the pragmatic inferencing posited by Relevance theory is drastically reduced compared to the inferencing required in the traditional Gricean view.

A similar process may happen when the child is tested for the use of the mutual exclusivity constraint in word learning. The child is presented with a couple of objects, one familiar (e.g. a spoon) and one unfamiliar (e.g. a whisk) and is asked, “Show me the fendle.” The child, even a preschooler, would readily point to the unfamiliar object. One way of explaining such success is to assume that the child will reason in this way (Bloom, 2000, p. 68):

I know that a banana is called a banana.
If the speaker meant to refer to a banana, she would asked me to show her the banana.
But she didn’t, she used a strange word, fendle.
So she must intend to refer to something other than the banana.
A plausible candidate is the whisk. Fendle must refer to the whisk.

Bloom maintains that this explanation is roughly what previous research on the use of the mutual exclusivity constraint seemed to support and that it makes unnecessary premises that are specific for word learning. Children, by contrast, appear to perform similarly when they need to infer many object properties, not just its name (Diesendruck & Markson, 2001). It is, however, possible that children use a ‘matching strategy’ both when they are inferring the name and when inferring the function, or other properties of an object. This matching strategy, like in the ‘clowns experiments’, may not require the level of mental state attribution and pragmatic reasoning outlined by Bloom. It would be more similar to the explanation based on the mutual exclusivity assumption as summarized by Markman, Wasow and Hansen (2003): “that can’t be a fendle, it’s banana”.

To summarize, the main reasons to be skeptical about the existence of an early emerging pragmatic module are that (1) there is no evidence that preverbal infants draw such metapsychological inferences, (2) the adaptive value of such putative module is obviously not a demonstration of its existence, and (3) the unconscious or non reflexive nature of (some) pragmatic inferences is coherent not only with a modular cognitive component but also with an automatized process, based on experience, as suggested by Bloom (2002).
Our view is that it may be worthwhile to differentiate among different types of pragmatic inferences and to propose that some of these inferences are drawn by dedicated modules whereas others are drawn by more general inferential processes involved in ToM reasoning. In other words, we do not propose that there is single ‘pragmatics module’ but rather a set of highly specialized modules for some instances of pragmatic inferencing. For example, the inferences involved in reference resolution when the direction of the speaker’s gaze is available may be carried out by a dedicated submodule that takes, as its input, the output of the Eye Direction Device postulated by Baron-Cohen (1995) and the output of the language module. We see no convincing reason to think that the mechanism dedicated to this kind of computation should also be the same mechanism exploited to interpret sarcasm and metaphors.

To make the sub-module envisaged by Sperber and Wilson work one also needs to endow it with access to a very wide sort of information, since potentially any piece of information may be relevant to utterance interpretation. This, however, is not compatible with domain specificity and informational encapsulation. A family or more specialized submodules may turn out to be more efficient, especially at the beginning of intentional communication.

Selective impairment is one indispensable source of evidence for modularity (Fodor, 1983). We have now a substantial body of evidence showing selective impairment of ToM, not only in children with autism, but also in brain damaged populations. Following damage to the right hemisphere, many adult patients have difficulty on ToM reasoning tasks though they retain grammar in their language (Happé, Winner, & Brownell, 1999; Surian & Siegal, 2001). By contrast, following damage to the left hemisphere language centers, many patients become aphasic as shown by loss of grammar though they retain ToM skills (Varley & Siegal, 2000; Siegal, Varley, & Want, 2001). The known selective breakdowns, either in impaired adults or in atypically developing children, do not provide clear evidence to support the idea of a dedicated pragmatic module. To date, there are no reported cases in which difficulties in inferential communication are not accompanied by differences also in non communicative mind-reading skills such as those tested in false belief tasks. The absence of relevant evidence, however, may simply be due to the fact that the methods used so far were not
suitable to detect such a selective impairment. Sperber and Wilson’s (2002) proposal may foster future studies that are specifically designed to assess the predicted dissociation, as was the case for the modular model of Theory of Mind (Baron-Cohen, Leslie & Frith, 1985).

2.2.2 Considerations at the ontogenetic level

The research on ToM reasoning in deaf children suggests that access to conversation is necessary for the development of ToM. One way of portraying the role of conversation in the development of ToM is to see conversation as the situation in which children are provided with crucial input to learn what a mental state is, what kinds of mental states people can entertain and how they come to entertain them. This process would lead them to abandon immature theories of actions that include teleological, but not mentalistic concepts. The poverty of the stimulus account of the expression of ToM reasoning is contrary to such a view. We argue for an alternative hypothesis which recognize both the presence of a rich innate competence and the necessity for specific experiences during a critical period. In our view conversation is a powerful, perhaps the most powerful source of exercise of metarepresentational skills. During conversation, speakers are required to constantly update their representations of their interlocutors’ mind and to infer their informative and communicative intentions (Grice, 1989). In comparison, the mind reading required in the interpretation of daily actions appears to be a much less intensive and demanding task. If children are deprived of this exercise they may be prevented from strengthening the links between ToM and central processes systems or the pathways required to access ToM representational resources; this in turn would result in poor performances on ToM reasoning tasks; however, such impairment would not be as severe as the metarepresentational deficit reported in most autistic children. This view can account for the experimental evidence of late signers difficulties in ToM tasks, which are very similar to the difficulties reported in autistic children, and also the naturalistic reports of deaf children ability to establish friendships and enjoy social interactions, which are in contrast with the poor social abilities reported in autistic children.

Evidence from infant communication, for example referential pointing and shared attention activities, does not reflect a metarepresentational understanding, but rather the ability to infer intentions, goals and possibly perceptual or attentional states. Children display an actional understanding of agents (Leslie,
1994) when they are involved in early forms of intentional communication, not yet a cognitive understanding. Although early forms of metarepresentation may not be necessarily tied to verbal communication (i.e., conversation), it is with verbal communication that the communicative activity comes to a level of sophisticated comprehension and complexity unmatched by other communicative activity. Competent participation in conversation requires continuous and very fast computation of participants’ mental states. It is therefore only with such activity that humans are required to exploit and instantiate fully their metarepresentational skills.

3 The genetics of language and ToM

Further evidence for the dissociation between grammar and ToM comes from studies in early infancy on the precursors of language. Holowka and Petitto (2002) videotaped 10 babies aged 5 to 12 months acquiring either English or French. They scored 150 randomly selected segments of babbles, nonbabbles, and smiles in terms of left, right, or equal mouth opening and found that babbling was accompanied by right mouth asymmetry, whereas smiling was accompanied by a left mouth asymmetry and nonbabbling by equal mouth opening. The left hemisphere cerebral lateralization for language indicated by right hemisphere asymmetry in babbling clearly indicates that left lateralized language functions in adults is present very early in human development before the actual onset of speech. These findings are in keeping with neuroimaging studies that point to left hemisphere dominance for speech perception in the first year of life (Dehaene-Lambertz et al., 2002), though greater plasticity in the maturing brains of infants may allow for more right hemisphere substitution than in adults following right hemisphere damage (Dehaene-Lambertz et al., 2004).

Given the presence of this very early asymmetry, is there support for a genetic basis to the left hemisphere structures that control language? Using a sample of 10 monozygotic and 10 dizygotic twins, Thompson et al. (2001) sought to investigate the influence of individual genetic differences on brain structure as shown on three-dimensional maps constructed from magnetic resonance images. Despite the underpowered sample size, they found highly significant heritability in the asymmetry of Broca’s and Wernicke’s language areas in the left hemisphere.

Similarly, twin studies using large samples have revealed substantial nonoverlapping genetic
influences on phenotypic measures of language and nonverbal intelligence in infancy and early childhood (Dale et al., 2000; Price et al., 2000), although the genetic overlap may be greater at later ages (College et al., 2002). In the case of specific language impairment, both behavioral and molecular genetic research indicate powerful and enduring genetic influences on grammar in both children and adults that nonetheless can spare nonverbal intelligence (Dale et al., 1998; Lai et al., 2001; Spinath et al., in press; Van der Lely, Rosen, & McClelland, 1998). Neuroimaging experiments using fMRI have been carried out involving members of the KE family who have a language disorder caused by a mutation in the \textit{FOXP2} gene. These have shown that affected family members display underactivation in Broca’s area on a verb generation task compared to unaffected members who have a typical left-dominant pattern of activity in this area, pointing the critical involving of the \textit{FOXP2} gene in the neural substrate of language (Liégeois et al., 2003).

Few studies have directly investigated genetic influences on language and ToM. In a study of 3-year-old twins, Hughes and Cutting (1999) reported substantial nonoverlapping genetic influences on measures of verbal intelligence and ToM reasoning such as the Sally-Anne task. By contrast, in a twin study of 5-year-olds, Hughes et al. (2004) found that environmental factors explained most of the variance on ‘advanced’ ToM reasoning measures and that the only genetic factors that influenced ToM were those that were shared with verbal ability. However, in the latter investigation, children were given ‘second-order’ ToM story tasks about a character’s beliefs about another character’s beliefs. In the predominantly low SES sample tested, children need to have attained sufficient verbal ability to bear in mind the complex premises of each second order task and to reason successfully. Such reasoning ability may be more likely to be influenced by the social environment of family, peers, and schooling than is the case of the simple first-order ToM reasoning of 3-year-olds preschoolers.

4 Mental modularity and cultural diversity

The evidence that we have reviewed indicates that grammar and ToM reasoning is the product of mechanisms that are modular to a significant degree. Dissociations between grammar and ToM performance on cognitive tasks, and precursors in infant behavior, provide evidence for cognitive modularity. Neuroimaging and patient lesion studies that demonstrate dissociations between grammar
and ToM in brain activation and function provide evidence for neural modularity. There is also evidence for genetic modularity insofar as there is a strong genetic basis for the left hemisphere language structures and performance on measures of verbal intelligence that do not overlap with performance on measures of nonverbal intelligence. Genetic studies also highlight largely nonoverlapping genetic basis for grammar and ToM as shown on false belief reasoning tasks. Thus while there is no necessary connection among forms of modularity – a dissociation between grammar and ToM on cognitive tasks need not have a corresponding neural or genetic substrate (Coltheart, 1999), data exists and continues to accumulate to support the specialized, modular nature of grammar and ToM at all three levels.

Based on this research, our view is that there are communalities in the capacity for - and emergence of - grammar and ToM in that both represent the elaboration of innate processes that are achieved automatically and effortlessly by typically developing children. In this sense, grammar and ToM can be seen as parallel modular systems that come together to provide a foundation for the transmission of culture (Sperber, 1996).

Drawing a parallel between the expression of grammar and the expression of ToM creates insight into the nature of culture in relation to universals in cognition. Humans regardless of culture acquire a grammar. Culture determines the specific nature of the native grammar to be acquired. Similarly, all humans regardless of culture acquire the concept of a belief that possesses true and false attributes once it is triggered by exposure to conversation about the mental states of others, and in doing so manifest a theory of mind. Culture influences the specific beliefs that people hold about the minds of others and shapes non-core aspects of ToM.

In this sense, ToM and grammar emerge as autonomous domain-specific systems that normally come online at set times in development despite wide variations in the environment. These systems interact to support word learning and the acquisition of specific beliefs.

In tandem with cues from the grammar of language, ToM in the form of the ability to interpret others’ intentions contributes substantially to how children learn the meaning of words (Bloom, 2000; Diesendruck & Markson, 2002). For example, Gelman and Eberling (1998) gave children aged 2-3 years drawings of various nameable objects (e.g. a man). Each drawing was described as illustrating a
shape that was created intentionally (e.g. someone painted a picture) or created accidentally (e.g. someone spilled some paint). Participants were simply asked to name each picture. Children used shape as the basis for their naming primarily when the shapes were intentional and substance (paint) when the shapes were accidental. In this way, they displayed evidence of sharing the speaker’s viewpoint in conversation that is vital for effective communication.

With the support of grammar and ToM, children also acquire the specific languages and beliefs of their community. These languages and beliefs are encrypted to be accessible to those within a culture – and function to protect it. As Baker (2001) remarks (see also Sperber, 1996), the parameters of variation in language in particular and in culture more generally have many of the same properties as engineered codes and ciphers (with a secret key) insofar as these properties function to conceal a message, rearrange its parts, and replace its symbols at different levels of structure. More generally, the factoring of language into a universal grammar available to everyone and parameters encrypted to be accessible to the few suggests that language variation is not an evolutionary accident. Instead, it is part of the inherent design specifications for communication that have the goal of producing messages that are easily understandable by the intended audience but not by those outsiders who may attempt to listen. Modular systems in theory of mind and grammar interact to form the basis of problem solving resources children use to acquire words and culture, but their autonomy is reflected in the domain-specific breakdown of function following brain lesions in adulthood.

One of the main functions of human culture is to clarify what people value, what they take seriously in their daily lives, what they will fight for to include or exclude others in their groups (Premack & Hauser, 2001). Founded on the capacity for grammar and ToM, enculturation involves specific languages and beliefs that are encrypted to be easily accessible only to those within a culture. The human mind possesses the capacity to marshal a series of autonomous modular systems such as grammar and ToM. In this way, as we have previously maintained (Siegal & Varley, 2002), the mind has created an extended functional processing system in which the sum is greater than its parts.


