Issues in Coevolution of Language and Theory of Mind

Description

The relation between language and theory of mind remains in need of clarification, both at the level of language evolution, language acquisition and the very content of theory of mind. This raises the question of the very nature of theory of mind. Is it a monolithic, more or less modular mental faculty; or is it a combination of different mechanisms, some of which may be rather low-level? How much theory is there in theory of mind and how much is needed to evolve a language? Very much the same questions apply to language acquisition. This workshop will attempt to analyse the coevolution of these two uniquely human capacities, their co-dependence and interaction.

The Workshop is organized by the Institut des Sciences Cognitives CNRS, Lyon. Starting from February 2004, a new paper will be put on line and open to discussion every two weeks. The research presented in this workshop is supported in the framework of the European Science Foundation EUROCORES programme.

French version:

La co-évolution du langage et de la théorie de l'esprit La relation entre le language et la psychologie naïve demande beaucoup de clarification, au niveau de l'évolution du langage, de son acquisition et du contenu même de la psychologie naïve. Cela soulève la question de la nature même de la psychologie naïve. S'agit-il d'une compétence monolithique, plus ou moins modulaire, ou plutôt d'une combinaison de mécanismes, dont certains de très bas niveau ? Combien de psychologie naïve a-t-on besoin pour évoluer un langage ? Les mêmes questions se posent autour de l'acquisition du langage. Ce colloque essayera d'analyser la co-évolution de ces deux capacités exclusivement humaines, leur co-dépendance et leur interaction.

Ce colloque est organisé par l'Institut des Sciences Cognitives CNRS, Lyon. A' partir de Février 2004 un nouveau papier sera ouvert à la discussion toutes les deux semaines. La recherche présentée dans ce colloque est réalisée dans la cadre du programme EUROCORES de European Science Foundation

Moderators:

Peter Ford F. Dominey (CNRS - Institut des Sciences Cognitives, Lyon), **Gloria Origgi** (CNRS, Institut Jean Nicod, CNRS), **Anne Reboul** (CNRS - Institut des Sciences Cognitives Lyon)

Guest Panel:

Ingar Brinck (Lund University), **David Caplan** (Harvard Medical School), **Stefano Cappa** (Center for Cognitive Neuroscience - Ospedale San Raffaele, Milan), **Peter Carruthers** (University of Maryland), **Robyn Carston** (University College, London), **Cristiano Castelfranchi** (CNR – Italy), **Rhiannon Corcoran** (University of Manchester), **Viviane Deprez** (Institut des Sciences Cognitives, Lyon), **Jean-Louis Dessalles** (Ecole Nationale Supérieure des Télécommunications), **Keith Frankish** (Open University), **Peter Gärdenfors** (Lund University), **Jim Hurford** (Edinburgh University), **Cristina Meini** (Università del Piemonte Orientale, Vercelli), **Jacques Moeschler** (Unversité de Genève), **Ira Noveck** (CNRS - Institut des Sciences Cognitives, Lyon), **Andrei Popescu-Belis** (ISSCO - Université de Genève), **Kim Sterelny** (Victoria University of Wellington)

The Workshop is organized by the Institut des Sciences Cognitives CNRS, Lyon. Starting from February 2004, a new paper will be put on line and open to discussion every two weeks. The research presented in this workshop is supported in the framework of the European Science Foundation EUROCORES programme.

- Evolution of Language from Theory of Mind or Coevolution of Language and Theory of Mind?
- Evolution du langage depuis la théorie de l'esprit ou coévolution du langage et de la théorie de l'esprit
Anne Reboul (CNRS - Institut des Sciences Cognitives Lyon)

- Constructions underlying theory of mind and language
Peter Ford F. Dominey (CNRS - Institut des Sciences Cognitives, Lyon)

-Cognitive and Functional Factors in the Evolution of Grammar
Frederick J. Newmeyer (Washington University - Institut des Sciences Cognitives, Lyon)

- Do sex differences in empathy account for sex differences in language acquisition?
Simon Baron-Cohen (Cambridge University)

- A Pragmatic Perspective on the Evolution of Langage and Languages.
- Qu'est-ce que la pragmatique peut apporter à l'étude de l'évolution du langage ?
Gloria Origgi (CNRS, Institut Jean-Nicod) and
Dan Sperber (CNRS, Institut Jean-Nicod)

- Imitation, Quoting and Theory of Mind
Tecumseh Fitch *(*University of St. Andrews)

- Why language first?
Jill De Villiers (Smith College)

- Theory of mind and language ability. Understanding the bigger picture
Ted Ruffman (University of Sussex)

- Word learning without Theory of Mind. Possible, but useless
Gil Diesendruck (Bar-Ilan University, Israel)

- On concepts and language
Véronique Boulenger (Institut des Sciences Cognitives, Lyon) and
Tatjana Nazir (Institut des Sciences Cognitives, Lyon)

- The Mirror System Hypothesis. Linking Language to Theory of Mind
Michael Arbib (University of Southern California)

- Putting all the strands together. Coevolution of language and theory of mind
Peter Ford F. Dominey (CNRS - Institut des Sciences Cognitives, Lyon) and
Anne Reboul (CNRS - Institut des Sciences Cognitives, Lyon)

Evolution of Language from Theory of Mind or Coevolution of Language and Theory of Mind?

Anne Reboul (CNRS - Institut des Sciences Cognitives Lyon)

(Date of publication: 13 February 2004)

Abstract: Though some contemporary theories of language evolution see it as taking as its foundation the pre-evolution of theory of mind, acquisition data seem to show that the ability to pass the false belief test does not appear much before four years of age. What is more, some recent studies claim that this ability is crucially linked to the acquisition of some linguistic structures and/or concepts. This seems to suggest that theory of mind might not be monolithic and that a full theory of mind needs language to develop. Thus a coevolution scenario might be preferable to a sequential one in which theory of mind evolution would preceed language evolution.

Introduction

For centuries, philosophers and scientists have been trying to answer the question of what is specific to humankind, as opposed to the rest of nature, i.e. the other animals. Among the many abilities which have been proposed, from bipedalism to laughter, through tools fabrication and usage, two stand out as those for which no animal can clearly and non controversially be said to possess them: language and theory of mind (i.e., the ability of interpreting and predicting others' behavior by the attribution of mental states, such as beliefs, desires, feelings, etc.). Thus, despite the by now fairly numerous attempts to teach them language, no great ape, even the deservedly renowned Kanzi, can be said to have mastered a language comparable in complexity (both qualitatively and quantitatively) to human languages. On the qualitative side, it still remains very doubtful whether apes can learn syntax and on the quantitative side, apes' vocabularies remain extremely limited. The same may be said for theory of mind (henceforth ToM): despite the undisputed fact that great apes (mostly chimpanzees of both species) live in sophisticated hierarchical societies, with a fair amount of social fluidity, it has not been conclusively proved that they have anything like the kind of ToM with which (normal) human beings are endowed. In fact, there seems to be fair evidence that they may, according to Sterelny's distinction (2000) between behavior-readers and mindreaders, belong to the former rather than to the latter category, while human beings clearly are mindreaders.

On the other hand, the evolutionary status of language remains unclear: is it an adaptation or some kind of exaptation (i.e., a by-product of other abilities which were adaptations in their own rights)? In the first case, has it evolved from animal communication or is it sui generis? In the second case, what was it exapted from? And what were those adaptations from which language emerged and how exactly did they allow it to emerge? The very fact, outlined above, that both language and ToM seem to be species-specific abilities, combined with the fact that ToM seems to have some role in any sophisticated communicative behavior, and with the idea that language evolved from animal communicative systems, have led some researchers to the idea that ToM was a prerequisite for language evolution, or, alternatively, that it was one of the adaptations from which language was exapted. (It is presumably important to note that the fact that language might have first appeared as an exaptation does not preclude it from having become an adaptation, i.e. from being selected for and biologically inscribed in the genotype through environmental pressure. This is indeed is a plausible scenario for bipedalism, see e.g. Berge & Gasc 2001, Picq 2003.) If the link between communication and ToM and the hypothesis that language evolved from animal communication are accepted then, language, whether it evolved or was exapted, is seen primarily as a communicative system, i.e. a system whose main function is communication. Note that this does not mean that language cannot have other functions: only the function it was adapted for (or that made it a useful exaptation) was communication. In fact, this agrees pretty well with the Social Cognition hypothesis, proposed by Humphrey (1976), and according to which social cognition evolved under the pressure of group size among social animals and other cognitive abilities are derived from it.

The idea that language and ToM are strongly linked does not come as a surprise to the pragmatician, especially in the Gricean line: it is one of the central tenet of contemporary pragmatics that what is being linguistically communicated is semantically underdetermined and that the decoding of the sentence must be accompanied by inferential processes yielding the complete interpretation of the utterance (see, e.g., Sperber & Wilson 1995, Levinson 2000, Reboul & Moeschler 1998). A good candidate for these inferential processes may be on the line of ToM. Indeed, it is hard to imagine that linguistic communication could take place if our species could not mindread. In two recent papers, Origgi & Sperber (2000, in press) have linked this peculiarity of linguistic communication to the whole problem of language evolution. Basically they point out that it was not language as such that was adapted, but rather the ability for language acquisition. They outline the role of ToM in solving the paradox that this raises, i.e. what is the use of language acquisition if there is no language to be acquired? This paradox, which is untractable on a codic view of linguistic

communication, dissolves on an inferential or mixed view, as inference steps in. Basically, what this means is that language could not have appeared without some sort of mindreading ability. Indeed, this is a fairly frequent hypothesis in the literature of language evolution (see the papers in Givon & Malle 2002 and, for a review, Reboul 2003a).
I will not be directly concerned here with the precise process of language evolution (but see the papers in Christiansen & Kirby 2003) but with the more delicate problem of whether the evolution of ToM did indeed precede the evolution of language, i.e. in the more precise chronological aspects of the problem, as well as in what sort of ToM, if any, did indeed precede language.

Successive evolutions or coevolution of language and ToM?

To sum up, the evolution of language could presumably not have taken place without a workable ToM and this suggests a scenario in which the evolution of ToM preceded and conditioned the evolution of language. As was pointed previously by Malle (2002), things are however not as simple as they seem: for instance, our nearest relatives in the great apes family, i.e. the two species of chimpanzees, do not seem to have anything like a human ToM (this is not to suggest that we might have inherited our ToM from chimpanzees, but that both they and we might have inherited it from our common ancestor of about 7 million years ago); what is more, the main test for ToM, the false belief test is not passed by normal children much before 4 years of age. In addition, a more recent test, the opaque context test (see Kamawar & Olson 2000, Robinson & Apperly 2003) is not passed before 5 years. However, language acquisition begins at aproximatively nine months, much before children can pass false belief, as shown in the following table (built from the data in Baron-Cohen 1995 and Bloom 2000):

Age	**Language acquisition**	**ToM acquisition**
From birth to 9 months		ID and EDD
9 months to 18 months	Going from 6 words to 40	SAM
24 months	311 words	Development of TOM
30 months	574 words	Development of TOM
48 months	Development of vocabulary	False belief test
60 months	Development of vocabulary	Opaque context test

(Where ID is the detector of intentionality, EDD is the detector of eye direction and SAM is the shared attention mechanism). The problem is that, just as a workable ToM is supposed to allow language evolution, a workable ToM is supposed to be necessary for language acquisition (Bloom 2000). The above table seems to suggest that either these hypotheses are wrong or that a more complex view, involving coevolution and coacquisition, rather than a succession of ToM and language, is needed. Indeed, there is independent experimental evidence that there is a link between passing the false belief test and linguistic abilities, whether these abilities are semantic, syntactic or both (see e.g. Yun Chin & Bernard-Opitz 2000, de Villiers & Pyers 2002, Ruffman et al. 2003). I will not try to adjudicate between syntactic and semantic hypotheses, as I think that it is probable that both syntactic and semantic factors play a role. I will however adopt the conclusion that whatever sort of ToM allows for language evolution (and acquisition), it is unlikely to be a full-blown ToM of the sort that allows passing the false belief and opaque context tests. What this means is not that ToM does not play a role in evolution/acquisition of language, but that the question of the nature of the ToM that facilitates both processes has to be carefully examined.

What kind of ToM allows both language evolution and acquisition?

In a recent and extremely fascinating book on the design of animal communication (see Hauser & Konishi 1999, and Reboul 2003b, for a review), Perrett devotes a chapter to "A cellular basis for reading minds from faces and actions", centering on primate abilities. Basically, he shows that there are three types of cells in the temporal cortex of macaques: cells that encode the visual appearance of body and face both static and mobile; cells that encode specific bodily and facial movements; cells that respond to particular faces and body movements as goal-directed actions. There are, as well, prefontal cells which code both the motor component and the visual appearance of specific movements, a sort of general concept of the action. This leads Perrett to a two-steps scenario in which temporal cells recognize the movement as intentional (in the vernacular sense) and alert prefrontal cells which identify the action. Additionally, a fourth type of cells in the temporal area seem to distinguish between self-generated and other generated dimensions of visual stimuli. According to Perrett, this sophisticated visual system may be sufficient for a complex social life without needing an additional mindreading capacity.
Finally, the temporal population of neurons seems to have analogues in the human brain. What would the system just described amount to in ToM terms? Well, it should cover at least one module associated with ToM, ID, and maybe also EDD. That is, an individual with such a neuronal system should be able to detect

intentionality in others' behavior and, possibly, to detect eye direction. Whether it covers SAM as well is unclear, as, on Baron-Cohen's account (Idem), SAM yields ternary representations of type [Mummy-see(I-see-girl)], that is embedded representations, and there is no reason to think that such a neuronal system would in and of itself allow embedding. What is clear, however, is that, at the beginning of language acquisition, though ID and EDD are acquired, SAM is not entirely operational. This suggests, that, despite the title of Perrett's paper, what is in question here might be more on the behavior-reading side than on the mindreading side. The question then is: if ID and EDD are necessary for language acquisition though SAM is not (at least at the beginning of acquisition), and given that there is some evidence of ID and EDD in primates, would not ID and EDD have been sufficient (in terms of behavior/mindreading) for language evolution too? Note that this is not a question of ontogeny recapitulating philogeny, but rather a question of necessary conditions for the evolution of a new ability, the ability to acquire a language. Note as well that I'm not claiming that there is nothing more to language acquisition/evolution than a behavior-reading ability. What I am suggesting rather is that the kind of social ability necessary for the evolution of the language acquisition device may have been of the order of behavior-reading rather than mindreading abilities. Thus SAM, which can be taken to be the first step in mindreading, would not be necessary to the acquisition (or evolution) of language at least at the beginning. The next question is: is a linguistic ability of some kind (however immature) necessary for the development of SAM? In other words, could a non-linguistic animal develop SAM? Well, chimpanzees have been shown to be capable of gaze following (see, e.g., Povinelli 2000) just as babies do. The question then is whether this indicates that chimpanzees and babies have SAM or whether it merely indicates that they have ID and EDD. The only common behavior that indicates that both chimpanzees and babies have SAM is that they are able to follow the gaze of another individual. It is not clear why this should entail a ternary representation of the kind described by Baron-Cohen. In other words, though ternary representations may appear on the basis of gaze following abilities, it is not clear that such abilities are based on ternary representations.
If this is right, then it appears that language evolution/acquisition, though it unsurprisingy rests on social abilities, may need behavior-reading rather than mindreading abilities. A full mindreading ability such as ToM would not be needed. This would enable us to restate the question of the relation between acquisition/evolution of language and acquisition/evolution of ToM.

Coevolution of language and ToM

As the table above makes clear, language and ToM are acquired in tandem, at least if one relies on false belief and opaque context tests as the acid tests for possession of ToM. This reliance has been contested by Bloom & German (2000), who pointed out that the false belief test tests plenty of other things in addition to ToM, such as e.g. memory. This criticism is probably right though it does not mean that the false belief task does not test ToM as well: this is because possession of ToM relies on the ability to attribute to other people different beliefs from one's own. There also is the methodological point that until now no better test seems to have been proposed. However, the issue regarding the false belief and opaque context tests might rather be that, as they depend on language, they obviously cannot be passed until mastery of language is well advanced. Thus, such an objection would go, it is no surprise that passing those tests is done late and they do not give us any information about the mindreading abilities of children previous to their having attained a linguistic ability sufficient for the task. This objection may well be right, but it is not clear that it goes through given the evidence that piles in favor of the existence of a strong link between mindreading and language. Supposing that there is indeed such a link, what more precisely could it be?
In a series of experiments, Povinelli (2000) has tried to demonstrate that chimpanzees, despite their social and tool using abilities, do not have ToM or naive physics comparable to human ToM or naive physics (Note however that these experiments have been criticized by, e.g., Hauser 2001). The difference is due to the fact that both abilities are supported in humans by abstract concepts corresponding to invisible but supposedly causally efficient "entities" such as force, belief, etc. Chimpanzees not only do not have these concepts, according to Povinelli: no amount of learning leads them to acquire them. Povinelli's theory regarding this major difference is that it can be explained by the presence in humans and absence in chimpanzees of language. In other words, language is, under that hypothesis, what allows human to develop abstract concepts of the sort described above. Note that though abstract concepts in ToM might depend on an embedding capacity of the sort described in Hauser et al. (2002), there is no reason to think that this is the case for abstract concepts in naive physics. Thus, it may be that we do indeed need language for mindreading, though we do not need mindreading for language at its inception, even though we need behavior reading abilities. This means that language may have a cognitive function, though I will not discuss here whether this was the basis for language evolution (see Newmeyer 2003 for a nice discussion).

Conclusion: what is a ToM useful for?

If we do not need a ToM to acquire language and did not need it to evolve language, what is the use of a full-blown ToM? A first (Gricean) suggestion might be that we need it for utterance interpretation. This position

has been criticized by Sperber & Wilson (2002) who propose that utterance interpretation rests on a dedicated comprehension module, which evolved especially for the comprehension of linguistic communication, exploiting some metarepresentational principles, as well at the relevance least cognitive effort principle. This would seem to rob ToM of any significant role in utterance interpretation. However, I would like to go back to a tripartite distinction introduced by Sperber (1994) between three interpretive strategies: Naive optimism, in which the hearer considers the speaker to be both competent and benevolent; Cautious optimism, in which the hearer considers the speaker to be benevolent but not necessarily competent; and Sophisticated understanding, in which the speaker is assumed to be neither competent nor benevolent. In the first strategy, no ToM is required, though some is necessary in the second strategy and a lot more in the third strategy. It is a common place to say that communication (especially linguistic communication) brings with it the possibility of deception. This may be where the coevolution scenario takes a new dimension: the evolution of a language acquisition device allowed the emergence of language and linguistic communication that respectively allowed the development of a full-blown ToM and of deception. And ToM may be the best tool for detecting and counter foiling deception, though I would not go so far as to claim any sort of Baldwin effect such as that described in Godfrey-Smith (forthcoming), in which a new situation (here linguistic communication) leads to the adoption of new behavioral skills (here ToM from language), leading to a change in social ecology (better mindreading individuals are favored), hence to changes in selection pressures, which might lead to the new skills being adapted, i.e. to an increase of genotypes predisposing individuals to acquire these skills (here full-blown ToM).

A final word: Dehaene (1997) has shown how sophisticated contemporary mathematics arises from and is still determined by numerosity, a relatively low level capacity which we share with a wide range of animal species, from birds to primates. Numerosity needs to be supplemented by language and a symbolic system of notation to be the foundation of mathematics. My suggestion is that full-blown mindreading may rest on relatively simple behavior-reading abilities supplemented by language to develop. It then may be used in sophisticated linguistic communication (such as fiction for instance) though it probably is not necessary for either language acquisition or evolution and for most common-place linguistic communication.

References

Baron-Cohen, S. (1995) *Mindblindness. An essay on autism and theory of mind*, Cambridge, MA, The MIT Press.

Berge, C. & Gasc, J-P. (2001) "Quand la bipédie devient humaine", in Picq, P. & Coppens, Y. (eds), Aux origines de l'humanité. Le propre de l'homme, Paris, Fayard, 80-125.

Bloom, P. (2000) *How Children Learn the Meanings of Words*, Cambridge, MA, The MIT Press.

Bloom, P. & German, T. (2000) "Two reasons to abandon the false belief task as a test of theory mind", Cognition 77, B25-B31.

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

De Villiers, J.G. and Pyers, J.E. (2002), "Complements to cognition: a longitudinal study of the relationship between complex syntax and false-belief understanding", Cognitive Development 17, 1037-1060.

Dehaene, S. (1997) *The number sense*, Oxford, Oxford University Press.

Givon, T. & Malle, B. (eds) (2002) *The evolution of language out of pre-language*, Amsterdam, John Benjamins Pub. Co.

Godfrey-Smith, P. (forthcoming) "Between Baldwin skepticism and Baldwin boosterism", in Weber, B. & Depew, D. (eds), Learning and evolution. The Baldwin effect reconsidered, Cambridge, MA, The MIT Press.

Hauser, M.D. (2001). "Elementary, my dear chimpanzee. Review of "Folk physics for apes" (D. Povinelli)", Science 291, 440-441.

Hauser, M., Chomsky, N. & Fitch, T. (2002) "The faculty of language. What is it, who has it, and how did it evolve?", Science 298, 1569-1579.

Hauser, M. & Konishi, M. (eds) (1999) The design of animal communication, Cambridge, MA, The MIT Press.

Humphrey, N. (1976) "The social function of intellect", in Bateson, P. & Hinde, R. (eds), *Growing points in ecology*, Cambridge, Cambridge University Press, 303-317.

Kamavar, D. and Olson, D.R. (2000) "Children's Representational Theory of Language: Problems of Opaque Contexts", Cognitive Development 14, pp. 531-548.

Levinson, S. (2000) *Presumptive meaning. The Theory of Generalized Conversational Implicature*, Cambridge, MA, The MIT Press.

Malle, B. (2002), "The relation between language and theory of mind in development and evolution", in Givon, T. & Malle, B. (eds), *The evolution of language out of pre-language*, Amsterdam, John Benjamins Pub. Co., 265-284.

Newmeyer, F. (2003) "Grammar is grammar and usage is usage", Language 79/4, 682-707.

Origgi, G. & Sperber, D. (2000) "Evolution, communication and the proper function of language: a discussion of Millikan in the light of pragmatics and of the psychology of mindreading", in Carruthers, P. & Chamberlain, A. (eds), *Evolution and the human mind. Language, modularity and social cognition*, Cambridge, Cambridge

University Press, 140-169.
Perrett, D. (1999) "A cellular basis for reading minds from faces and actions", in Hauser, M. & Konishi, M. (eds) *The design of animal communication*, Cambridge, MA, The MIT Press.
Picq, P. (2003) *Au commencement était l'homme. De Toumaï à Cro-Magnon*, Paris, Odile Jacob.
Povinelli, D. (2000) *Folk Physics for Apes*, Oxford, Oxford University Press.
Reboul, A. (2003a) "Review: Givon, Talmy and Bertram F. Malle, eds (2002) *The Evolution of Language out of Pre-language*. John Benjamins Publishing Company, paperback ISBN 1-58811-238-1, ix+392", Linguist list issue 14.1734.
Reboul, A. (2003b) "Review: Hauser, Marc D. and Mark Konishi, ed. (2003) *The Design of Animal Communication*, MIT Press.", Linguist list issue 14.3088.
Reboul, A. & Moeschler, J. (1998) *La pragmatique aujourd'hui. Vers une nouvelle science de la communication*, Paris, Le Seuil.
Robinson, E.J. and Apperly, I.A. (2001) "Children's difficulties with partial representations in ambiguous messages and referentially opaque contexts", Cognitive Development 16, pp. 595-615.
Ruffman, T., Slade, L., Rowlandson, K., Rumsey, C. and Garnham, A. (2003) "How language relates to belief, desire and emotion understanding", Cognitive Development 18, pp. 139-158.
Sperber, D. (1994) "Understanding verbal understanding", in Khalfa, J. (ed.), *What is intelligence?*, Cambridge, Cambridge University Press, pp. 179-198.
Sperber, D. & Origgi, G. (in press) "Qu'est-ce que la pragmatique peut apporter à l'étude de l'évolution du langage?", in Hombert, J-M. (ed.) *L'origine de l'homme, du langage et des langues*.
Sperber, D. & Wilson, D. (1995) *Relevance. Communication and cognition*, Oxford, Basil Blackwell, 2nd edition.
Sperber, D. and Wilson, D. (2002) "Pragmatics, modularity and mind reading", Mind and Language 17, pp. 3-23.
Sterelny, K. (2000) "Primate worlds", in Heyes, C. & Huber, L. (eds), *The evolution of cognition*, Cambridge, MA, The MIT Press, 143-162.
Yun Chin, H. and Bernard-Opitz, V. (2000) "Teaching Conversational Skills to Children with Autism: Effect on the Development of a Theory of Mind", Journal of Autism and Developmental Disorders 30/6, pp. 569-583.

French version:
traduction de l'original en anglais de Anne Reboul

Evolution du langage depuis la théorie de l'esprit ou coévolution du langage et de la théorie de l'esprit

Résumé:
Bien que certaines théories contemporaines de l'évolution du langage la conçoivent comme prenant pour fondation l'évolution préalable de la théorie de l'esprit, les données de l'acquisition semblent montrer que la capacité à passer le test de la fausse croyance ne se développe pas avant quatre ans. Qui plus est, des études récentes affirment que cette capacité est liée de façon cruciale à l'acquisition de certaines structures linguistiques et/ou de certains concepts. Ceci suggère que la théorie de l'esprit pourrait ne pas être monolithique et qu'une théorie de l'esprit complète a besoin du langage pour se développer. Ainsi un scénario coévolutionniste pourrait être préférable à un scénario séquentiel où l'évolution de la théorie de l'esprit précède celle du langage

Introduction

Pendant des siècles, les philosophes et les savants essayent de répondre à la question de ce qui est le propre de l'homme, par opposition au reste de la nature, c'est-à-dire aux autres animaux. Parmi les nombreuses capacités qui ont été proposées, de la bipédie au rire, en passant par la fabrication et par l'usage des outils, il n'en reste que deux dont on peut dire de façon claire et indiscutable qu'aucun animal ne les possède : le langage et la théorie de l'esprit (c'est-à-dire la capacité à interpréter et à prédire le comportement d'autrui par l'attribution d'états mentaux, comme la croyance, le désir, les sentiments, etc.). Ainsi, malgré les nombreuses tentatives pour apprendre aux primates à parler, aucun d'entre eux, même le célèbre Kanzi, n'a obtenu la maîtrise d'un langage comparable du point de vue de la complexité (à la fois quantitativement et qualitativement) aux langues humaines. Du point de vue qualitatif, il paraît extrêmement douteux que les primates puissent apprendre la syntaxe et, du point de vue quantitatif, leur vocabulaire reste très limité. La même chose peut être dite de la théorie de l'esprit (ci-après ToM) : malgré le fait indiscutable que les primates (principalement les deux espèces de chimpanzés) vivent dans des sociétés hiérarchisées sophistiquées, avec une dose non négligeable de fluidité sociale, on n'est pas arrivé à montrer de façon indiscutable qu'ils aient le type de ToM dont jouissent les êtres humains (normaux). De fait, il semble y avoir de bonnes raisons de penser qu'ils pourraient appartenir, selon la distinction de Sterelny (2000) entre

lecteurs de comportement et lecteurs d'esprit, au premier groupe plutôt qu'au second, alors que les êtres humains sont clairement des lecteurs d'esprit.

D'un autre côté, le statut du langage quant à son évolution reste peu clair: s'agit-il d'une adaptation ou d'une quelconque exaptation (c'est-à-dire d'un sous-produit d'autres capacités qui seraient, quant à elles, le fruit d'une adaptation) ? Dans le premier cas, le langage a-t-il évolué à partir de la communication animal ou est-il sui generis ? Dans le second cas, à partir de quoi a-t-il été exapté ? Et quelles seraient les adaptations à partir desquelles le langage a émergé et comment, exactement, lui ont-elles permis d'émerger ? Le fait même que le langage et la ToM semblent spécifiques à l'espèce humaine, combiné au fait que la ToM semble avoir un rôle dans n'importe quel comportement communicatif sophistiqué, et à l'idée que le langage a évolué à partir des systèmes de communication animaux, ont conduit certains chercheurs à l'hypothèse que la ToM était un pré requis pour l'évolution du langage, ou, alternativement, que c'était une des adaptations à partir desquelles le langage a été exapté. (Il faut noter que le fait que le langage pourrait d'abord avoir été une exaptation ne lui interdit pas d'être devenu, par une pression environnementale, une adaptation, par exemple en étant sélectionné puis inscrit dans le génotype. C'est par exemple un scénario plausible pour la bipédie, cf. Berge & Gasc 2001, Picq 2003). Si le lien entre la communication et la ToM ainsi que l'hypothèse selon laquelle le langage a évolué à partir de la communication animale sont admis, alors le langage, qu'il ait évolué ou qu'il ait été exapté, est vu principalement comme un système de communication, c'est-à-dire un système dont la fonction principale est la communication. De fait, ceci s'accorde bien avec l'hypothèse de la Cognition sociale, proposée par Humphrey (1976), et selon laquelle la cognition sociale a évolué sous la pression de la taille des groupes chez les animaux sociaux et que les autres capacités cognitives en seraient dérivées.

L'idée selon laquelle il y a un lien entre le langage et la ToM n'est pas une surprise pour le pragmaticien, particulièrement lorsqu'il est d'obédience gricéenne : une des hypothèses centrales de la pragmatique contemporaine est que la communication linguistique est sémantiquement sous déterminée et que le décodage de la phrase doit s'accompagner de processus inférentiels livrant l'interprétation complète de l'énoncé (cf., entre autres, Sperber & Wilson 1995, Levinson 2000, Reboul & Moeschler 1998). Ces processus inférentiels pourraient être basés sur la ToM. En effet, il est difficile de penser que la communication linguistique pourrait exister sans que notre espèce puisse lire l'esprit. Dans deux articles récents, Origgi & Sperber (2000, à paraître) ont lié cette particularité de la communication linguistique au problème de l'évolution du langage. Ils font remarquer que ce n'est pas tant le langage lui-même qui a évolué, mais la capacité à l'acquisition du langage. Ils indiquent le rôle de la ToM dans la solution du paradoxe que ceci soulève : à quoi sert l'acquisition du langage s'il n'y a pas de langage que l'on puisse acquérir ? Ce paradoxe, qui est insoluble si l'on adopte une vision codique de la communication linguistique, se dissout si l'on adopte une vision inférentielle ou mixte, parce que l'inférence intervient. A la base, ce que ceci signifie est que le langage n'aurait pas pu apparaître sans une capacité quelconque à lire l'esprit. De fait, c'est une hypothèse très fréquente dans la littérature sur l'évolution du langage (cf. Givon & Malle 2002 et, pour un compte-rendu, Reboul 2003a).

Je ne m'intéresserai pas ici au processus précis de l'évolution du langage (cf. Christiansen & Kirby 2003), mais au problème plus délicat de savoir si l'évolution de la ToM a effectivement précédé l'évolution du langage, c'est-à-dire aux aspects chronologiques du problème, tout autant qu'au type de ToM qui a dû précéder le langage.

Une succession d'évolutions ou une co-évolution du langage et de la ToM ?

En bref, l'évolution du langage n'aurait probablement pas pu se faire sans une ToM opérationnelle et ceci suggère un scénario dans lequel l'évolution de la ToM a précédé et a été une condition de l'évolution du langage. Comme l'a indiqué précédemment Malle (2002), les choses ne sont cependant pas aussi simples : par exemple, nos plus proches parents dans la famille des grands primates, les deux espèces de chimpanzés, ne semblent pas avoir de ToM comparable à la ToM humaine (ceci ne veut pas suggérer que nous aurions pu hériter notre ToM des chimpanzés, mais que les chimpanzés et les êtres humains auraient pu l'hériter de leur ancêtre commun d'il y a environ sept millions d'années) ; qui plus est, le principal test de la ToM, le test de la fausse croyance, n'est pas réussi par les enfants avant quatre ans. Par ailleurs, un test plus récent, le test des contextes opaques (cf. Kamawar & Olson 2000, Robinson & Apperly 2003), n'est pas réussi avant cinq ans. Cependant, l'acquisition du langage commence approximativement vers 9 mois, très longtemps avant que les enfants puissent passer le test de la fausse croyance, comme le montre le tableau suivant (construit à partir des données de Baron-Cohen 1995 et Bloom 2000):

Age	Language acquisition	ToM acquisition
From birth to 9 months		ID and EDD
9 months to 18 months	Going from 6 words to 40	SAM
24 months	311 words	Development of TOM
30 months	574 words	Development of TOM
48 months	Development of vocabulary	False belief test
60 months	Development of vocabulary	Opaque context test

(Où ID est le détecteur d'intentionnalité, EDD est le détecteur de direction du regard et SAM est le mécanisme d'attention partagée). Le problème est que, exactement comme une ToM opérationnelle est supposée permettre l'évolution du langage, une ToM opérationnelle est supposée nécessaire pour l'acquisition du langage (Bloom 2000). Le tableau ci-dessus suggère que soit ces hypothèse sont fausses soit une vision plus complexe, impliquant la co-évolution ou la co-acquisition plutôt qu'une succession entre la ToM et le langage, est nécessaire. De fait, il y a des données expérimentales qui indiquent un lien entre succès au test de la fausse croyance et capacités linguistiques, que ce lien soit de nature syntaxique, sémantique ou les deux à la fois (cf. entre autres Yun Chin & Bernard-Opitz 2000, de Villiers & Pyers 2002, Ruffman et al. 2003). Je ne vais pas essayer de choisir entre les hypothèses syntaxique et sémantique parce que je pense que ces deux facteurs jouent un rôle. Je vais cependant adopter la conclusion selon laquelle quelle que soit la ToM qui permet l'évolution du langage (et son acquisition), ce n'est probablement pas la ToM complète qui permet de passer les tests de la fausse croyance et des contextes complexes. Ce que ceci signifie, ce n'est pas que la ToM ne joue pas un rôle dans l'évolution/acquisition du langage, mais que la question de la nature de la ToM qui facilite ces deux processus doit être soigneusement examinée.

Quelle sorte de ToM permet à la fois l'évolution et l'acquisition du langage ?

Dans un livre récent et tout à fait fascinant sur la communication animale (Hauser & Konishi 1999 et Reboul 2003b pour un compte-rendu), Perrett consacre un chapitre à la base cellulaire de la lecture de l'esprit à partir des visages et des actions. Il montre qu'il y a trois types de cellules dans le cortex temporal des macaques : des cellules qui encodent l'apparence visuelle du corps et du visage, statiques ou dynamiques ; des cellules qui encodent des mouvements corporels et spatiaux spécifiques ; des cellules qui répondent à des mouvements corporels et spatiaux particuliers interprétés comme des actions dirigées vers un but. Il y a aussi des cellules pré-frontales qui encodent à la fois le composant moteur et l'apparence visuelle de mouvements spécifiques, quelque chose comme un concept général d'action. Ceci conduit Perrett à un scénario en deux étapes dans lequel les cellules temporales identifient le mouvement comme intentionnel (dans le sens non technique) et alertent les cellules pré-frontales qui identifient l'action. De plus, un quatrième type de cellules dans l'aire temporale semblent distinguer les actions initiées par soi et d'autres dimensions générées des stimuli visuels. Selon Perrett, ce système visuel sophistiqué pourrait être suffisant pour une vie sociale complexe, sans qu'une capacité de lecture d'esprit additionnelle ne soit nécessaire. Enfin, cette population temporale de neurones semble avoir des analogues dans le cerveau humain.

A quoi le système qui vient d'être décrit se ramenerait-il en termes de ToM ? Il devrait couvrir au moins un des modules associé à la ToM, ID, et peut-être aussi EDD. En d'autres termes, un individu pourvu d'un tel système neuronal devrait être capable de détecter l'intentionnalité dans le comportement d'autrui et, peut-être, de détecter la direction du regard. Il n'est pas clair qu'il couvre aussi SAM, étant donné que, d'après Baron-Cohen (Idem), SAM livre des représentations ternaires du type [Maman-voit (je-vois-fille)], c'est-à-dire des représentations enchâssées, et il n'y a pas de raison de penser que ce système neuronal puisse, par lui-même, permettre l'enchâssement. Ce qui est clair, cependant, c'est que, au début de l'acquisition du langage, bien que ID et EDD soient en place, SAM n'est pas entièrement opérationnel. Ceci suggère que, malgré le titre de l'article de Perrett, ce qui est en question ici pourrait être davantage du côté de la lecture du comportement que de celui de la lecture de l'esprit. La question est alors : si ID et EDD sont nécessaires à l'acquisition du langage bien que SAM ne le soit pas (au moins au tout début de l'acquisition), et étant donné qu'il y a des indications qu'ID et EDD existent chez les primates, pourquoi ID et EDD ne suffiraient-ils pas aussi (du point de la lecture du comportement ou de l'esprit) à l'évolution du langage ? Remarquons que ce n'est pas la question de savoir si l'ontogénie récapitule la philogénie, mais plutôt une question de conditions nécessaires à l'évolution d'une nouvelle capacité, la capacité d'acquérir un langage. Remarquons aussi que je ne prétends pas qu'il n'y a pas plus dans l'évolution/acquisition du langage qu'une capacité à lire le comportement. Ce que je suggère est plutôt que la sorte de capacité sociale nécessaire pour l'évolution du module d'acquisition du langage pourrait avoir été de l'ordre de la lecture du comportement plutôt que de celle de l'esprit. Ainsi SAM, que l'on peut considérer comme la première étape dans la lecture de l'esprit, ne serait pas nécessaire à l'acquisition (ou à l'évolution) du langage au moins à ses débuts. La question suivante est alors : une capacité linguistique d'une sorte quelconque (quelque immature qu'elle puisse être) est-elle nécessaire au développement de SAM ? Autrement dit, est-ce qu'un animal non-

linguistique pourrait développer SAM ?

On a montré que les chimpanzés sont capables de suivre le regard (cf. Povinelli 2000) exactement comme le sont les bébés. On peut donc s'interroger pour savoir si ceci indique que les chimpanzés et les bébés ont SAM ou si cela indique simplement qu'ils ont ID et EDD. Le seul comportement commun qui indiquent que les chimpanzés et les bébés ont SAM est qu'ils sont capables de suivre le regard. La raison pour laquelle ceci impliquerait une représentation ternaire du type décrit par Baron-Cohen n'est pas claire. En d'autres termes, bien que des représentations ternaires puissent apparaître sur la base d'une capacité à suivre le regard, il n'est pas clair que cette capacité soit basée sur des représentations ternaires.

Si ceci est correct, il semble que l'acquisition/évolution du langage, bien qu'elle s'appuie de façon peu surprenante sur des capacités sociales, pourrait avoir besoin d'une capacité à lire le comportement plutôt que d'une capacité à lire l'esprit. Une capacité à lire l'esprit complète, comme la ToM, ne serait donc pas nécessaire. Ceci nous permettrait de poser différemment la question de la relation entre l'évolution/acquisition du langage et l'évolution/acquisition de la ToM.

La co-évolution du langage et de la ToM

Comme le montre le tableau ci-dessus, le langage et la ToM sont acquis en tandem, au moins si l'on s'appuie sur les tests de la fausse croyance et des contextes opaques comme pierre de touche de la possession d'une ToM. Cette confiance a été critiquée par Bloom & German (2000), qui ont remarqué que le test de la fausse croyance teste beaucoup d'autres choses que la ToM, comme, par exemple, la mémoire. Cette critique est probablement juste, mais elle ne signifie pas que le test de la fausse croyance ne teste pas la ToM : en effet, la possession d'une ToM s'appuie sur la capacité à attribuer à autrui des croyances différentes de celles que l'on a soi-même. On remarquera aussi la question méthodologique qui est qu'il n'y a pas de test plus approprié à l'heure actuelle. Cependant, la question en ce qui concerne les tests de la fausse croyance et des contextes opaques pourrait plutôt être que, comme ils dépendent du langage, ils ne peuvent bien évidemment pas être réussis avant que la maîtrise du langage ne soit bien avancée. Ainsi, une telle objection pourrait être formulée en disant qu'il n'y a rien de surprenant dans le fait que la réussite à ces tests soit tardive et qu'ils ne nous donnent pas d'information sur les capacités à lire l'esprit des enfants avant qu'ils n'atteignent une maîtrise linguistique suffisante pour passer le test. Cette objection est peut-être correcte, mais les indications de plus en plus nombreuses qui s'accumulent en faveur de l'existence d'un lien fort entre lecture de l'esprit et langage lui retirent beaucoup de son poids. Si l'on suppose qu'il y a un tel lien, en quoi consisterait-il précisément ?

Dans une série d'expériences, Povinelli (2000) a essayé de montrer que les chimpanzés, malgré leurs capacités sociales et leur performance d'utilisateurs d'outils, n'ont pas une ToM ou une physique naïve comparable à celle des humains (cependant ces expériences ont été critiquées, notamment par Hauser 2001). La différence serait due au fait que ces deux capacités sont centrées chez les humains autour de concepts abstraits correspondant à des "entités" invisibles mais supposées causalement efficaces comme la force, la croyance, etc. Selon Povinelli, non seulement les chimpanzés ne possèdent pas ces concepts, mais aucun entraînement ne leur permet de les acquérir. La théorie de Povinelli en ce qui concerne cette différence majeure est qu'elle s'explique par la présence chez les humains et l'absence chez les chimpanzés du langage. Autrement dit, dans cette hypothèse, le langage est ce qui permet aux êtres humains de développer des concepts abstraits du type que je viens de décrire. Remarquons que, bien que les concepts abstraits dans la ToM puissent dépendre d'une capacité à l'enchâssement du type de celle que décrivent Hauser et al. (2002), il n'y a pas de raison de penser que ce soit le cas pour les concepts abstraits de la physique naïve. Ainsi, il se pourrait que nous ayons effectivement besoin du langage pour lire l'esprit, bien que nous n'ayons pas besoin de lire l'esprit pour le langage à ses débuts, même si nous avons besoin de capacités à lire le comportement. Ceci signifie que le langage pourrait avoir une fonction cognitive, bien que je ne souhaite pas discuter ici de la possibilité que ce soit la base de son évolution (pour une proposition intéressante dans cette voie, cf. Newmeyer 2003).

Conclusion: à quoi sert la ToM ?

Si nous n'avons pas besoin de la ToM pour acquérir le langage et si nous n'en avons pas eu besoin pour l'évolution du langage, à quoi sert une ToM complète ? Une première suggestion (gricéenne) consisterait à dire que nous en avons besoin pour l'interprétation des énoncés. Cette position a été critiquée par Sperber et Wilson (2002) qui avancent l'hypothèse que l'interprétation des énoncés s'appuie sur un module dédié à la compréhension, module qui aurait évolué spécifiquement pour la compréhension linguistique et qui exploiterait des principes métareprésentationnels ainsi que le principe de pertinence dont on sait qu'il est basé sur un principe d'économie cognitive. Ceci semble priver la ToM de tout rôle significatif dans l'interprétation des énoncés. Je voudrais cependant en revenir à une distinction introduite par Sperber (1994) entre trois stratégies interprétatives : l'optimisme naïf, dans lequel l'interlocuteur considère que le locuteur est tout à la fois compétent et bienveillant ; l'optimisme prudent, dans lequel l'interlocuteur considère que le locuteur est bienveillant, mais pas nécessairement compétent ; la compréhension sophistiquée, dans

laquelle le locuteur n'est considéré ni comme compétent, ni comme bienveillant. Dans la première stratégie, il n'y a pas besoin de ToM, alors qu'un peu de ToM est nécessaire dans la seconde et qu'une ToM complète est requise dans la troisième. Il est commun de dire que la communication (particulièrement la communication linguistique) amène la possibilité de la tromperie. C'est peut-être ici que le scénario de la co-évolution prend une nouvelle dimension : l'évolution d'un mécanisme d'acquisition du langage permet l'émergence du langage et de la communication linguistique qui permettent respectivement le développement d'une ToM complète et de la tromperie. Et la ToM serait le meilleur outil pour détecter et contrer la tromperie, bien que je n'irai pas jusqu'à avancer un quelconque effet Baldwin comme celui que décrit Godfrey-Smith (à paraître), dans lequel une nouvelle situation (ici la communication linguistique) conduit à l'adoption de nouveaux comportements (ici la ToM depuis le langage), conduisant à un changement dans l'écologie sociale (les meilleurs lecteurs d'esprit sont favorisés), et donc à des changements dans les pressions sélectives, qui pourraient conduire à l'évolution de ces nouveaux comportements, c'est-à-dire à une augmentation de génomes prédisposant les individus à acquérir ces comportements (ici une ToM complète).

Un mot pour finir: Dehaene (1997) a montré comment les mathématiques sophistiquées actuelles sont issues et sont encore déterminées par la numérosité, une capacité de relativement bas niveau que nous partageons avec un grand nombre d'espèces animales, depuis les oiseaux jusqu'aux primates. La numérosité doit être complétée par du langage et par un système de notation symbolique pour être le fondement des mathématiques. Ma suggestion est qu'une capacité à lire l'esprit peut s'appuyer pour se développer sur des capacités de lecture du comportement relativement simples, augmentées par le langage. Elle peut ensuite être utilisée dans la communication linguistique sophistiquée (comme la fiction par exemple) bien qu'elle ne soit probablement pas nécessaire pour l'acquisition ou l'évolution du langage et pour la majeure partie de la communication linguistique quotidienne.

References

Baron-Cohen, S. (1995) *Mindblindness. An essay on autism and theory of mind*, Cambridge, MA, The MIT Press.

Berge, C. & Gasc, J-P. (2001) "Quand la bipédie devient humaine", in Picq, P. & Coppens, Y. (eds), Aux origines de l'humanité. Le propre de l'homme, Paris, Fayard, 80-125.

Bloom, P. (2000) *How Children Learn the Meanings of Words*, Cambridge, MA, The MIT Press.

Bloom, P. & German, T. (2000) "Two reasons to abandon the false belief task as a test of theory mind", Cognition 77, B25-B31.

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

De Villiers, J.G. and Pyers, J.E. (2002), "Complements to cognition: a longitudinal study of the relationship between complex syntax and false-belief understanding", Cognitive Development 17, 1037-1060.

Dehaene, S. (1997) *The number sense*, Oxford, Oxford University Press.

Givon, T. & Malle, B. (eds) (2002) *The evolution of language out of pre-language*, Amsterdam, John Benjamins Pub. Co.

Godfrey-Smith, P. (forthcoming) "Between Baldwin skepticism and Baldwin boosterism", in Weber, B. & Depew, D. (eds), Learning and evolution. The Baldwin effect reconsidered, Cambridge, MA, The MIT Press.

Hauser, M.D. (2001). "Elementary, my dear chimpanzee. Review of "Folk physics for apes" (D. Povinelli)", Science 291, 440-441.

Hauser, M., Chomsky, N. & Fitch, T. (2002) "The faculty of language. What is it, who has it, and how did it evolve?", Science 298, 1569-1579.

Hauser, M. & Konishi, M. (eds) (1999) The design of animal communication, Cambridge, MA, The MIT Press.

Humphrey, N. (1976) "The social function of intellect", in Bateson, P. & Hinde, R. (eds), *Growing points in ecology*, Cambridge, Cambridge University Press, 303-317.

Kamavar, D. and Olson, D.R. (2000) "Children's Representational Theory of Language: Problems of Opaque Contexts", Cognitive Development 14, pp. 531-548.

Levinson, S. (2000) *Presumptive meaning. The Theory of Generalized Conversational Implicature*, Cambridge, MA, The MIT Press.

Malle, B. (2002), "The relation between language and theory of mind in development and evolution", in Givon, T. & Malle, B. (eds), *The evolution of language out of pre-language*, Amsterdam, John Benjamins Pub. Co., 265-284.

Newmeyer, F. (2003) "Grammar is grammar and usage is usage", Language 79/4, 682-707.

Origgi, G. & Sperber, D. (2000) "Evolution, communication and the proper function of language: a discussion of Millikan in the light of pragmatics and of the psychology of mindreading", in Carruthers, P. & Chamberlain, A. (eds), *Evolution and the human mind. Language, modularity and social cognition*, Cambridge, Cambridge University Press, 140-169.

Perrett, D. (1999) "A cellular basis for reading minds from faces and actions", in Hauser, M. & Konishi, M.

(eds) *The design of animal communication*, Cambridge, MA, The MIT Press.
Picq, P. (2003) *Au commencement était l'homme. De Toumaï à Cro-Magnon*, Paris, Odile Jacob.
Povinelli, D. (2000) *Folk Physics for Apes*, Oxford, Oxford University Press.
Reboul, A. (2003a) "Review: Givon, Talmy and Bertram F. Malle, eds (2002) *The Evolution of Language out of Pre-language*. John Benjamins Publishing Company, paperback ISBN 1-58811-238-1, ix+392", Linguist list issue 14.1734.
Reboul, A. (2003b) "Review: Hauser, Marc D. and Mark Konishi, ed. (2003) *The Design of Animal Communication*, MIT Press.", Linguist list issue 14.3088.
Reboul, A. & Moeschler, J. (1998) *La pragmatique aujourd'hui. Vers une nouvelle science de la communication*, Paris, Le Seuil.
Robinson, E.J. and Apperly, I.A. (2001) "Children's difficulties with partial representations in ambiguous messages and referentially opaque contexts", Cognitive Development 16, pp. 595-615.
Ruffman, T., Slade, L., Rowlandson, K., Rumsey, C. and Garnham, A. (2003) "How language relates to belief, desire and emotion understanding", Cognitive Development 18, pp. 139-158.
Sperber, D. (1994) "Understanding verbal understanding", in Khalfa, J. (ed.), *What is intelligence?*, Cambridge, Cambridge University Press, pp. 179-198.
Sperber, D. & Origgi, G. (in press) "Qu'est-ce que la pragmatique peut apporter à l'étude de l'évolution du langage?", in Hombert, J-M. (ed.) *L'origine de l'homme, du langage et des langues*.
Sperber, D. & Wilson, D. (1995) *Relevance. Communication and cognition*, Oxford, Basil Blackwell, 2nd edition.
Sperber, D. and Wilson, D. (2002) "Pragmatics, modularity and mind reading", Mind and Language 17, pp. 3-23.
Sterelny, K. (2000) "Primate worlds", in Heyes, C. & Huber, L. (eds), *The evolution of cognition*, Cambridge, MA, The MIT Press, 143-162.
Yun Chin, H. and Bernard-Opitz, V. (2000) "Teaching Conversational Skills to Children with Autism: Effect on the Development of a Theory of Mind", Journal of Autism and Developmental Disorders 30/6, pp. 569-583.

Discussion

Joint attention as a scaffold for language
Peter Gärdenfors
Feb 18, 2004 10:40 UT

I find the topic of Anne Reboul's paper extremely important. However, I have some problems with the way she describes a ToM, in particular the function of shared attention, and its role for the evolution of language.

First of all, ToM does not come in a single package. In my recent book *How Homo Became Sapiens*, I distinguish several levels of a theory of mind: having a theory of emotions, having a theory of attention, having a theory of intentions, having a theory of beliefs, and having self-consciousness. Reboul seems to do something similar in her paper since she distinguishes the intentionality detector (ID), the eye direction detector (EDD) and the shared attention mechanism (SAM). Separating ToM into several components means that one cannot only ask one question about the connection between ToM and the evolution of language, but one must separate the roles of the different components. Second, there is an ambiguity in what can be meant by "shared attention". On the one hand, it can mean just "I see what you see" which is a second order attention. This is presumably what Reboul means by eye direction detection. On the other, it can mean what Tomasello and others call joint attention. The latter involves that I see that you are looking at an object and you see that I see the same object, but also that I see that you want me to look at what you look at. Thus joint attention also involves a theory of intention.

The most fundamental form of intentional communication is to manipulate the attention of others. Tomasello (1999, p. 102) writes that in order for me to understand that you want to communicate about an object, I must understand that you intend us to jointly attend to the object. To achieve this requires a second order intention and a second order attention. But not even this is necessary to reach intentional communication. Ingar Brinck (2001) argues that if I aim at manipulating your attention it is sufficient that I intend that you and I jointly attend to something. This only requires a first order intention and a second order attention, since it is not necessary that you understand that I intend to communicate. Such a form of communication can be achieved, for example, by pointing and joint gaze contact – something which children manage from an age of about one year and which presumably also apes are capable of.

Even though intentional communication can be achieved without joint attention, the mechanism is very important in language learning. First of all, joint attention is triadic in that the two attendees and the attended object form a "referential triangle" (see picture on p. 104 in Tomasello (1999)). But, most importantly, joint attention scaffolds language learning. If an adult points to a frog and the child sees what the adult is pointing to and reaches joint attention by shifting its gaze between the adult and the frog, the child will connect the word "frog" with the perception of the object. Normally, one or two instances are sufficient for a child to learn this connection. And once it has coupled the word with the object, the child can use the word itself to direct the attention of somebody else. Such a role reversal is possible once you master joint attention. Tomasello's claim is that pointing is an essential component in the language learning mechanism and unless the child masters it, the rest of language learning will not work either. I am not implying that language is learnt by pointing and joint attention only – there are many other mechanisms (Bloom 2000) – but it is a mechanism that plays an important role in language acquisition and presumably played a similar role in the evolution of the early stages of language. The available evidence indicates that apes and other animals never reach joint attention. They do not understand that a human or another ape has intention towards their attentional state. This may a crucial factor in why they have difficulties in learning language, let alone develop a language of their own. In brief, the relation between ToM and language evolution (or language acquisition for that matter) is an entangled matter. However, the mechanism of joint attention, which is a crucial stage in the development of a ToM, is necessary for developing a language that goes beyond merely signaling. I submit that at least for this component of ToM, it must precede language, evolutionarily as well as developmentally.
References

Bloom, P. (2000): How Children Learn the Meanings of Words, MIT Press, Cambridge, MA.

Brinck, I. (2001): "Attention and the evolution of intentional communication." Pragmatics and Cognition 9(2), pp. 255-272.

Gärdenfors, P. (2003): How Home Became Sapiens: The Evolution of Thinking, Oxford University Press, Oxford. Tomasello, M. (1999): The Cultural Origins of Human Cognition, Harvard Unversity Press, Cambridge, MA.

Reply to Peter Gärdenfors
Anne Reboul
Feb 18, 2004 13:28 UT

I am thankful to Gärdenfors' throughtful comment which will allow me to clarify some of the points I tried to make in my paper. First of all, I certainly do not think that ToM comes in a single package and I fully endorse Baron-Cohen's distinction between different modules which make different contributions to mindreading, i.e. the intentionality detector (ID), the eye direction detector (EDD), the shared attention mechanism (SAM) and theory of mind (ToM). I tend to think however that this is not the whole story to mindreading and that other mechanisms, such as, for instance, the detection of emotion, certainly play a role in mindreading. In my paper, I made the mistake of using ToM indiscriminately to refer to mindreading in general rather than to what Baron-Cohen calls ToM, which in my paper was called "full-fledged ToM". Let me begin again by calling full-fledged ToM simply ToM (as Baron-Cohen does) and calling the full set of abilities mindreading, by acknowledging that ID and EDD jointly amount to second order attention, and by calling what Gärdenfors calls joint attention SAM. I agree with most of what Peter Gärdenfors says, including what he says, following Tomasello, regarding intentional communication. Where we may partly disagree is on the last two lines of his comment where he claims that SAM "must precede language, evolutionarily as well as developmentally". ID and EDD jointly enable an individual to read the behavior of others. They certainly do not enable her to mindread others (i.e. behavior reading is necessary but not sufficient for mindreading). SAM is necessary for that (on this point, we fully agree, I think). What is more, I agree that SAM "is necessary for developing a language that goes beyond mere signaling". What my hypothesis amounts to is simply that language may have begun as signaling and that it is the development of intentional communication (of which Gärdenfors admit that it can take place without SAM) through whatever form of proto-language one cares to propose that allowed the development of SAM and SAM (jointly with the evolution of the language acquisition device) then allowed the development of language as more than mere signaling. In other words, ID and EDD precede intentional communication through proto-language, and then there is a coevolution of language and SAM and language is instrumental in ToM. As regards language acquisition, children acquire their first words at around nine months, before they have mastered SAM, which may indicate that their first use of language could be mere signaling. Their progressive mastery of linguistic communication as intentional leads them to SAM, and eventually to TOM. On that point, one should perhaps note that apes, who learn those reduced forms of linguistic communication which I alluded to in my paper, do not develop SAM (or obviously language) which means that their "linguistic" communication is presumably more on the side of mere signaling than on anything else, just as that of children at the first stage of language acquisition. Thus we may agree more than Peter Gärdenfors thinks…

Do we still need a False Belief Task?
Gloria Origgi
Feb 18, 2004 12:30 UT

In her provocative text, Anne Reboul claims that there's no need of a full-fledged *Theory of Mind* to acquire language. Neither this capacity is required for utterance interpretation, for which a more fine grained capacity of recognizing speaker's meaning is required.

She suggests at the end of her paper that a full-blown Theory of Mind could matter for sophisticated interpretation of social interactions.

But if this is the picture, then what kind of mindreading ability are people testing through the false belief task? Which capacity are psychologists measuring in this task? What makes a false belief task still a mindreading task, given also the evidence that the non-mental versions of the task provide the same results (see the results for the "false photograph task" in Leslie [2000] and Slaughter [1998])?

Wouldn't it be possible to claim just that with the acquisition of language our abstract thoughts about things and people become more sophisticated and we become able to deal with counterfactual situations and embedded reasoning?

To sum up, should we abandon False Belief Task to understand the mindreading capacities the underlie language acquisition and discourse comprehension? I understand that Anne suggests we should, and I agree with her. But I am interested in knowing what other people think of this option.

References

A. Leslie (2000) “How to acquire a “representational theory of mind” in D. Sperber (ed.) *Metarepresentations*, Oxford University Press.

Slaughter, V. (1998) “Children’s understanding of pictorial mental representations” *Child Development*, 69, pp. 321-332.

Reply to Gloria Origgi
Anne Reboul
Feb 18, 2004 14:01 UT

It's difficult to answer a comment on which one is in global agreement. Indeed, I do think that the false belief task does not test our ability to acquire or evolve language, because, even if it does test mindreading abilities, it does not test the correct ones. Discussing what it is exactly that the false belief tests is beyond the topic of this reply (and may indeed go beyond the topic of the whole conference). What seems more interesting is to try to delineate more precisely how linguistic and non-linguistic abilities interact, and, in this specific case, how non-linguistic mindreading abilities interact with either linguistic abilities or abilities acquired through language. Here, again, I globally agree with Gloria on the idea that there is a strong (and unsurprising) link between language and abstraction (where abstraction concerns invisible entities in the sense alluded to in my paper) and between language and such relatively abstract cognitive faculties as counterfactual and embedded reasoning (for instance, make-believe games do not develop much before 18 months, when language is well under way). Perhaps a final additional remark: Bloom (2000, 68; see ref. in my paper) explains lexical contrast (the fact that children, when presented with a new word, will systematically consider it to apply to an as yet nameless object rather than to an object for which they already have a name) through a 6 steps mindreading process, clearly involving ToM and counterfactual reasoning (cf. Reply to Gärdenfors). Interestingly, most (if not all) of the experiments that Bloom discusses involve three-year-olds, i.e. children who certainly have SAM and who may be expected to have acquired quite a lot of ToM. If evidence of a similar phenomenon could be found in children between 9 and 18 months it would be much more difficult to explain it through ToM and counterfactual reasoning…

What does it take to learn a word
Gil Diesendruck
Feb 18, 2004 16:08 UT

Reboul does a great service to our understanding of the relation between language and "theory of mind" (ToM) by decomposing the full-fledged ToM into specific components (Baron-Cohen's). The components might be conceptually modular, and are even more reasonably developmentally separable. I don't want to commit to this particular decomposition, and I agree with Gardenfors that perhaps other analyses are welcome.

What I want to point out, however, is that just as we need an analyis of ToM, we also need an analysis of what it means to learn a language, or even simply acquire a word. Reboul presents a table summarizing milestones in the development of mindreading and language acquisition. Noticeably, however, while the milestones in mindreading are qualitatively different, the ones on language development are just quantitatively so, having to do with the number of words in children's vocabulary. But just as the capacity to mindread might be composed of separate skills, so might the capacity to acquire words.

Here's just a sketch of what these steps might look like: 1- There are intentional acts directed at objects; 2- Some sounds uttered by people are examples of such intentional acts; 3- One function of intentional acts is to refer to things; 4- Words are the best instances of such intentional-referential acts; 5- Words are common knowledge.

Again, this is just a sketch to make the point that when discussing the possible relation between ID, EDD, SAM, or joint attention to word learning, we need to specify what level of word learning capacity we are trying to explain.

On a final note, Reboul refers to the studies by Diesendruck & Markson (2001) (cited by Bloom, 2000) showing that 3-year-olds seem to engage in some sort of mindreading when disambiguating the meaning of a word. Citing Bloom, Reboul asks whether indeed children need to go through what appears like counterfactual reasoning, to avoid lexical overlap. I believe that is not necessary. Children could reason "positively" that "I expect speakers to use familiar words (e.g., "banana") if they want to refer to familiar things (a banana)." In these studies, children hear an unfamiliar word from which they infer that the speaker is probably not intending to refer to the familiar object.

Markson and I have data that 2-year-olds respond the same way 3-year-olds do, and Markman, Wasow, & Hansen (2003) have recently reported that even 15-month-olds to some extent avoid lexical overlap. But pushing the age even younger may not be required, once we consider my earlier comments. That is, this task may tap onto a fairly sophisticated understanding of words, and thus may not be the kind of task to assess the earliest capacities to acquire words. We need a systematic analyses of both full-fledged mindreading AND word learning in order to look for the relationship between them.

References

Bloom, P. (2000). How children learn the meanings of words. Cambridge, MA: MIT Press.

Diesendruck, G., & Markson, B. (2001). Children's avoidance of lexical overlap: A pragmatic account. Developmental Psychology, 37, 630-641.

Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. Cognitive Psychology, 47, 241-275.

Reply to Gil Diesendruck
Anne Reboul
Feb 19, 2004 14:15 UT

Basically, I'm in agreement with Gil Diesendruck on her comment. I agree on the discrepancy between the qualitative aspect of milestones in mindreading as opposed to the quantitative aspects in language development, though that quantitative aspect might be more obvious in the acquisition of the lexicon than in that of grammar, e.g. syntactic structures (which incidentally makes studies linking ToM and syntax very relevant and interesting). Coming back to lexical acquisition, I think that of the five steps described by Gil, the fifth might be the more interesting as it taps directly in the notion of convention. Words are (subject to obvious morpho-phonological constraints) by and large arbitrary, which means that they are shared through the establishment of a convention linking all speakers in a linguistic community (note that this does not exclude idiosyncrasic diversity). This is presumably true of word learning in the acquisition of the lexicon. However, it is still not entirely clear how such conventions get established. I take it that clearly the establishment of such lexical conventions should include the first four steps described by Diesendruck. The fifth and last one should be given a more precise description, however, for two reasons: first the notion of common knowledge may (in some versions) imply infinite regress (and hence is not very plausible from a psychological point of view); what is more, it is presumably the decisive step in the establishment of the lexical convention considered and might involve more sophisticated mindreading than the previous ones. All in all, I agree with Diesendruck that 'we need a systematic analyse of both full-fledged mindreading AND word learning in order to look for the relationship between them.

Bricolage in the Evolution of Language and Mind-Reading
Michael Arbib
Feb 21, 2004 18:52 UT

1. I agree with Anne Reboul's conclusion that "full-blown mind reading may rest on relatively simple behavior-reading abilities supplemented by language to develop … [and] probably is not necessary for either language acquisition or evolution ..." I agree, too, with Gärdenfors that we must treat ToM as a complex of

structures rather than a single one, and with Diesenbruck that language acquisition must also be subdivided, but think that the emphasis on word acquisition may be mistaken and that “fully fledged” ToM (ffToM) is impossible without more complex utterances. Here are some specific comments on different aspect of Reboul’s discussion:

2. Reboul asserts that “it is hard to imagine that linguistic communication could take place if our species could not mind read”. I would agree that linguistic communication would seem to require some form of shared attention (cf. Gärdenfors) but do not see why it would require mind reading rather than behavior reading to be of great social benefit. Indeed, Reboul quotes Perrett’s [1999] analysis of "A cellular basis for reading minds from faces and actions" which may perhaps be more akin to behavior reading than mind reading, and yet be sufficient for a complex social life without needing an additional mind reading capacity. I will return to this topic in my own paper discussing the Mirror System Hypothesis later in this series.

3. I would like to see a more critical discussion of ID as “the detector of intentionality”. I think there is compelling evidence to distinguish what is called imitation in the newborn from what I would consider “true” imitation in the one-year old infant. Perhaps a similar distinction between “the early appearance of something with some of the characteristics of intentionality” and “later appearance of an understanding of the intentions of others” might be helpful in our multi-aspect approach to ToM.

4. I would like to see a thread discussing why mind reading is useful, as distinct from behavior reading. If I see that you are angry and behave in a more guarded fashion because of this, one might explain it in “mind talk” but in this case analysis of behavioral disposition in relation to facial expression seems more than adequate. Why then do mental states figure in folk psychology as more than dispositions for current behavior? It seems that the ability to follow quasi-causal chains to link a variety of behaviors by an individual seem crucial, and it may require both good episodic memory and the ability to use language to organize it to ground a folk psychology that makes these mental states feel real. This seems to support Reboul’s position.

5. Reboul notes Sperber & Wilson’s (2002) notion of a dedicated comprehension module, which evolved especially for the comprehension of linguistic communication, but this seems too all-or-none. Language is regenerative - not simply expressing extant meanings but also creating new realities as experience crystallizes around new words and expressions and stories. However, Sperber’s (1994) distinction between three interpretive strategies that she cites does seem helpful. Whereas the speaker of early language (in either sense) might simply be uttering a description or command, effective use of modern language will often (not always) have a communicative goal of affecting the speaker in a certain way. With experience, one builds increasingly subtle models of individuals to tailor one’s utterance not only to one’s general experience with them but also to one’s sense of their current state. This is true for behavior, but becomes amplified in subtlety because language can communicate so many more messages which can thus go wrong in so many ways. Basic animal behavior in terms of hierarchy seem to be the basis for this - the protoToM, if you like, once coupled to, e.g., the facial expression of the emotions or other communicator (e.g., pheromones) of current state.

6. Let me just add that much of the “evolutionary process” described in Reboul’s article and here may well be a historical process building on a biology adequate to support protolanguage, rather than resting on specific biological changes.

Reply to Michael Arbib
Anne Reboul
Feb 23, 2004 13:16 UT

As Michael Arbib had the foresight to give numbers to different parts of his comment, I'll answer each of these parts separately, referring to them by the same numbers. 1. I too agree that ToM is a complex of structures: indeed I should have thought that this was a central part of my paper. So, we're all of us, Michael, Gil, Peter and I, in agreement on this point. 2. Linguistic communication would not (at all) need mindreading if it entirely rested on an encoding-decoding mechanism. This, by the way, seems to be how quite a lot (if not all) of animal communication systems work. Linguistic communication is not like that because what a given speaker wants to communicate may go farther than what the sentence (s)he uses means from a purely linguistic point of view. This was why Grice came up with two notions, the notion of implicature and the notion of conversational maxims. Both rested implicitly on mindreading notions. This, however, as Sperber and Wilson (1995) pointed out was not necessary: a cognitive economy principle, coupled with a simple inference system, can do the job of recovering the speaker's intention. What I claim, however, is that, though this is quite enough for most of linguistic communication, it may not be enough for some rare — but not all that rare — situations in which one

or both of the communicators attempt to manipulate the other. Note that detecting lies is not necessarily the point: a robust sense of contradiction (i.e. elementary logic) is enough for that. ToM however is necessary to discover the specific intent of the lier and, frequently, to deflect the effects of the lie. It's also necessary for that ubiquitous human activity, i.e. the production and consumption of fiction, from Henry James and Proust to cheap romantic novelettes. A short remark: Peter Gärdenfors advocates the version of shared attention which is mindreading in an at least elementary form. On this point, Arbib and Gärdenfors disagree and I probably agree with both: i.e. I agree with Arbib that mindreading is not necessary in the first stages of language acquisition, but I also agree with Gärdenfors that joint attention (or SAM in Baron-Cohen's terms) is probably necessary for further stages of language acquisition and presumably was necessary to make language more than a system of signs (indeed to lift linguistic communication to its present levels). 3. There I fully and without reservation agree with Arbib. 4. I agree with Arbib on the necessity of a discussion on why ToM is necessary. My own view has just been expressed above: it is necessary to survive the widespread possibility of deception that language brings. 5. There, I think, there is a misunderstanding. Sperber and Wilson's comprehension module is quite strongly not a mindreading module. It is, as said in point 1, an inference mechanism coupled with a cognitive economy principle, not involving mindreading abilities. However, I would agree with Arbib on two points: I'm still not entirely convinced by the idea that comprehension of linguistic communication is done through a dedicated module, rather than built on more general abilities — note that on Sperber and Wilson's view (1995), the relevance principle comes in two versions, one general, the other dedicated to comprehension of linguistic communication: this suggests to me that, taking into account the specific character of communication as attention grabing, the general principle might be enough and the comprehension principle might be dispensed with. If this is the case, then there would be no evolution of a specific comprehension module. On the rest of point 5, I agree with Arbib, though see my reply to Cristina Meini for more details. 6. Michael may be right on that one, but I tend to think that some recent genetic evidence (see the website of the SLI consortium for details and reference) does seem to give some (admittedly still fragile) evidence for a biological change specific to language (though most certainly not to ToM). Thus, there is again no major disagreement here, I think (aren't we all nice people!). Reference Sperber, D. and Wilson, D. (1995) Relevance: communication and cognition, Oxford, Basil Blackwell (2nd edition).

UNDERSTANDING ATTENTION: IS IT A PSYCHOLOGICAL MATTER?

Cristina Meini

Feb 22, 2004 17:51 UT

I find Anne Reboul's paper extremely interesting and I am inclined to agree on her conclusion that proto-language development could require no folk psychological capacities. At the same time, Gärdenfors' reply has to be seriously taken into account. According to him, joint attention is crucial in language learning because it is a precondition for role reversal, which is a crucial step in language development . Since he adopts Tomasello's (1999) notion of joint attention, it follows that folk psychological capacities (more precisely, a psychological theory of attention) are necessary for early language development.

Consider the following well known experiment. A child and an adult each take an object in their hands. Both objects are unknown to the child. When the adult utters a word unknown to the child, we would expect her to associate this new noun with the object in her hand. On the contrary, what she does is first to monitor the adult's gaze direction and then associate the noun with the object the experimenter is looking at (Baldwin, 1991; 1993). (Notice that these data concern not exactly proto-language talkers, but nineteen month-old children. At this age, the acquisition of lexicon has started long since; moreover, independent data suggest that SAM and ToMM are already in place [see e.g. Leslie, 1987]) . A child who can learn the meaning of a new word in Baldwin's situation will soon be able to reverse the roles and use pointing to name objects. Moreover, pointing is soon performed with repeated gaze-shifting. Contrary to Tomasello (1999), I am not persuaded that these performances (and, in particular, role reversal) involve metapsychological capacities from the start. Nobody doubts that a full grasp of the attentive mechanisms requires an understanding of their psychological properties. Nevertheless, early lexicon acquisition and early pointing-and-naming could need just a geometrical notion of attention, for which both the direction of gaze and the direction of the adult's self-generated movement are relevant per se. That is, the child need not to know that there is a mind beyond the eyes. She is just (not intentionally) mimicking the adult. In this developmental framework, it is by pointing with a geometrical conception of attention that the child can come to understand the selective nature of attention, which is a genuinely psychological property. JA (in Tomasello's sense) would develop later, so that attention turns out to be a psychological affair.

References

Baldwin D. (1991), “Infant’s contribution to the achievement of joint reference”, Child Development, 63, 875-890.

Baldwin D. (1993), “Infant’s ability to consult the speaker for clues to word reference”, Developmental Psychology, 29, 832-843.

Leslie A. (1987) “Pretense and representation: The origins of ‘theory of mind’”, Psychological Review, 1, 94, 412-426.

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

Reply to Cristina Meini
Anne Reboul
Feb 23, 2004 10:31 UT

My answer to Cristina Meini will be brief as, once again, I don't think that there is any obvious disagreement between us. First of all, I think (and I believe that there is more or less general agreement on that point) that whatever modules mindreading abilities may consist in (and as pointed before in the present debate, they include behavior reading abilities, though they do not reduce to these behavior reading abilities), these modules are psychological in Meini's sense (that is, if I understood her rightly, non linguistic). Language acquisition (or at least lexical acquisition) begins with behavior reading abilities such as, for instance, ID (intentionality detector) and EDD (eye direction detector) — though defensors of other behavior reading modules should feel free to replace them with their own choice. What this means is that mindreading is not necessary for language acquisition in this very first stage. This however does NOT mean that mindreading (e.g. SAM) is not necessary at later stages or that it was not necessary in language evolution to go beyond protolanguage. Meini's reply raises a further question: if role reversal is, as Gärdenfors claims, necessary for language acquisition/evolution, and if it rests on joint attention capacities and if joint attention capacities are mindreading capacities, then language acquisition/evolution needs mind reading abilities. Thus, in the end, the main question in Meini's reply is: does role reversal require metapsychological capacities? This is a difficult question for a very simple reason: the answer will depend on the theoretical description that one gives of the cognitive (or functional) mechanisms which underlie role reversal. This, by the way, is one of the major problem in any description of any behavior as involving or not mindreading: depending on one's theory, one will say that this or that behavior (e.g., the ability to recognize oneself in a mirror) indicates or does not indicate mindreading ability. A final point: the central question is whether one can have mindreading abilities without language. If one can, then there is no reason to think that young children cannot have mindereading abilities at a very early age (e.g. 18 months). If one can't, then neither young children nor chimpanzees have mindreading abilities, though young children will acquire them with language. Right now, let's face it, there is no obvious end to the discussion… A final point: if mindreading abilities, or at least ToM (in the false belief test sense) need language to develop, this does not make them linguistic abilities, anymore — to return to the comparison at the end of my initial paper — than mathematics is a linguistic capacity. ToM still would be a psychological (in Meini's sense) ability, but an ability which could only develop with language.

Language facilitates constructing more powerful modules such as TOM
Eric Baum
Feb 22, 2004 20:52 UT

My just-released book *What is Thought?* (MIT Press, 2004, see http://whatisthought.com for more information) illuminates these issues within the context of its overall model of mind. Space limitations will unfortunately render the following sketch both imprecise and somewhat telegraphic.

Based on a variety of arguments, *What is Thought?* describes a model in which the mind is execution of a computer program composed of semantically meaningful modules. (What meaning is is discussed at some length but elided here.) These modules call one another, and lower level modules comprising a large fraction of the program of animal mind are essentialy coded into the genome. Words are seen as labels for such meaningful computational modules, thus facilitating communication of programs. Words are thus easy to learn, since the concepts are already programmed in, and common, and often salient by virtue of their meaningful nature, and and one must merely learn to attach the correct label. Creatures build additional

modules onto the program during life, building on top of the genetically coded structure. The problem of discovering meaningful and useful new code is computationally hard, i.e. requires much computation. Humans have built vastly more additional structure into the program by virtue of their ability to make cumulative progress, based on computations by billions of individual minds that can pass partial results through language, on the problem of new meaningful code construction.

A module such as TOM is too complex to spring fully formed, but rather builds on previous modules coded into the genome. For example, plovers are known to be able to estimate intent. A module for estimating intent is a necessary submodule for a more complex TOM. While creatures like plovers and apes do not have a full human TOM, they have necessary submodules. The work of Perrett cited by Reboul, of which I was not previously aware, seems to fit nicely into this picture of genetically predisposed submodules. Humans acquire a more complex TOM by building new concepts on top of these submodules, and thus refining the computer code-- the program (developed over human history) being passed on through such instructional devices as bedtime stories and fiction. This is thus consistent with the table of language and TOM abilities by age presented by Reboul.

A difficulty with this theory is the question of why language took so long to evolve, if the concepts existed and the words are simply attached to them. The answer proposed is that evolution was caught in a local optimum where a full language was not fit. The proposal of Martin Nowak and colleagues that this local optimum arose through reasons arising from information theory is discussed, as well as a (different, but possibly complimentary) proposal based on evolutionary computing considerations that the key step was getting started in expecting to hear language-- before this one couldn't create a word with the expectation it would be learned and thus useful. This appears related to the discussion in Reboul's paper of Origgi & Sperber(2000, in press) in a slightly different context.

If these arguments are correct, the main mental differences, including TOM, between human and apes are likely due to additional programming constructed on top of that in the genome and passed through language, and thus are primarily due to nurture. However, such developments can feed back into the genome in various forms of coevolution such as the Baldwin effect. *What is Thought?* discusses in particular how the Baldwin effect may have facilitated the development and genetic coding of human gramatical competence. However, one does not need to posit a linguistic use for TOM to understand why it would be discovered. As Reboul notes, there are various other reasons why developing a strong TOM renders one fitter. And as TOM is a semantically meaningful module, and as its program is built on submodules present before humans, it is quite natural that it would be discovered once language is available to facilitate passing the code.

Eric Baum http://whatisthought.com

Reply to Eric Baum
Anne Reboul
Feb 23, 2004 10:59 UT

Eric Baum's comment is welcome as it enables me to say a few things which I would have liked to say in my paper but didn't for reasons of space. Let me begin with what I agree on in Baum's comment: I agree on the idea that animals have concepts (I'm not among the philosophers such as, e.g., Davidson or Gauker, who think that there are no concepts without language). I also agree on the fact that animals are behavior readers of varying skills. The first point of disagreement may be that I don't think that all concepts can appear independently of language. There, my position would agree with Povinelli's: I think that his suggestion that animals, including chimpanzees, cannot have abstract concepts, in the sense of concepts for invisible things (e.g. force, belief, knowledge, etc. — note that this does not deny them an ability to generalize —) might well be right. In other words, though most if not all animals have concepts, there are some concepts that nonhuman animals will not have, because their existence depends on language. Among such concepts are, I think, those concepts which are taken to be central to ToM, i.e. epistemic concepts among which the more prominent is belief. This explains quite enough why, though a lot of concepts existed before language, there is a necessary connection between language and ToM: without language, the concepts which are basic for ToM cannot appear. On the other hand, given the opportunity for lies and deception which linguistic communication offers, ToM may be a necessary addition to linguistic abilities. Here, I would like to refer to Carruthers's (2002) BBS paper on "the cognitive functions of language". In this paper, Carruthers claims that concept forming as a cognitive function of language is trivial and uninteresting and that the cognitive function of language is to be a medium of communication between different cognitive modules. I will not pronounce here on his positive thesis, but I would like to insist that concept forming as a cognitive function of language if one thinks that all concepts are attainable

without language. If this is not so, then concept forming can be a highly nontrivial cognitive function of language. A final word: it has been claimed that ToM is an innate ability because of poverty of input. I would like to point out that the fact that false belief test is not successfully passed much before four years does suggest an alternative account. During the period when young children spend their time making their parents' life miserable by asking continuous "why" questions, quite a few of these questions are devoted to why people do this or that. And parents answer these questions using quite explicitly ToM explanations. This suggests that ToM may be a late development because it is learned through explicit linguistic communication bearing on the explanation of other people's behavior.

References: Carruthers, P. (2002) "The cognitive functions of language", Behavioral and Brain Sciences 25, 657-726.

reply to reply: language and representing vs discovering complex concepts
Eric Baum
Feb 23, 2004 15:58 UT

We, of course, agree that people have concepts animals don't (e.g. the concept of quantum field theory). The question then is whether (among other possibilities) (a) language and such concepts are unrelated, (b) language and such concepts are both made possible by some other evolutionary discovery, (c) language itself is necessary for representing these concepts, and/or (d)(as I suggested) such concepts could be represented without language, but weren't discovered-- these concepts are too difficult to discover in one lifetime without being programmed in, or at least being guided to partial results, through language, which thus facilitates discovery over many lifetimes. *What is Thought?* reviews such various possibilities and evidence, including the possibility that grammar is, or was made possible by, a better interface for computational modules. (This last possibility would be intriguing for AI, since it would imply there is an enormously valuable discovery waiting to be made.)

While I don't think there are sufficient grounds for a confident resolution, I think option (d) is both obviously one way language facilitates more complex concepts (we wouldn't have quantum field theory without our ability to pass partial results, thus facilitating cumulative progress) and sufficient to explain the available data (and e.g. specifically TOM). Thus the principle of parsimony suggests we don't need to venture into other explanations.

Moreover, there is some evidence against the notion that language is necessary for representing complex concepts (as opposed to discovering them.) (1) numerous mathematicians and scientists, introspecting, argue that they think non-verbally. (2) There is significant evidence against the oft-posed hypothesis that animals can not consider objects not present, for example: in experiments where dolphins and monkeys are taught "to speak", they seem to have no problem referring to objects not present; dogs seem to dream about chasing rabbits (which are not physically present in their bedroom) (and monkeys trained to press levers with their feet when viewing movies spontaneously press when they dream). As to invisible (and odorless) objects, it seems evident dogs can form a concept of an invisible fence. (3) Some concepts that might be thought abstract are understood by some simple creatures-- for example bees understand the concept of same and different(Giurfa et al,2001) and solve many other complex mental tasks (Gould and Gould, 1994). (4) (Donald 1991) reported the case of Brother John, a monk subject to epileptic seizures that completely suspended all verbal abilities. Yet he was cogent though these seizures and had no difficulty remembering them.

Eric Baum http://whatisthought.com

Reply to Baum's reply
Anne Reboul
Feb 24, 2004 14:54 UT

The problem is not that we agree that animal don't have such concepts as quantum field theory. We do. That's not the question. But, obviously, quite a lot of human beings do not have the concept of quantum field theory either. Admittedly, they could acquire it, though it might mean acquiring quite a

lot of other, background concepts. What I'm saying is not that animals cannot have scientific concepts which, anyway, most human beings do not have either. What I'm saying (following Povinelli 2000) is that animals do not have some concepts that all normal older children and adults manipulate on a daily if not hourly basis and that these concepts are abstract in the sense that they are not and cannot be "perceived". Following again Povinelli's suggestion, I make the hypothesis that such concepts cannot appear without language, precisely because they are abstract in that sense. It is important to note that the ability of "talking" animals to refer to absent objects is not relevant here: these ABSENT objects are absent precisely in the sense that they have been perceived in the past even of they are not presently perceived. Such commonplace abstract concepts as BELIEF are different in that they CAN'T be perceived at all. There is no evidence of animals constructing representations of non-existent things (as human beings very commonly do in fiction or counterfactual thinking). Again, the introspective evidence to the effect that savants and scientists think non-verbally is neither here nor there: for one thing, this is true mainly of mathematicians and physicists, the most notable example being Einstein who claimed that he discovered relativity by imaginatively seeing himself riding a light ray. First of all, regarding mathematicians, Dehaene (see ref. in my original paper) explains quite well such evidence. Regarding physicists, most examples have to do with research on space-time, where at least the spatial dimensions are perceptible or may give rise to mental imagery. As far as I know, regarding BELIEF, the only spatial metaphor is that of the "belief box", and that supposes that the concept of BELIEF is already known. "Same" and "different" are presumably necessary for generalisation, i.e. for concept formation for concrete objects and I certainly would not want to deny that animals have concepts for concrete objects. However, SAME and DIFFERENT presumably can be attributed on the basis of perception. Finally, the epileptic monk is a nice example, but there are examples of aphasic people performing well at cognitive tasks (among others, the false belief test). There are a few explanations for that (namely that aphasia might impair linguistic performance rather than competence). Another one is quite simply that though language might be necessary to acquire some concepts, it may not be necessary to use them once they are built. In other words, as Arbib (point 5) argued in his comment, "language is regenerative — not simply expressing extant meanings but creating new realities", in this specific instance, epistemic concepts. Thus, I would like to add a fifth possibility to Raum's enumeration (let's call it e): e) language may be needed, in the case of concepts which are abstract in the above sense, not only to represent the concept but to build it.

Not Only Language and Not Only for Language
Cristiano Castelfranchi
Feb 24, 2004 14:18 UT

I do basically agree with Reboul's thesis that ToM is not a pre-requisite for human language. However I think that all this issue (and the related argumentation) is somewhat invalidated or distorted. We should first of all discuss the possible forerunners of linguistic verbal communication in human species, and their characters. The main issue is - before that of the evolution of language - that of the evolution of intentional communication, being it behavioral, gestural or vocal. If we do not identify the cognitive requirements of primitive forms of intentional communication before language we cannot identify which level of social cognition and which limited and preliminary forms of ToM are necessary before language and can be part of its prerequisites for evolution. We can have a good model of co-evolution when we will have a more analytic idea of partial and limited proto-forms of both: ToM and intentional communication. For example, compared with other forms of intentional communication - that seem to presuppose some model about what the other can see, recognize, react to, believe, expect, want, etc. - linguistic communication has important additional features that seems to add cognitive complexity and requisites. Let me give some examples. Fully developed linguistic communication necessarily is meta-communication (Grice): you have to understand that I intend - by my act - that you understand that I intend... Not all forms of intentional communication (when I have the goal that you understand something from my act, for example that I'm on rush, or that I like your dinner, or that I will pass on your left) are meta-communicative, i.e. I also want that you understand that I'm communicating with you. In deception (simulation, bluff) my intention to communicate should even be hidden. Analogously, linguistic communication has the special requirement of not being based on 'natural' signs but on conventional and artificial ones. This might require additional cognitive capabilities, also in ToM terms: is it necessary for example tacit meaning and rule negotiation? Surely conventions (that are definitely very relevant in human life independently from language) put specific requirements on social cognition, in terms of conditional reciprocal expectations about what the other expects, intends, understands (Castelfranchi et al. 2003), and perhaps in terms of mutual beliefs (a very complex recursive ToM representation). Moreover, linguistic communication basically is 'about' world objects and events, not just about the behaviors and related internal states of the agent. This 'symbolic' capability - which already

appears in 'symbolic' gestures (rather different - to me - from gestures deriving from and denoting actions) might be cognitively complex: you do not simply recognize and understand my behavior (and I intend so) but I have the idea that you will derive from my sign a mental representation of something else, of an object or a process in the world. These brief examples are just to argue that we need a much more careful anatomy of cognitive prerequisites of communication (in its different forms), of language, and of other forms of social interactions. In the same perspective I agree with Reboul's view of ToM as a non unitary capacity but as a set of cognitive skills of different complexity that can evolve in steps.

Although I agree with the author's thesis, and I find very relevant the decomposition of ToM acquisition in cognitive skills, I have some doubt about how Reboul builds her argument. She claims that "this is not a question of ontogeny recapitulating philogeny" but in fact her reasoning seems to go in this direction. It would be better to explicitly discuss some tacit presuppositions of the argument. It seems clear to me that the development or acquisition of a given capability C1 can precede the acquisition of another capability C2 without the former being a pre-requisite of the latter neither in ontogenetic terms nor in evolutionary terms. It also seems clear to me that if C1 is a developmental prerequisite for C2 it must precede C2 in acquisition. But I'm not so sure and equally persuaded that if C1 has been an evolutionary prerequisite for the philogenetic emergence of C2 the manifestation of C1 in ontogeny must precede the manifestation of C2. This seems on the contrary be presupposed as obvious.

Castelfranchi, C., Giardini, F., Lorini, E., Tummolini, L. (2003) "The Prescriptive Destiny of Predictive Attitudes: from Expectations to Norms via Conventions" in R. Alterman, D. Kirsh (Eds) Proceedings of the 25th Annual Meeting of the Cognitive Science Society, Boston, MA.

Reply to Cristiano Castelfranchi
Anne Reboul
Feb 25, 2004 11:40 UT

There are quite a few interesting points in Castelfranchi's comment. I'll take them in succession. First Castelfranchi says that "we should first of all discuss the possible forerunners of linguistic verbal communication". I agree that this would be desirable, if it were possible. Unfortunately, as archeologists know (see Picq 2003, Picq and Coppens 2001, ref. in my paper), behavior, whether communicative or not, does not keep well. This means that, unless we want to rely on (more or less educated) guesses, which might turn out to be as many 'just-so' stories, we tend to rely on what is available, i.e. animal communication systems and language acquisition. The next point is Castelfranchi's claim that "fully developed linguistic communication necessarily is meta-communication". This has been (successfully I think) contested by Sperber and Wilson (1995, 2002, refs. in my paper) on the basis of vicious regress (psychologically implausible) and by Glüer and Pagin (2003). Some linguistic communication may need meta-communication (as e.g. the sophisticated understanding strategy described by Sperber 1994, ref. in my paper), but not all linguistic communication does. I'm not sure about what the point on linguistic communication being basically 'about' objects and events in the world is supposed to show. Clearly, some events in the world are (parts of) behavior, and I never claimed that language was not about objects and events in the world in any case. The most interesting point made by Castelfranchi is, I think, the one in the last paragraph. So, while I want to reiterate that my argument wasn't a simple "ontogeny recapitulate philogeny" argument, I think that Castelfranchi's point raises questions which are interesting in their own right. Let me begin by pointing out that my argument was not positive (as Castelfranchi seems to think) but negative: if language acquisition begins when neither ToM nor SAM are in place, then neither ToM nor SAM are necessary (in the logical sense) for the beginnings of language acquisition. Note that (as I have frequently pointed out in the present discussion in answers to various comments) this does not mean that neither SAM nor ToM play a role at later stages of language acquisition. What I think is interesting in Castelfranchi's comment is that it agrees with something which seems clear to me (though I didn't mention it in either the paper or previous replies): language acquisition is easier than language evolution, mainly because there is a language to be acquired in the first case, while this is not true in the second case. An accompanying difference (which may or not reduce to the first, depending on whether or not syntax is taken to be the product of evolution in the biological sense) is that presumably learning a convention is easier than establishing a convention. What this means is that SAM and ToM, though not (logically) necessary for language acquisition might have been (logically) necessary for language evolution. However, this leaves unexplained the strong links which have been found between ToM and language mastery…

Reference Glüer, K. and Pagin, P. (2003) "Meaning theory and autistic speakers", Mind and Language 18/1, 23-51.

Reply to Reboul's reply
Cristiano Castelfranchi
Feb 26, 2004 10:32 UT

Let me make my 700-words contribution a bit clearer (although not more interesting). My point was quite obvious: it would be better to take into account and discuss the cognitive requirements of forms of communication that precede or develop in parallel with language (like gestures) and that presumably (following several scholars like Corballis, Li and Hombert, and others) precede language in philogeny. We have a lot of data about at least their acquisition (Bates, Volterra, and many others). Then, I simply gave some examples of possible differences - in cognitive complexity - between language and other forms of communication that might precede it. I maintain my claim: If we do not identify the cognitive requirements of primitive forms of intentional communication we cannot identify which level of social cognition and which limited and preliminary forms of ToM are necessary before language and can be part of its prerequisites for evolution. As for my point about meta-communication in speech acts an Ipse Dixit remark (reference to Wilson and Sperber) is not such a conclusive argument. I maintain my claim, that was not so clear because of the shortness and a possibly biasing reference to Grice. For sure (to me) any speech act addressed to a mind-reader 'signifies' the speaker's intentions ('to communicate to you', 'that you do something', etc.); and for sure the speaker expects so, i.e. he knows this and intends so (at least in speech acts like imperatives, requests, promises, etc.). The speaker intends that the hearer recognises her intention of communicating, that the other does something, or assume something. In my framework this is enough for having 'meta-communication'. The problem is that such a level of communication about the agent's intention about the other's understanding is also (already?) present in pointing and grasping (used as request) - at least at the mind-reading level- and also in the act of 'handing/offering' (*), or in several (but not all) cases of behavioural communication like repeatedly looking at the watch in order the other understands that I want that he leaves.

(*) In handing/offering something to you I want that you take it also because you understand that I want so, and I expect that you will understand also that I want that you understand.

A final reply to Cristiano Castelfranchi
Anne Reboul
Feb 26, 2004 15:37 UT

The hypothesis that language began with gesture is just that, an hypothesis, or, in the words of my preceding reply to Cristiano, an educated guess, which may turn out to be a just-so story. To reiterate: we don't know how language began (whether it had gestural origins or not). This is not meant to deny either (1) that linguistic communication is accompanied by gestures or (2) that gestures preceded language in philogeny. The point is mainly that we don't know. As for the remark concerning the fact that we know about the acquisition of gestures, I agree but as long as we are not sure that there was a gestural preliminary to language, we might as well have a direct look at language acquisition. By the way, quite a few hypotheses about language acquisition have it that proto-language or pre-language may have had the form of the first utterances in acquisition (i.e. no or very little morphosyntax). I'm rather bewildered about the next point: the reference to Sperber and Wilson was not an Ipse Dixit remark, but just a reference to a rather well-known work which pointed out difficulties to a generalized meta-communicative approach to linguistic communication and offered a sophisticated alternative. I'm not quite sure what the offering/taking argument is exactly supposed to show. My understanding of the example is that when you offer something to someone, you expect him/her to understand that this is a positive act and that you expect him/her to take the offering to show you that he/she takes it as a positive act. This can of course be seen as an example of mindreading but I fail to see its significance to language evolution. Interestingly, though chimpanzees use food present to cement alliances, this seems almost entirely ritualized, as is grooming. Of course, one could claim that grooming crucially involves mindreading, but that's not obvious. In more general terms, present giving/receiving is a very ritualised type of act and, as such, not necessarily implying mindreading. In other words, all reciprocal or potentially reciprocal acts do not necessarily involve mindreading. And it is not clear what exact relevance giving/receiving has to the coevolution of language and ToM.

At least two routes into language
Simon Baron-Cohen
Feb 24, 2004 23:33 UT

Anne Reboul's paper raises deep and important issues, and she is to be congratulated for such a clear exposition. Here, I throw some questions into the debate concerning the necessary precursors for language.

First, are any of ID, EDD, or SAM needed for language acquisition? EDD can't be necessary for language, since EDD operates solely in the visual modality, and we know that congenitally blind individuals do acquire language. So we can eliminate one of these candidates. If only it was all this easy.

What about ID? One might imagine that if one couldn't recognize that someone else's utterance was goal-directed or volitional, then it would just sound like sound, without any obvious function. So, ID might be necessary, or certainly useful, for acquiring language.

Finally, what about SAM? SAM is quite special because it allows minimal gricean behaviour. The possessor of a SAM, listening to an utterance, can infer that the speaker is attending to x, and that the speaker has checked that the listener has perceived that the speaker is attending to x. This could go quite a long way to helping the listener understand linguistic reference, and the need for mutual awareness between speaker and listener in communication. So this would be pretty useful in the acquisition of language too.

But back to the question: Are these necessary for language acquisition (in ontogeny, and therefore by implication, in phylogeny)? One can imagine language acquisition without an ID or a SAM. But such a route into language would rely on a learned association strategy (this sound occurs in the presence of this event/object). So understanding that the speaker intends x or is referring to y could be difficult.

Language acquisition along this second route would be somewhat bizarre. It would probably be very inflexible, since only surface behaviour would be judged. (If he made sound x last time he was in situation A, then he has to make sound x again). Anyone who has worked with young children with profound autism might recognize this kind of odd language acquisition.

My suggestion, then is that if in ontogeny it is possible for language to be acquired through the mindreading route (i.e, with the benefit of ID and SAM) or through the paired assocation route, then it might also have been possible to get so far in evolution without these mindreading mechanisms. But clearly, once such mechanisms are available, the mutual benefits they confer would rapidly mean that cause and effect could not be disentangled. Language clearly assists mindreading (since you can ask people what's on their mind) and mindreading clearly assists language (since if they change their surface behaviour you can still hazard a guess at their underlying intended meaning).

> **Reply to Simon Baron-Cohen**
> Anne Reboul
> Feb 25, 2004 11:54 UT
>
> I agree of course with Simon Baron-Cohen well taken point that EDD is not necessary for language acquisition because blind children do acquire language (and ToM by the way). I also agree that ID and SAM are not mandatory to learn a language given that an association system pairing sounds and objects may be enough to learn language (that is presumably how some autistic children do learn language). However, I think that (and I take Simon's final comments to mean just that) one should presumably distinguish between learning a language (which can be done at any age, however laboriously, and which is up to a point an artificial process) and acquiring a language which seems to be a natural process, not involving any heavy tutoring. Now, I never said that one could acquire language (in the above sense) without ID (ID is in place at the beginnings of language acquisition) nor did I claim that SAM (and later ToM) don't play any role at later stages of language evolution and thus I fully agree on Baron-Cohen's final comments on the ontogeny of language. Just a final comment: it may be that ID is in place from birth on, though it is notable that just-born infants dicriminate between language and other sounds. Though this might be ID's work, it might equally be some inborn ability, necessary for language acquisition. This, however, does not mean that ID does not play a role in the early stages of language acquisition, just as the fact that SAM and ToM do not seem to play a role in early language acquisition does not mean that they don't play a role at later stages.

Verbal comprehension is metarepresentational
Dan Sperber
Feb 26, 2004 18:01 UT

I am happy to see that Anne and most of the participants in this discussion agree that naïve psychology – 'theory-of-mind' is such an inappropriate name for it, or even for a part or a stage of it – is comprised of several components, with distinct phylogenetic and ontogenetic patterns of emergence. Taking this for granted, I would like to make three points:

1 – One important distinction among components of naïve psychology is that between those that involve the attribution and metarepresentation of mental conceptual content, and those, such as Shared Attention, that don't. I believe that there are several metarepresentational component of naïve psychology, and not just the kind of full-fledged belief-desire psychology evidenced in chidren who pass the false belief task. For instance, before this, children exhibit what might be called 'knowledge-desire psychology' with genuine attributions of contents that are hardly controvertible, at least on the desire side.

2 – In particular, Deirdre Wilson and I have argued that the attribution of meanings to speakers (Gricean's speaker's meanings) is metarepresentational and is well-evidenced well before children pass the false belief task. So, for us - as for Anne -, verbal communication does not involve 'full-fledged ToM' (but see next point), but it does involve metarepresenational abilities. We may be wrong, of course, but what is the evidence, experimental or otherwise, that trumps, or at least weakens, the ubiquitous observational evidence that young speakers (three years and less) understand speaker's meaning and not just sentence meaning, and aim at having their meaning understood?

3 – I am not sure that we want a notion of full-fledged ToM, or, for that matter, full-fledged naïve psychology. I am sure that if we want such a notion, there is little reason to assume that passing the false belief task is evidence of possession of this full-fledged psychology. It is not implausible, for instance, that some components of naïve psychology kick in only at adolescence.

> **Reply to Dan Sperber**
> Anne Reboul
> Feb 27, 2004 10:30 UT
>
> Once again, I find myself in agreement with most if not all of Dan's comment. I entirely agree with his first comment. I also basically agree with his second comment which I take to mean that, though the process of understanding does not necessarily (but may) involve mindreading, the outcome of the process is couched in metarepresentational terms. As said previously, I was merely pointing out that the first stage of language acquisition does not imply true mindreading abilities (such as SAM for instance), not that later stages and linguistic communication as such never need them. I think that Dan's comment points out a new direction: i.e. not only the necessity to distinguish between different components of mindreading with distinct phylogenetic and ontogenetic patterns, but to distinguish between the components which would lie more on the behavior reading side (if I do this, then (s)he will do that) and those which would lie more on the mindreading side (if I do this, then (s)he will believe p, and, as a result, (s)he may do that). As well, it is necessary to distinguish (presumably) among both behavior reading and mindreading components that imply conceptual (as opposed to non-conceptual) metarepresentation. What is more, this is not only necessary for theoretical reasons, it's also necessary for experimental reasons: though I still think that the false belief test is important, I think that it's very far from being sufficient and that many other tests (not necessarily implying language) should be developped. This leads me to the last and third point of Dan's comment: the term of "full-fledged" ToM or naive psychology probably is unfortunate. I entirely agree that naive psychological abilities presumably go on developping past the age at which the false belief test is passed — indeed this is rather obvious given that the opaque contexts tests are not passed until a year later —, and, what is more, I think that the development of naive psychology presumably goes on in adulthood, just as does the development of vocabulary and other abilities. I'm not sure about other components of naive psychology kicking at adolescence, though I'd be ready to bow to evidence if any is available. I just want to end up by a short and limited defence of the false belief test. I'm afraid that my paper led people to think that I advocated dropping it altogether. That was not my goal and neither is it my opinion. I think that the false belief test has a limited but real utility in that it allows the test of the ability of individuals to attribute to others beliefs that differ from their own. I'm not defending the present form of the false belief test — I'm perfectly open to the idea that it should be modified, for instance to test whether the age limit which the present form seems to yield is a reality or an experimental artefact —, I'm just saying that the ability to attribute to other people beliefs different from one's own is a good test

of (some) conceptual metarepresentational abilities, mainly because attributing to other people the same beliefs as one holds does not seem economic: as long as beliefs are reasonably similar, you can reason from your own. This is common, for good reasons I think, to both theory ToM and to simulation theory of mindreading and should be kept. My goal was only to point out that it is not clear that merepresentational abilities (even non-conceptual language abilities) were necessary (in the logical sense) for the first stages of language acquisition and that they may not have been necessary in the first stages of language evolution. This led me to coevolutionary scenario, rather than to successive evolution scenario, not such a big revolution after all.

Constructions underlying theory of mind and language

Peter Ford F. Dominey (CNRS - Institut des Sciences Cognitives, Lyon)

(Date of publication: 1 March 2004)

Abstract: Together, language and theory of mind are distinguished by their recursive compositional nature, posing the difficult question of how such a recursive representation capability could have evolved. The current article provides the outline for the evolution of a mechanism that allows representation of social events, via their mapping onto intentional schema representations in a progression from holistic to increasingly abstract and compositional mappings. This progression is analogous to one that has previously been described for grammatical constructions in language. The implied shared use of this structure mapping capability for language and ToM will be developed.

1. Introduction

Both language and ToM are uniquely human, both involve manipulation of complex embedded structures, and both are subject to debate concerning the articulation of genetic and developmental processes. Likewise, there appears to be a crucial co-evolutionary relation between them that is at the crux of this conference. Does this relation extend to the sharing of a common underlying processing mechanism? The objective of this paper is to argue for the idea that the notion of "construction" as a mapping between representations can be extended from language and grammatical constructions to the domain of social behavior and theory of mind. In this context, we can consider that language is about mapping sentences to meaning, and theory of mind is about mapping social/behavioral contexts onto other's behavior (in terms of their underlying mental states).

From the language perspective, the construction grammar framework has been an important component of the functionalist approach to linguistics (see Goldberg 1995, Croft 2001, Tomasello 2003). The essential claim in this framework is that language is learned as a structured inventory of form to meaning mappings. In making this claim the framework exploits the inherent richness of the structure of meaning in form to meaning relations, and thus reduces the requirements on an innate universal grammar. The question posed here will be to what extent can the benefits of this construction model be applied to ToM. That is, can some or all of the ToM capability be provided by a learning mechanism that acquires the mappings between social/behavior contexts and their outcomes in progressively abstract and compositional manner?

Corcoran and Frith (2003) present results supporting the theory that at least part of a ToM capability might rely on analogical mapping between the current situation and autobiographical memory of similar situations in order to draw inferences about the mental states of others. This suggests the notion of *ToM constructions*, that is, mappings between behavioral scenario structure and the corresponding social/intentional outcome providing the point of departure for a construction based analysis of ToM, analogous to that in language. The rest of the paper will sketch out a brief overview of the construction grammar framework for language, and then a view of how the underlying structure mapping mechanism can be applied to provide a construction based theory of mind, ConTom.

2. Brief Review of Language in the Construction Grammar Framework.

As mentioned above, a significant scholarly effort has gone into the elaboration of the construction grammar framework for language that can be found in references including but clearly not limited to Goldberg (1995, 2003), Croft (2001), and Tomasello (2003). One of the most attractive elements of this framework is that the poverty of the stimulus argument, that has been so central in evoking the need for a highly pre-specified Universal Grammar (UG), is significantly weakened by reconsidering the powerful learning and attention sharing mechanisms and the richness of the stimulus (Goldberg 2003, Tomasello 2003, Pullman & Schultz 2002). In this context, the current analysis will attempt to identify perceptual primitives required to extract meaning from the environment, and structure mapping mechanisms that construct the mappings between language and this meaning. The goal will be to accommodate (sentence, meaning) pairs that embody some of the interesting aspects of language including its embedded recursive structure, as characterized in the relativised sentence “The boy kicked the ball that broke the window”.

In the technical sense, both words and larger phrasal patterns are constructions, as both pair form with meaning. The mapping from words to meaning retains a rather classical lexical mapping, and phrasal constructions can also have this holistic nature as in idiomatic phrases such as “Gimme that”, or “Don't beat around the bush”. Phrasal constructions can also be abstract and generative, as in the English transitive construction in which the arguments AGENT, ACTION and OBJECT can be instantiated by an open set of nouns and verbs to generate diverse sentences such as “John kicked the ball” and “The cat chased the dog”. The transition from these abstract argument constructions to generative, recursive compositionality results from the statistical pattern finding and extraction of phrasal and clausal structures such as noun phrases that

can then occupy argument positions in existing constructions, thus yielding a flexible compositional capability. This type of capability has been demonstrated in a hybrid neural network architecture by Miikkulainen (1996).

As stated by Goldberg (2003) a characteristic aspect of the construction grammar framework is that it is a "what you see is what you get" approach to syntax, in which there are no underlying levels of syntax, nor phonologically empty elements posited. In contrast to formally oriented generative frameworks, this view of language places a large emphasis on general structural mapping mechanisms, and the importance of the structure of meaning. The next paragraphs outline a "mechanistic" view of how this could work.

2.1 Perceptual Primitives

Events, Agents and Objects: In order to learn sentence to meaning mappings, the child/system must have a capability to extract meaning from perception. This includes the ability to discriminate between distinct objects and to perceive and represent events in terms of their agents and objects in a predicate-argument format. Given these capabilities, the infant can represent the event described by the sentence "The boy kicked the ball" in some kind of predicate-argument format "kick(boy, ball)". This corresponds to the first level of Leslie's theory of agency, the Theory of Body mechanism (ToBy) that characterizes the "mechanical" aspects of agents and objects and their interactions, as characterized for example by physical collision events. From this perspective, already at 6 months of age, children are capable of processing causal events with agents, objects and actions and using these "naive physics" representations to understand simple action scenarios that involve goal-directed reaching for objects (e.g. Woodward 1998). Similarly, infants in this same age range display rather sophisticated knowledge of the physical properties of objects that allows them to "parse" and understand dynamic scenes with multiple objects (Carey and Xu 2000).

Demonstrating this functionality, computer vision systems now exist that can extract force dynamic information including physical contact and relative velocity of objects in order to parse events and their thematic arguments from visual scenes (Siskind 2001, Dominey 2003a).

Referential Ambiguity: While it seems apparent that infants can extract and construct meaning from the perceptual world (Mandler 1999), there will potentially be massive meaning to choose from in that perceptual world, and so in a language learning context, the infant must somehow zero-in on what the speaker is talking about. Joint attention refers to this capability for the infant to follow (and later direct) the gaze of the speaker in order to establish a shared frame of reference around a third object (reviewed in Tomasello 1999). This capability begins to emerge well before the first birthday, and allows the formation of a triadic relation between the speaker, the infant, and the object of shared attention. It should be made clear that this joint attention mechanism makes a substantial and crucial reduction in referential ambiguity that will be required for language acquisition. It is important to distinguish this joint attention from the shared attention mechanism (SAM) of Baron-Cohen (1995) that generates embedded relations such as "John sees(I see the girl)", and instead to associate it with the eye direction detection (EDD) capability proposed by Baron-Cohen. It is also worth mentioning that even in blind children who learn language, joint attention still plays a crucial role, though not – of course – in the visual modality.

2.2 Structure mapping

Given these meaning extraction and attentional capabilities, the goal of the structure mapping system is to learn to associate different grammatical forms (e.g. active, passive, relative) with their associated meaning structures. Concretely such a system is provided with (sentence; meaning) pairs, such as (John hit the ball; hit(John, Ball)), (The ball was hit by John; hit(John, Ball)), (The ball that John hit broke the window; hit(John, ball), broke(ball, window)). Based on corpora-scale exposure to such examples, the system should extract the underlying structural mapping from grammatical form to meaning. In order to achieve this, the system should exploit the cross-linguistic regularity (Bates et al. 1982) that cues including case and grammatical markings (fixed or free) and word order uniquely identify each grammatical form and thus allow each construction type to be uniquely associated with its corresponding form to meaning mapping. We have demonstrated that such a mechanism can accommodate a variety of grammatical constructions in English, including embedded relative clauses, and demonstrates the ability to systematically generalize to new sentences (Dominey 2003b), and to extend without modification to accommodate Japanese, despite the significant typological differences between these languages (Dominey & Inui submitted). This limited overview is not at all meant to be complete, but rather to outline the theoretical framework CG for potential application to ToM.

3. A Construction Based Approach to Theory of Mind (ConTom)

Thus, given the above analysis of language in the construction context, we can now apply an analogous analysis to the domain of ToM. Again, from a functional perspective we will consider that ToM refers to the capacity to interpret, predict and explain the behavior of others (in terms of their underlying mental states). The inputs to the ToM system are behavioral situations, the processing is the recognition and analysis of these situations and the output is the prediction/explanation of the of the future behavior. Among the desired

behaviors will be to interpret the goals and behavior of others, as revealed by imitation, and to demonstrate ability to exploit embedded propositional attitudes as revealed by the use of interpreting other's actions based on their false beliefs. As suggested by Corcoran and Frith (2003), ToM capacities appear to rely at least in part on reference to autobiographical memory that allow inferences to made about the mental states of others. The next paragraphs provide an outline of how this could work in the construction framework.

3.1. Perceptual Primitives

In addition to the joint attention and event processing capabilities required for language processing as described above, it will be seen below that the child should also be equipped to detect "satisfaction" or "happiness" when an observed agent has successfully completed a goal. This will be important for post-hoc linkage of preconditions of actions with their goals.

3.2 Structure mapping

Here we will consider a capability that allows mapping of initial states or situations to behavior outcome states. Again, as suggested by Corcoran & Frith (2003) these mappings can be stored in an "autobiographical" memory that we can refer to as a "social construction inventory". Then, new situations can be interpreted based on (analogical) reference to constructions in this inventory. These constructions will be of the form (initial observed behavioral state; outcome behavioral state). When confronted with a new situation, by finding the closest match to this new situation in the "initial state" component of the social construction inventory, the child can then use the associated outcome in order to infer or predict the outcome for the new situation. Pattern finding processes will operate on the contents of the social construction inventory in order to generate progressively more abstract and generalized ToM constructions in the same manner as in the grammatical construction framework. Interestingly, in a population of schizophrenic patients, Corcoran and Frith (2003) observed correlated impairments in autobiographical memory, and in theory of mind task performance. We now consider examples of how a construction based theory of mind system might operate.

Goal attribution: Here we consider a behavioral scenario in which the child observes his Mother look at, reach for and pick up a bottle. Repeated exposure to this kind of event sequence will allow the child to learn to predict that Mother will subsequently take the bottle that she is currently looking at and reaching for. Technically, this corresponds to the mapping ((Look-at(Mother, bottle), Reach-for(Mother, bottle)); Pick-up(Mother, bottle)). As suggested by Leslie (1994), the outcome of the action can be entered into the construction representation as the goal or outcome state. Initially this may correspond to a "holo-construction" strictly linked to the Mother and bottle. As in grammatical constructions, with exposure and pattern finding, this holo-construction be extended first to include Mother and other objects, and then to other agents, and actions as well. Functionally, this capability should allow the child to imitate an uncompleted action in which the adult tries and fails to achieve a well know action, thus indicating the child's ability to infer the intended goal (Bellagamba & Tomasello 1999).

Attitude attribution: How can this attribution of goals be extended to behavior that reflects attribution of attitudes, allowing the child to have the meta-representation that "John wants Bill to give him an apple." Let us work through an example: Suppose that the child observes a scene in which Bill takes an apple out of a sack, and John approaches, looks at the apple then at Bill. Bill then gives the apple to John, who takes it and smiles. If the infant is sensitive to the meaning of the smile, then she will conclude that John is happy to have the apple, and that a goal has been satisfied. In the construction framework, the initial state is characterized by the Bill's possession of the apple, and John's approach behavior. The state transition is the giving of the apple from Bill to John, and the outcome is John's possession of the apple, and his satisfaction. In a repetitive learning situation, the child will learn the mapping from the initial state to outcome state. The backward linking of John's final satisfaction to his approach behavior in the initial state to allows the child/system to define a perceptual correlate of "wanting" in terms of goal satisfaction. Again, the resulting ability to determine that John wants Bill to give him the apple will start as a fixed holo-construction, that can subsequently generalize to different objects and participants, based on the exposure of the child to training examples, resulting in a generalized approach-based goal-attribution construction.

Extension to false beliefs: Imagine now that one day, the child sees that Bill has no apple in his sack. As usual, John approaches Bill in the same manner, "wanting" Bill to give him the apple that is (not) in his sack. Though John approaches Bill in his standard manner, Bill has no apple to give, and John is left dissatisfied. John approaches Bill and expects an apple based on his false belief that there is an apple in the bag. This learning example for a false belief situation provides data for the autobiographical social construction inventory that Bill acts on his false belief that John has an apple, despite that fact that the child knows that there is no apple. Again, through the operation of statistically based pattern finding, this will result in the progressive development of a generalized false belief capability.

4. Discussion and Conclusions

From the perspective of co-evolution, one result of this exercise is the suggestion that the meta-representational aspect of theory of mind was not a precursor for language with recursive embedded structure (see Reboul, this conference, and discussion). Leslie likewise notes that structural linguistic knowledge and language processing mechanisms are essentially independent of ToMM, though ToMM's development may impact on communicative language use (Leslie 1994). What is required however, is a joint attention capability – visual or otherwise, that allows teacher and learner to talk about a common referential object.

With respect to the proposed construction based theory of mind (ConTom), we can observe that this is comparable to the teleological reasoning in infancy as described by Gergely and Csibra (2003). This is a non-mentalistic system that allows inference about other's goal directed actions based on perceptual aspects of reality without attributing intentional mental states to the actor's mind. The question then is whether this type of teleological system can extend naturally to a mentalistic system that allows meta-representations including propositional attitudes about propositional attitudes. There are two components to the response.

First, purely from the perspective of structural mapping, given a representation of the form (believes(John, wants(Mary, Apple))) can the ConTom system accommodate such a representation as a component the outcome of a ToM construction? I believe that the response is an obvious yes: the construction framework is based on a generic analogical mapping capability that maps well formed inputs onto their corresponding outputs, whether these (input; output) pairs are (sentence; meaning) pairs in language, or (social context; behavioral/mental outcome) in the ToM domain. As long as the outcome is a well formed and reproducible it can be paired with its initial state in the social construction inventory.

Second, from the perspective of generating meta-representations like (believes(John, wants(Mary, Apple))), can the ability to generate such representation be developed directly from the teleological system, or does it require an additional mind reading module corresponding to something like Leslie's ToMM2? In a certain sense we approach issues related to Searle's (1980) Chinese room in that this construction based approach can likely be extended to yield behavior indicative of a m-representation theory of mind (as indicated in the false belief example above), with the open issue of whether "understanding" has really been captured in computational terms (Horst 2003). While the response is beyond the scope of this paper, we can speculate that in the developing child, if self beliefs (e.g. "I thought it was raining") become accessible as elements contributing to social constructions, then they can provide the grounding for understanding and generalization to the application of such representations to others.

In conclusion, the stated goal of this exercise was to demonstrate that a structure mapping capability that has been described for grammatical constructions in language can generalize to explain aspects of human theory of mind. In agreement with the theory and results of Corcoran and Frith (2003) on the potential role of autobiographical in theory of mind, I have worked through a demonstration sketch that supports the stated goal. It is of significant interest that this further elevates the status of a generalized analogical structure mapping capability in the hierarchy of cognitive functions by demonstrating the effectiveness of this analogical mapping as a basis for language and theory of mind construction frameworks.

References:

Baron-Cohen, S. 1995 *Mindblindness* , MIT Press.

Bates E, McNew S, MacWhinney B, Devescovi A, Smith S (1982) "Functional constraints on sentence processing: A cross-linguistic study" *Cognition*, 11, 245-299.

Bellagamba F, Tomasello M (1999) "Re-enacting intended acts: comparing 12- and 18-Month-Olds", *Infant Behavior and Development* 22(2) 277-282.

Carey S, Xu F (2001) "Infant's knowledge of objects: beyond object files and object tracking" *Cognition*, 80, 179-213.

Corcoran R, Frith CD (2003) "Autobiographical memory and theory of mind: evidence of a relationship in schizophrenia" *Psychological Medicine*, 33 897-905

Croft W (2001) *Radical construction grammar: syntactic theory in typological perspective*, Oxford: Oxford University Press.

Dominey P.F. (2003) "Learning Grammatical Constructions from Narrated Video Events for Human-Robot Interaction", IEEE Conf. On Humanoid Robots, Karlsruhe Germany.

Dominey, P.F. (2003a) "Learning Grammatical Constructions in a Miniature Language from Narrated Video Events", Proceedings of the 25th Annual Meeting of the Cognitive Science Society, Boston.

Dominey P.F., Inui T (2004) "Miniature Language Learning via Mapping of Grammatical Structure to Visual Scene Structure in English and Japanese", submitted.

Frith,C.D.,and Frith,U.1999 "Interacting minds - a biological basis" Science 286:1692 -1695.

Gergely G, Csibra G (2003) "Teleological reasoning in infancy: the naïve theory of rational action", *Trends in Cognitive Science*, 7(3)287-292.

Goldberg, A. (1995) *Constructions* University Chicago Press, Chicago and London.

Goldberg, A. (2003) "Constructions: a new theoretical approach to language", *Trends in Cognitive Science*,

Volume 7, Issue 5 , May 2003, Pages 219-224.
Horst, S., "The Computational Theory of Mind", The Stanford Encyclopedia of Philosophy (Fall 2003 Edition), Edward N. Zalta (ed.), URL = http://plato.stanford.edu/archives/fall2003/entries/computational-mind/>.
Kotovsky L., Baillargeon R., " The development of calibration-based reasoning about collision events in young infants". 1998, *Cognition*, 67, 311-351.
Leslie A.M. (1994) "ToMM, ToBy, and Agency: Core architecture and domain specificity", in *Mapping the mind: domain specificity in cognition and culture* (LA Hirschfeld & SA Gelman, Eds.)
Mandler J. (1999) éPreverbal representation and language", in Bloom et al. (eds) *Language and Space*, 365-384.
Miikkulainen R. (1996) "Subsymbolic case-role analysis of sentences with embedded clauses". *Cognitive Science*, 20: 47-73.
Pullman G.K., Sholz B.C. (2002) "Empirical assessment of stimulus poverty arguments" *Linguistic Review* 19, 9-50.
Reboul A (2004) "Evolution of Language from Theory of Mind or Coevolution of Language and Theory of Mind?" at: http://www.interdisciplines.org/coevolution .
Searle, John. (1980) "Minds, Brains and Programs" Behavioral and Brain Sciences 3:417-424.
Siskind JM (2001) "Grounding the Lexical Semantics of Verbs in Visual Perception Using Force Dynamics and Event Logic", *Journal of Artificial Intelligence Research*, volume 15, pp. 31-90.
Tomasello, M. (2003) *Constructing a language: A usage-based theory of language acquisition.* Harvard University Press, Cambridge.
Tomasello, M. (1999) *The cultural origins of human cognition* Harvard University Press, Cambridge.
Woodward A.L. (1998) "Infants selectively encode the goal object of an actor's reach" *Cognition* 69 1-34.

Discussion

Meta-representation takes more than embedding
Dan Sperber
Mar 2, 2004 15:41 UT

Compare:

(1) (a) Mary arranged that John could come to the party

(b) Mary wished that John could come to the party

Both (1a) and (1b) contain what looks like the same embedded clause, *(that) John could come to the party*. However, in (1a), this embedded clause represents a state of affairs, whereas in (1b) it represents the content of Mary's wish. If John happens to be Jack the Ripper, then (1a) and (2a) are truth-conditionally equivalent. Not so with (1b) and (2b), if Mary does not know that John is Jack the Ripper, then (1b) may be true and (2b) false:

(2) (a) Mary arranged that Jack the Ripper could come to the party

(b) Mary wished that Jack the Ripper could come to the party

This is standardly described by saying that *(that) John could come to the party* is used transparently in (1a) and opaquely in (1b). Equivalently, *(that) John could come to the party* is used meta-representationally in (1b).

The point is that the ability to embed a construction in another is not sufficient for the ability to metarepresent. What is required is the ability to use a representation (mental or linguistic) in two ways: transparently to represent some state of affairs, and opaquely – or meta-representationally -- to represent other representations with identical or similar contents. The ability to metarepresent is a distinct *semantic* ability.

Peter does not address the issue. Could he explain, in constructionist terms, how a cognitive system capable of just a transparent use of representations would acquire the novel semantic ability to use its representations meta-representationally?

Behavioral Grounding of Meta-Representation in a Construction Based Theory of Mind
Peter Ford Dominey
Mar 3, 2004 21:43 UT

In his comment “Meta-representation takes more than embedding,” Sperber asks “how a cognitive system capable of just a transparent use of representations would acquire the novel semantic ability to use its representations meta-representational?” As stated elsewhere by Sperber (1) *Metarepresentations are representations of representations, but not all representations of representations are metarepresentations in the relevant sense. The human metarepresentational capacity we are interested in here is, first and foremost, a capacity to represent the content of representations*.

In the discussion of the target paper I suggested that the use of self-beliefs (e.g. “I thought it was raining”) might provide a semantic bootstrap for this metarepresentation. I thank Dan for providing the opportunity to enlarge on this. The response will be in the spirit of the construction framework, that is, with specification of an initial and limited metarepresentation capability, highly grounded in relevant behavior, that can then be generalized, analogous to generalization from holoconstructions to abstract constructions in language.

Sperber provides the initial capability by supposing that *an animal might detect the fact that a conspecific has been suitably impressed with some display of strength, and represent it by means of a single unarticulated symbol knows-that-I-am-stronger-than-him as in 4b*

(4) (b) he knows-that-I-am-stronger-than-him

Such animals would possess a very rudimentary metarepresentational capacity, lacking compositionality and recursion. They could only metarepresent a short and fixed list of representations.”

Interestingly, by allowing this, it seems that Sperber has made the leap of metarepresentation, albeit rudimentary, and that the issues of compositionality and recursion can be handled separately, by the same type of mechanism through which grammatical constructions start as holistic and become increasingly abstract and compositional. In this direction, Sperber proceeds to extend this picture to a species with a rich ability to form and use mental representations of indefinitely varied contents. In this case the fixed set of holo-constructions above will not do for *representing indefinitely many of the mental states of their conspecifics*

The question is, can a system learn to use conceptual representations "opaquely" or metarepresentationally in terms of the contents of conspecifics’ mental states. Sperber holds that *Such an organism might be capable of representing, say, the fact that it is raining, but never of representing the fact that it is representing the fact that it is raining*.

I will propose a possible pathway by which this impasse can be traversed. Reconsider the same organism capable of mental representations with indefinitely varied contents. Imagine further that this organism has memory of its past behavior and past mental representations, and the ability to compare and detect differences between contents of different representations in the context of its behavior. At 12:00 the organism believes (represents the fact that) it is raining and thus takes its umbrella, and leaves its house. At 12:02 the organism emerges from the house into the bright sunshine, and at the same time the organism is burdened by the weight of the umbrella. The organism then attempts to determine why it has this heavy umbrella, and recalls that at 12:00 it (falsely) believed (represented the fact) that it was raining. This leads to the detection of the violation between the contents of the recent false belief and the contents of the current observation-based correct belief. Exposure to this type of violation should allow a statistical learning/categorization mechanism to develop a fixed holo-social-construction corresponding to “I took my umbrella because I thought it was raining”. In this case the “it was raining” represents the content of the organism’s belief, rather than a simple state of affairs, thus satisfying Sperber’s opacity requirement in a special case. The structural requirement on the system to enable this capability is the ability to store, retrieve and compare mental representations, a

relatively straightforward capability (from a functional neurophysiology perspective) that provides a non-linear change in performance.

Imagine now that the organism encounters its colleague, Jane, walking around in the bright sunlight with an umbrella. The organism will interrogate its autobiographical store of social constructions and make the analogical mapping between the states “Jane in the sun with her umbrella” and “Me in the sun with my umbrella,” in order to metarepresent “Jane brought her umbrella because she thought it was raining” via analogical mapping. Through the operation of pattern finding mechanisms on such experiences, this type of social construction will become progressively more abstract, general and compositional (as for grammatical constructions in language), thus satisfying the opacity requirement in the general case.

(1) References to Sperber are cited from: Sperber, Dan: "Metarepresentations in evolutionary perspective" (in Dan Sperber (ed.) *Metarepresentations: A Multidisciplinary Perspective*, Oxford University Press, 2000, pp.117-137). http://www.dan.sperber.com/metarep.htm

It ain't that easy
Dan Sperber
Mar 5, 2004 21:12 UT

In order to explain how one could *acquire* a genuine metarepresentational ability (with representation of the contents of representations metarepresented), Peter asks us to imagine an *“organism capable of mental representations with indefinitely varied contents. Imagine further that this organism has memory of its past behavior and past mental representations, and the ability to compare and detect differences between contents of different representations in the context of its behavior.”* This “solution” presupposes what it is meant to resolve. An organism capable of attributing to itself past representations or to compare contents of different representations is already metarepresenting contents, it already has a metarepresentational ability. Peter’s proposal boils down to having such a metarepresentational ability start with self-attribution, and then generalize to attributions to others. I see little reason to assume that attributions of mental states to oneself are prior, phylogenetically or ontogenetically, to attributions to others, but even if they are, this does not tell us how a metarepresentational ability could be acquired in the first place, which is what I was asking. How can one explains the acquisition of a new representational format (for instance, a metarepresentational format) that can represent things (for instance representations) which were not representable in previously available formats? To me, the task is of a comparable difficulty as would be that of explaining how an organism capable of seing in black and white could acquire the ability to perceive colours. My hunch is that a genuine metarepresentational ability has to be just as innate as colour vision.

Meta-Representation as Context Dependence
Peter Ford Dominey
Mar 9, 2004 11:27 UT

In his comment “It ain’t that easy” Sperber says “An organism capable of attributing to itself past representations or to compare contents of different representations is already metarepresenting contents, it already has a metarepresentational ability.” He then asks “How can one explain the acquisition of a new representational format (for instance, a metarepresentational format) that can represent things (for instance representations) which were not representable in previously available formats?”

The fundamental question has now been posed in a more precise and reduced manner, and our objective now is to determine indeed if the format change can be made in the context of the existing machinery, which Dan believes is not possible.

I believe that we can all agree that in the pre-meta-representing organism, there exists a capability to represent something like predicate argument relations with a content such as "An apple is in the bag" in a form such as IN(APPLE, BAG). The question now is to determine if this capability can be extended to represent a propositional attitude such as "John believes the apple is in the bag" corresponding to BELIEVE(JOHN, IN(APPLE, BAG)). While we're at it, the system should also be able to simultaneously represent that Mary believes that the apple is on the table, and that Paul believes

Mary believes that the apple is in the refrigerator. That is, the system must be capable of simultaneously entertaining a variety of possibly divergent and opaque representations, of the form:

1. BELIEVE(JOHN, IN(APPLE, BAG)) 2. BELIEVE(MARY, ON(APPLE, TABLE)) 3. BELIEVE(PAUL (BELIEVE(MARY, IN(APPLE, FRIDGE))))

From a purely functional perspective, the representation problem here can resolved by a context-dependant associative memory. Allowing a system to respond, behave, or remember differently depending on the current behavioral context is classic and well studied behavior (see Dominey & Boussaoud 1997). In the current context, when the system considers (e.g. attends to) each separate individual (John, Mary) it enters into a different pattern of neural activity that serves as a context. In each of these different contexts, the associated proposition can be linked with the context so that for each individual (i) the appropriate belief or set of beliefs (or propositional attitudes in general) PA(i) will be indexed and retrieved. In other words, the system will employ a set of context dependant representational frames. Each representational frame can be linked to a different individual (or an individual at a different time). These contexts can be nested, and can thus provide a concrete and realistic first step to meta-representation. The neural implementation of such representations has been studied by Pollack (1990).

For the example (1) above, for the context “JOHN” the representation IN(APPLE, BAG) is activated. Similar for example (2). For example 3, in the context “PAUL” there is another context “MARY” and in this context is represented IN(APPLE, FRIDGE). This implies a sequential activation of contexts for PAUL and then MARY, and the neural control of sequential activity is well studied and clearly available to the organism.

While Dan is looking for something qualitatively different, I think it is worth the effort to consider whether or not this kind of mechanism can provide the representational capabilities required for a representational theory of mind. The exposition above outlines a system that allows for the representation of different propositions in a context dependant manner in which contexts correspond to different agents. This provides a system for representing propositional attitudes of self and others, and thus appears to explain how meta-representation can be acquired in an existing representational format if we allow the introduction of context dependence. Again, context dependant learning is well documented from the behavioral and neurophysiological perspective in non-human primates (see Dominey & Boussaoud 1997), so the step can apparently be made.

Dominey PF, Boussaoud D (1997) Encoding behavioral context in recurrent networks of the frontostriatal system: A simulation study. Cognitive Brain Research, 6, 53-65

Pollack JB (1990) Recursive Distributed Representations, Artificial Intelligence, 46, 77-105

The objection stands
Dan Sperber
Mar 9, 2004 18:51 UT

I am afraid I have failed to convey the gist of my objection, namely, to repeat, that **meta-representation takes more than embedding**. Peter writes:

I believe that we can all agree that in the pre-meta-representing organism, there exists a capability to represent something like predicate argument relations with a content such as "An apple is in the bag" in a form such as IN(APPLE, BAG). The question now is to determine if this capability can be extended to represent a propositional attitude such as "John believes the apple is in the bag" corresponding to BELIEVE(JOHN, IN(APPLE, BAG)). While we're at it, the system should also be able to simultaneously represent that Mary believes that the apple is on the table, and that Paul believes Mary believes that the apple is in the refrigerator. That is, the system must be capable of simultaneously entertaining a variety of possibly divergent and opaque representations

The contrast I am trying to convey is that between (to use Peter’s formalism) BELIEVE(JOHN, IN(APPLE, BAG)) and CAUSE (JOHN, IN(APPLE, BAG)). In plain English, “John believes that the apple is in the bag” and “John makes it so that the apple is in the bag”. In the latter statement, “the apple is in the bag” is used transparently. It denotes an apple in a bag, or else it is misused. In the “John believes...” case, “the apple is in the bag” is used opaquely. It might be that the apple is not in

the bag, or that what is in the bag is a pear that John mistook for an apple, and so on, and still, “the apple is in the bag” might be used appropriately. How so? Because it represents not a state of affairs but the content of John’s belief. The same words represent different things (e.g. apples in bags and beliefs in minds) in their transparent and their opaque interpretations. They don’t have the same semantic value. *(TIP: some philosophers such as Davidson deny this -- try to use them in your defense)*. It is not however a case of homonymy. It is a case of using a special semantic resource: opacity or metarepresentation. Without the ability to interpret “the apple is in the bag” meatrepresentationally, it is not an interpretable object for the verb “believe”, just as “air” would not be an interpretable subject for it.

My question – still unanswered – is how could an organism capable of transparent use and, let me add, of all the embedings you want, ACQUIRE the ability to interpret embedded expression metarepresentationally?

A necessary distinction between necessary and sufficient conditions
Anne Reboul
Mar 10, 2004 8:15 UT

Unless I'm much mistaken, the issue of the discussion between Peter and Dan Sperber is Dan's pointing that though embedding is a necessary condition of metarepresentation, it is not a sufficient condition. As Dan points out, one also needs an ability to distinguish between opaque and transparent uses of words. I agree and I would like to say that this highlights the importance of distinguishing between necessary and sufficient conditions for the possession of a given ability. This, I think, might be a general comment on Peter's proposal both for language acquisition or evolution and for ToM acquisition or evolution. I read Peter's paper as aiming to show that both language and ToM might not have evolved in any biological sense of new structures specific to them being selected for, but that both might emerge (in a non-biological sense) from previous structures not dedicated to them. This is a perfectly legitimate position, provided that one can show that these previous structures are not only necessary but sufficient conditions for language and ToM. What Dan is pointing out in his successive comments is that embedding, whether linguistic or non-linguistic, is a necessary but not a sufficient condition for metarepresentation (which is a central ingredient in any type of mindreading). For language, Peter adopts the construction grammar framework. Though I'm sympathetic with some aspects of the framework, it has always struck me that on this view the generative power of language is either downplayed and/or not adequately explained. For instance, Bybee (2002) makes the claim (made on the basis of statistical analysis of corpora) that 55% of spoken and written texts are made of prefabricated units. This seems to be an argument for the fact that language acquisition and language use are made by "chunking" and that no other structural explanation of language is necessary. It strikes me that what this statistics shows may be exactly the reverse: in other words, there remains 45% of linguistic production which is NOT made of prefabricated units and which does not seem adequately explained by "chunking". In other words, what such statistics make clear is the generative power of language. What is more, they make it clear, but construction grammar does not clearly explain it. Thus, there is some temptation to think that the processes described by Peter and more generally in Construction Grammar may be necessary to language acquisition/evolution, but that they are not sufficient. And, unless they are shown to be sufficient, there is still room for the evolution (or coevolution) of specific (and possibly modular) structures for language and mindreading.

References Bybee, J. (2002), "Sequentiality as the basis of constituent structure", in Givon, T. & Malle, B. (eds) The evolution of language out of pre-language, Amsterdam, John Benjamins, 109-134.

What is Sufficient
Peter Ford Dominey
Mar 10, 2004 17:20 UT

Anne Reboul makes the highly useful remark that the construction based approach might be able to identify processing capabilities that are necessary for language and ToM, but that the case has not been made for these capabilities being sufficient for language and ToM.

Indeed. Let me first say that I with respect to ToM, I clearly acknowledge that embedding does not imply opaque meta-representation. In this context, I have now proposed the conditions under which a transparent capability could become opaque. Likewise, in response to Anne on compositionally in the language domain, I would again clearly acknowledge that while a fixed set of constructions will be of great use, any theory of language must account for compositionally. In my target paper I did gloss

over the transition from abstract to compositional constructions. A good example case is the processing of relative clauses. As suggested by Tomasello (2003) the operation of pattern finding functions can isolate phrasal structure (e.g. relativised noun phrases) such that these fragments can then be inserted as if they were nouns into existing constructions. This kind of process would explain the specific case of generalized compositionally for relative phrase structures, and provides an example for the more general case. Such statistically based pattern finding has been demonstrated in various studies of grammar induction and computational linguistics and remains to be applied in the construction grammar domain.

This said, however, I do not want to advocate the position that a structure mapping capability in the construction framework will be both necessary and sufficient for language and theory of mind. Clearly there will be functions that are required for these two capabilities that fall outside of the general construction structure mapping domain. Thus two goals are identifiable: First to specify and demonstrate the functionality of the construction mechanism, and second to do the same for those additional necessary functions required for language and ToM that are outside of the construction domain.

Acquiring Opaque Representations
Peter Ford Dominey
Mar 10, 2004 17:22 UT

Sperber asks how a system with a transparent representation capability (i.e. the ability to represent the state of affairs in which the semantic value of the representation is linked to the actual state of affairs) can acquire the ability to represent opaquely (i.e. to represent candidate states of affairs that are not directly linked with the actual state of affairs) and to interpret opaque representations.

Before asking how a system can interpret opaque representations, one should first ask under what conditions a system would itself generate or encounter opaque representations. In other words, what are the social and environmental pressures that would drive a system capable of manipulating transparent representations to become capable of manipulating opaque representations.

Consider the two statements "I see that it is raining" and "I think that it is raining". In the first case "it is raining" is used transparently, and in the second case it is used opaquely. Imagine an organism that saw cloudy skies last night, hears dripping, and is convinced that it is raining. That is, it represents (RAINING). Then the organism steps outside and observes that it is not raining.

In such a context I proposed that the system could compare memories of past representations with current representations. Sperber countered by saying that "An organism capable of attributing to itself past representations or to compare contents of different representations is already meta-representing contents, it already has a meta-representational ability". Good. Let me now describe how this can be achieved.

Specifically, based on the context dependant memory mechanism I described previously, we will see that the organism is capable of recalling its own previous (false) representation of the state of affairs and its current representation of the state of affairs, comparing the two, and detecting that one does not reflect the actual state of affairs.

The conflict or mismatch between predictions and state of affairs representations provides a pressure on the organism to exploit its context dependant memory mechanism in order handle the situation in which its representations turn out to NOT reflect the state of affairs. The system will likely have a performance advantage if it is capable of recognizing that its representations may or may not indeed reflect the actual state of affairs, and taking appropriate measures in such cases. This is the crucial step from transparent to opaque that Sperber is looking for, and this can be directly extended to embedded representations concerning the representational states of others. I am thus claiming that an organism that can make predictions (which can sometimes be wrong), and that has a context dependent memory and representation capability will be able to detect its own false predictions, thus encountering opaque representations. I hold that this provides the basis for opaque meta-representation as required by Sperber.

In detail, in the "raining" example: In the memory system of the organism, associated with the temporal context of the near past in the house, is the representation (RAINING). Associated with the current

temporal context is the representation (NOT-RAINING). The availability of these two representations can be explained by straight forward neurophysiology of context sensitive memory. But these are still transparent representations. The recognition of possible conflict between these representations is the source of opacity. That is, the organism will be confronted with situations in which its representations do not reflect the actual state of affairs. The special semantic resource: opacity will emerge as a correlate of the statistical fact that not all representations (e.g. those resulting from predictions, or from faulty perception) will be correct. This results in representations like "it is raining" that have lost their transparent relation to the state of affairs, but that can still be used appropriately, e.g. to explain behavior. Note that statistically, these representations will be more often linked with predictions than with actual observations or actions. This will lead to the separation of propositional attitude predicates (predict, believe, hope) vs. actual state predicates (see, hear, make-so).

Not there yet
Dan Sperber
Mar 12, 2004 9:58 UT

Fido the dog remembers having hidden a bone at the foot of the tree, expects the bone to be still there, digs and fails to find the bone. There is a mismatch between his prediction and his observation, and this leads him to acquire the ability to metarepresent. Or it should, since the conditions of Peter's scenario are fulfilled. Of course, we don't believe that Fido does anything of the sort. It is much more plausible that, if the present perception contradicts the previous predicition, then the prediction is erased. In a well-functionning cognitive system, novel evidence contradicting previous assumption must lead to updating of assumptions, but it need not lead to self-attribution of mental states. In fact, there is no clear route through which it would lead to self-attribution of mental states in a system that didn't have in the first place a metarepresentational format to achieve such attribution begin with.

For Peter's scenario to work, his human character would have to remember not just that it would rain, but that *she had thought that* it would rain, and then to confront this mistaken prediction to her present perception. But if she could remember that *she had thought that* it would rain, then she already had a metarepresentational ability. So, as before, what Peter is explaining is how an organism already endowed with metarepresentational abilities could come to do new things with it (attribute mental state to others in earlier postings, realize it had been mistaken in the last posting), not how an organism could become capable or metarepresenting in the first place.

Believe me, it's a tough one!

Statistical learning pervades cognition
Sergio Navega
Mar 3, 2004 13:25 UT

Peter Dominey's target article proposes an appealing vision relating the co-evolution of mechanisms able to process complex embedded structures and the emergence of high level constructs required by a Theory of Mind. Apart from a minor quibble related to the use of the predicate-argument format, I concur with Dominey's main claims. What follows is another path to reach similar conclusions.

Since Saffran et al. (1996) seminal study with 8-month-old infants, language learning has been put into a new perspective. Universal grammars and innate language modules are constructs that don't seem necessary to explain how children acquire linguistic expertise, although the whole issue is still under heavy discussion. Following Saffran's results, a number of studies provided similar accounts about the statistical learning on the visual domain (Fiser & Aslin 2002; Kirkham, Slemmer, Johnson 2002) and also auditory and touch (Conway & Christiansen 2002).

These statistical learning abilities have also been found to be present in nonhuman primates (Hauser, Newport & Aslin 2001; Conway & Christiansen 2001), which are important results to support a potential evolutionary path of these fundamental cognitive mechanisms.

However, one hypothesis that is rarely explored considers these statistical mechanisms acting not only on lower level signals, but also during the interpretation of high level behaviors. This idea may be supported by a study by Baldwin et al. (2001) that shows how infants can learn typical sequences of acts in an unsupervised way, reminding us of Saffran's results.

Baldwin et al. concludes that this ability to extract relevant "units of action" from continuous behavior is of great developmental importance. As a typical situation, they refer to the kind of learning that might happen when an infant observes cashier's actions during a supermarket checkout. At this point in an infant's mental development, he/she may not have the necessary sets of high level constructs such as intentions, beliefs or desires of others. However, even without these high level concepts, the infant could be thought to be using the sequences of learned actions as part of the necessary grounding for the later development of such concepts, in a way similar to the lexical and grammatical learning. Under a constructivist viewpoint, this grounding would be a prerequisite to the formation of a set of notions about intentions of others. Thus, it seems reasonable to entertain the hypothesis that part of the mechanisms responsible for language learning are also in use during the development of ToM. References

Baldwin, Dare A.; Baird, Jodie A.; Saylor, Megan M.; Clark, M. Angela (2001) Infants parse dynamic action. Child Development, May/June 2001, Vol 72, No. 3, 708-717.

Conway, C. M. and Christiansen, M. H. (2001). Sequential learning in non-human primates. Trends in Cognitive Sciences, 5(12):539--546.

Conway, C.M., & Christiansen, M.H. (2002). Sequential learning by touch, vision, and audition. Twenty-fourth Annual Meeting of the Cognitive Science Society.

Fiser, József; Aslin, Richard (2002) Statistical learning of new visual feature combinations by infants. PNAS Vol 99, No. 24, 15822-15826.

Hauser, Marc D.; Newport, Elissa L.; Aslin, Richard N. (2001). Segmentation of the speech stream in a nonhuman primate: Statistical learning in cotton-top tamarins. Cognition, 78, B53-B64.

Kirkham, Natasha Z.; Slemmer, Jonathan A.; Johnson, Scott P. (2002) Visual statistical learning in infancy: evidence for a domain general learning mechanism. Cognition 83 (2002) B35-B42.

Saffran, Jenny R.; Aslin, R. N; Newport, E. L. (1996) Statistical learning by 8-month-old infants. Science 274, 1926-1928.

A Unified Construction Approach Motor control, Theory of Mind, and Language

Peter Ford Dominey

Mar 8, 2004 14:52 UT

Sergio Navega's discussion provides an interesting point of departure for an even larger generalization. We can consider that from certain formal "generativist" perspective there are three “impossible” problems solved by the nervous system. Impossible in the sense that they are ill-defined, with too many degrees of freedom for generalized learning mechanisms, and thus require some heavy domain-specific machinery to render the problems solvable. The first two are language acquisition and acquisition of theory of mind, and the third for our purposes (though the first from an evolutionary sense) is the "inverse" problem of motor control. The problem here has to do with the fact that to move your hand from your keyboard to your cup of coffee, there are an infinite number of solutions, due to the extra-degrees of freedom in your shoulder, elbow, wrist and fingers. Thus there is an infinite set of possible trajectories that could get your hand from one place to another. Furthermore, for each of these trajectories there are an infinite number of velocity profiles that can be used, and the calculation of the inverse dynamics (the force effects of moving masses) is likewise unbounded for even such a simple movement. Thus, from this sad perspective we would never move, never have that sip of coffee. This inverse (kinematics and dynamics) problem rings familiar from the "poverty of the stimulus" perspective, and a viable solution appears similarly from the construction perspective. Indeed, data suggest that the nervous system chooses a workable solution, based on the calculation of motor commands from a kind of parameterized construction inventory, in which specific input-output mappings (constructions) are parameterized and combined to provide an acceptable level of extension and generalization, while allowing for learnability (see Kawato 1999).

In this context, the idea is that during sensorimotor exploration, the infant begins to develop an inventory of visual-transformation-movement constructions that link parameters of visual target orientation (e.g. Looking at a target in space - that cup of coffee) and the (initially randomly generated) motor commands that bring the hand to that region of space. These motor constructions will initially be similar to the holoconstructions in language and theory of mind, and with experience they become

compositional. It is interesting that this range of compositional motor control appears to have co-evolved with the development of cortical influence on the striatum in the parallel loops of cortex, basal ganglia and thalamus. We (Dominey et al. 1995, 1998, 2003) have demonstrated how this canonical circuit provides capabilities for aspects of motor control and language, and neurophysiological investigations reveal indeed that this canonical circuit is reduplicated throughout the neocortex and striatal complex providing an ensemble of motor, cognitive and emotional state related functional circuits. It will be of interest to determine if a more systematic analysis of this construction based learning capability can indeed demonstrate its proposed generalization across these three domains of motor control, theory of mind and language.

References:

Dominey PF, Hoen M, Blanc JM, Lelekov-Boissard T (2003) Neurological basis of language and sequential cognition: Evidence from Simulation, Aphasia and ERP Studies, Brain and Language, 86(2):207-25

Dominey PF, Lelekov T, Ventre-Dominey J, Jeannerod M (1998) Dissociable Processes for Learning the surface and abstract structure sensorimotor sequences. Journal of Cognitive Neuroscience, 10 :6 734-751

Dominey PF, Arbib MA, Joseph JP (1995) A Model of Cortico-Striatal Plasticity for Learning Oculomotor Associations and Sequences, J Cog Neuroscience, 7:3, 311-336

Kawato M (1999) Internal models for motor control and trajectory planning., Curr Opin Neurobiol. Dec;9(6):718-27.

The importance of interaction
Sergio Navega
Mar 9, 2004 20:43 UT

I am again in full agreement with Peter Dominey's reply. It was opportune to remind that generativist's arguments concerning the "impossible problems" dissipate on a closer analysis of how human cognition might have developed. It pays to remind generativists that if one proposes innate organs instead of learned linguistic abilities, this leaves open the bigger issue of how these innate organs could have evolved in the first place. In the case of human symbolic language, it seems untenable to sustain that those proposed "innate organs" evolved in such a reduced period, in the time scale of Homo Sapiens evolution. A much more plausible explanation seems to be the idea that language emerges spontaneously as the result of the activity of a community of sufficiently intelligent agents (Kirby 2000).

Dominey rightly mentions that the problem of obtaining good performance from a sensorimotor standpoint is similar to the "poverty of stimulus" issue. It is, likewise, difficult to understand how a learning mechanism is capable of presenting any learning at all given the immense number of degrees of freedom (motor controls) and dimensionality (sensory signals). On the side of nativists, we can agree that innate circuitries capable of solving these problems occur frequently in nature. A gazelle is able to walk a few hours after birth. However, it seems intriguing that we, humans, take several months to accomplish such a feat, which suggest that our native endowance in this area is quite restricted.

Perhaps here is a point where I could add some additional ideas to Dominey's remarks. It really seems useful to consider the workings of parametrized input-output constructions as one of the fundamental processes in the control of such complicated apparatuses as the human arm and hand, in its effort to grasp a cup of coffee. However, one issue that appears to be missing is how the learning of these abilities take place. Given the huge amount of possible variations (the dynamics of arm movements is altered if one is wearing a thick coat), it seems necessary the learning of invariant aspects of these controlled movements (which, to be coherent with my previous post, requires statistical learning abilities, although not only relative to "surface aspects"). These constructs (or sensorimotor schemas, to use another wording) must necessarily be the result of abstractions of several prior experiences. But they must also be seen as being learned by trial and error methods, and, in my view, even this is not enough.

A recently born infant isn't aware of his/her own arms, demanding months to learn the covariation of movement of hands in relation to the corresponding perceptual identification of these movements. More to the point, an infant appears to be learning, in these episodes, not only the correct motor schemas and parametrized input-output sequences, required to command the limbs and legs in a certain way, but also all the concurrent perceptual activity resulting from these movements. This task is not a simple one, because perception must process not only visual stimuli, but also proprioception and frequently tactile feedback. An important implication of this process is that the schemas that emerge from such learning must represent simultaneously both kinds of information: one of a perceptual nature and another of "motoric" nature. Also, these structures must be firmly associated one to another, otherwise one wouldn't know how to "check" if what one is doing is correct or not (tennis players often hit the ball without looking at it, using only proprioception as feedback). This binding between perception/action structures seems essential to the development of efficacious and controlled movements. However, not only for that.

What happens to be a reasonable next step is to consider these "bound schemas" as the grounding over which several high level constructs may be supported. This is where can lead us such publications as Rizzolatti and Arbib (1998). Another indication that children use abstractions of perception/action schemas to ground meaning can be found in David Bailey's Phd thesis (1997). [on a relatively off-topic matter, this is also one of the greatest challenges of purely symbolic AI: the lack of a body from which to derive meaning conveyed by sensorimotor schemas]. From here, we go to "intentionality detection", as is suggested by Rizzolatti et al. (2000), which again appears to support part of Dominey's initial thesis in his target article.

References

Bailey, David (1997) When push comes to shove. Phd Thesis, University of California, Berkeley.

Kirby, Simon (2000) Syntax without natural selection: how compositionality emerges from vocabulary in a population of learners. In: Knight, C.; Studdert-Kennedy, M., Hurford, James; The Evolutionary Emergence of Language. Cambridge University Press.

Rizzolatti, G., and M. A. Arbib. (1998) Language within our grasp. Trends in Neurosciences 21(5): 188-194, 1998

Rizzolatti G, Fogassi L, Gallese V (2000) Mirror neurons: Intentionality detectors? Int J Psychol 35: 205-205

Cognitive and Functional Factors in the Evolution of Grammar

Frederick J. Newmeyer (Washington University - Institut des Sciences Cognitives, Lyon)

(Date of publication: 15 March 2004)

Abstract: Both cognitive and functional factors have played important roles in the evolution of grammar. Human language originated with the linking of conceptual structures and the vocal output system. With the passage of time, the needs of communication came to play an ever-more important role in grammar. Human language today hence reflects the influence of both types of factors.

Introduction

Languages (and hence the grammars that characterize them) are used for two very different purposes. On the one hand, language plays a central role in cognition. Our thoughts are often (but not always) formulated in language and most linguists would agree that a central part of grammar is a level at which meanings are represented. On the other hand, language serves as the primary medium of communication among members of our species. While one might reasonably take the position that vocal communication involves, to a certain extent, the transfer of meaning from one person to another, that is not one of its necessary features. Indeed, it has been shown that a sizeable percentage of our utterances are primarily social in character, rather than involving an exchange of (propositional) information.

Now, clearly, a system 'ideally' designed for cognition is not necessarily a system that is 'ideally' designed for verbal interaction. On the one hand, the former admits to a complexity not found in the latter. That is, introspection tells us that we are capable of forming in our heads complex concepts that we could never have the ability (or, often, the desire) to communicate to another individual. On the other hand, verbal interaction is subject to constraints of real time that are not found in the simple representation of meaning.

The purpose of paper, then, is to puzzle out different contributions of cognition-aiding factors and vocal interaction-aiding factors in the origins and evolution of grammar. In keeping with standard practice among linguists — albeit a practice that has often created confusion among non-linguists — I will refer to the latter factors as 'functional' ones. This usage has arisen as a result of the belief of a large number of linguists that the primary 'function' of language is communication (rather than knowledge representation). Hence a 'functional explanation', for many linguists, is one that is based on enabling more efficient communication among speakers and hearers. In the remainder of this paper, then, a factor will be said to be 'cognitive' if it is based on representations of meaning or thoughts, and a factor will be said to be 'functional' if it is based on pressure for (more) efficient use of language among members of the species. I will argue that both played a role in evolution, but an imbalanced one. Natural language grammars at their onset reflected primarily cognitive factors. Functional pressure exerted itself only after the basic structure of grammar was in place. Furthermore, those factors based in cognition are reflected in the genetically-transmitted aspect of grammars (or 'Universal Grammar', as it is often called), while communicative factors are primarily historical and have become more and more manifest with the passage of time.

On the functional grounding of grammar

At first blush, the structure of language seems good for communication and at same time not particularly well designed for the representation of meaning or thought. There are several ways that functional factors seem to have shaped language. Perhaps the most important is derived from pressure to process sentences rapidly. For example, cross-linguistic generalizations about language structure reflect the fact that it is in the language user's interest to recognize the major grammatical elements of the sentence as rapidly as possible. One of the most longstanding typological generalizations in syntax (see Greenberg 1963) is that languages in which the verb precedes the object tend to have prepositions and that languages in which the verb follows the object tend to have postpositions (i.e., preposition-like elements that follow the noun that they are associated with). As Hawkins 1994 has shown, in languages in which this generalization holds, the length of time that it takes the hearer to identify all of the parts of the verb phrase is rather short, in that it is simply a function of the distance between the verb and the preposition. All that lies between these two elements is the object noun phrase. But in that minority of languages where the generalization linking verb-object and preposition-object order does not hold, the identification time is longer, since the object of the preposition as well as the object noun phrase lie between the verb and the preposition. The reasonable conclusion, then, is that the typological generalization about the correlation between verb-object order and adposition type reflects the preference of language users to process input rapidly.

Iconic motivation for grammatical structure is a theme in much functionalist writing. For our purposes this means that the form, length, complexity, or interrelationship of elements in a linguistic representation reflects the form, length, complexity or interrelationship of elements in the concept that that representation encodes. For example, it is well-known that syntactic units tend also to be conceptual units. In his classic study of the

effects of iconicity in syntax, Haiman 1985 points to a multitude of cases where grammatical distance and conceptual distance are correlated. Consider, for example, the following well-known pair of examples:

(1)

a. John caused Bill to die by inadvertently buying him a ticket on a plane that ended up crashing.

b. ?John killed Bill by inadvertently buying him a ticket on a plane that ended up crashing.

The oddness of (1b) results from the fact that causatives that are single words (e.g. 'kill') tend to convey a more direct causation than periphrastic causatives (e.g. 'cause to die'). So, where cause and result are formally separated, conceptual distance is greater than when they are not.

The third type of usage-based explanation appeals to the flow of information in discourse. Such explanations start from fact that language used to communicate and communication involves the conveying of information. Therefore, it is argued, the nature of information flow should leave and has left its mark on grammatical structure. Information flow has been appealed to, for example, in order to explain the ordering of the major elements within a clause. There are three ways to say: 'Daddy has brought a Christmas tree' in Russian, a typical so-called 'free word-order' language:

(2)

a. *Pápa prinyós yólku.* (Daddy — bought — a Christmas tree)

b. *Yólku prinyós pápa.* (A Christmas tree — bought — Daddy)

c. *Yólku pápa prinyós.* (A Christmas tree — Daddy — bought)

Each sentence is interpreted with the initial element representing old information and the final element representing new information. A common functionalist claim is that the discourse principle of Communicative Dynamism governs the ordering. The passage of time from past to present to future is mirrored iconically in discourse by the ordering of old information before new information (see, for example, Firbas 1987).

Not only do grammars in certain respects seem well designed functionally, but there are also a number of ways that they seem extremely poorly designed for the representation of meaning or thought. Hurford 2002 has catalogued some examples of such (seeming) poor design. Most importantly, since we do use language to communicate, grammatical structure necessarily has phonological properties as well as syntactic ones. Phonology is necessary for the expression of ideas, but not their formulation. Likewise, language is replete with morphological complexity, which serves no obvious cognitive function. For example, in English a concept can be represented by a stand-alone word ('book'), by a prefix ('un-able'), or by a suffix ('king-dom'); indeed a single word can contain all three ('under-talent-ed'). If language evolved solely in the service of cognition, it is hard to imagine why it would manifest morphological complexity.

Hurford points to several other ways that grammars make distinctions that serve no evident cognitive function. For example, virtually all languages distinguish in their grammars between the grammatical relation 'Subject' and the grammatical relation 'Direct Object'. These grammatical notions conflate a relatively large number of semantic ones. Hurford also notes that all languages contain what he calls 'quirky mismatches', that is, they express parallel semantic notions in quite different grammatical ways. Hence past and present tense in English are encoded by suffixes, but future tense by a modal verb. Along the same lines, markers of clause boundaries such as complementizers serve no obvious role in cognition, nor does the 'displacement' of an element from another element with which it is associated semantically (e. g., in 'Who did you see?', 'who' is displaced from the verb 'see', of which it is the direct object). Finally, Hurford points to the phenomenon of grammatical agreement as one which serves no cognitive function. So in Spanish, the article and the adjective must agree in gender and number with the noun that they modify — a puzzling fact from the perspective of a cognitive account of the origins of grammar.

Grammar as a reflection of cognitive processes

The core of my argument for the importance of cognitive factors is based on the many ways that grammars seem overdesigned for efficient communication and at the same time well designed from the standpoint of cognition. First, consider the importance of full argument structure. An interesting fact about actual utterances produced by language users is that they rarely contain a subject, a verb, and an object, where the subject and the object (the 'arguments' of the sentence) are full non-pronominal words. Most utterances consist of a verb with one full argument, which is either the subject of an intransitive verb or the object of a transitive verb. Other arguments are either reduced to pronomimal or affix status or omitted entirely (the latter in languages like Spanish and Chinese which allow the wholesale omission of arguments). One's first thought might be that this fact presents another piece of evidence against the idea of the origin of grammar in cognition. However, when one looks at things more closely, we see that grammars are in fact 'propositional', that is they consist of structures specified by formal rules that take the sentence to be the basic unit of grammar, where sentences are in a rough mapping with propositions, verbs with predicates, and noun phrases with logical arguments. Evidence for this claim is provided both by the fact that the

process of speech production involves calling upon the full argument structure of the sentence (Levelt 1989) and by the fact that sentence fragments can be interpreted only by reference to full grammatical structure (Newmeyer 2003). In other words, whatever one might do in actual speech, one's cognitive representation embodies all the arguments of the sentence. These are ingredients of cognition, not communication.

Second, every language on earth allows for the possibility of recursion, that is, sentences embedded inside of sentences inside of sentences, ad infinitum. For example, in principle, there is no limit to the number of times that another subordinate clause can be added in sentences like the following:

(3) Mary thought that John said that Sue insisted that Paul believed that …

Is recursion necessary for communication? Apparently, it is not. We virtually never have any reason to utter complex sentences like (3). And the desired message conveyed by sentences with recursion like (4a) can easily be communicated by a sentence like (4b), employing juxtaposition of two clauses:

(4)

a. Mary thought that John would leave.
b. Here is what Mary thought. John was going to leave.

Why does human language have recursive properties? The obvious answer is that human thought has recursive properties.

Third, human languages are horribly designed for communication from the point of view of the amount of ambiguity that they allow. Virtually any sentence imaginable is loaded with potential ambiguity. Of course, we deal with this problem in actual language use by means of complex systems of inference and implicature, conveyed meanings, and so on. Hence, in actual conversation, real ambiguity is normally a minor problem. But our concern here is whether languages are well shaped for communicative purposes. Based on the ambiguity that they permit, the conclusion has to be that they are not well shaped. If we focus on cognitive representations instead of on communication, however, structural ambiguity is a much less serious problem. The reason is that many (communicatively) ambiguous sentences are disambiguated by their structures. From the point of view of language use, the possibility of dual (i.e. ambiguous) representations for the same sequence of words is not communicatively desirable. But since the different meanings are represented differently from the cognitive standpoint, we must conclude that in this respect grammars seem well adapted to cognition.

Fourth, a central fact about language that it allows us say anything that we can conceptualize, regardless of whether we would actually have any need, desire, or likelihood to convey the information conceptualized. No more effort is required to say an obviously false sentence or an obviously true one than one that might be genuinely communicatively relevant. Along the same lines, we have no more trouble uttering pure nonsense sentences than grammatically parallel ones that contain an easily accessible semantic content. In other words, communicatively useless sentences provide another example of how language is 'overdesigned' for communication.

Fifth, grammatical categories tend to have a closer relation to cognitive categories than to communicative ones. As we have seen, the mental representation of transitive sentences contains a subject, a verb, and an object, even though actual utterances are pared down considerably. The classical definition of parts of speech in terms of meaning ('a noun is a person, place, thing, or idea') is close to correct. Units of word formation (morphemes) are almost alway definable semantically. On the other hand, communicative categories such as 'topic' and 'focus' are less likely to be marked in languages by a special category than semantically-defined categories are. In other words, the properties of the basic grammatical building blocks of a sentence suggest that in an important respect, grammar is better designed for cognition than for communication.

Finally (and more controversially), covert levels of grammatical structure represent aspects of meaning. In the classic Government-Binding model of grammar, the covert level of D-structure is the pure representation of predicate-argument structure and the covert level of Logical Form represents quantification structure, that is, the relations between quantifiers and the variables that they bind. If this picture is right, it suggests that cognitive relations are at the heart of grammar in a way that communicative aspects are not.

Cognition, functional pressure, and the origins of language

Where are we now? On the one hand, we have seen six important design features of language that pertain little — if at all — to communication. Importantly, none of them are 'learnable' in the ordinary sense of the word, suggesting that they were there from the dawn of human language itself. On the other hand, those aspects of language that seem designed to better aid communication are historical in nature. That is, unlike

predicate-argument structure and so on, we can see how they developed over time. For example, take a communicative aspect of language par excellence, namely discourse markers. These are expressions like the following:

(5) then, I mean, y'know, like, indeed, actually, in fact, well, …

Even though they are essential to the makings of a coherent discourse, they invariably arise historically from something else, most typically out of conceptual meanings and uses constrained to the argument structure of the clause. This fact is not surprsing if if vocal communication itself is derivative. Nouns and verbs trace back to nouns and verbs, because they were there from the start. The derivative nature of discourse markers points to a time when we had structured conceptual representations, but they had not yet been coopted for communication.

There is another reason to posit that cognition left its mark on language before communication. We have learned that the conceptual abilities of the higher apes are surprisingly sophisticated. However, their communicative abilities are remarkably primitive. There is very little calling on their conceptual structures in communicative settings. These facts suggest a three-stage process in language evolution. First, there was the inherited level of conceptual structure. Secondly, the level became linked to the vocal output channel, creating for the first time a grammar that was independent of the combinatorial possibilities of conceptual structure per se and making possible the conveying of thought — in other words, communication. And once grammars started to be drawn upon for real-time purposes, the constraints of real-time use begin to affect their properties.

Conclusion

This paper began by raising the question of the relative roles of cognitive and functional factors in the evolution of grammar. The conclusion is that they have both played important roles, though unbalanced ones. Human language was jump-started by the linking of conceptual structures and the vocal output system. In other words, cognitive factors were the first to shape grammars. But with the passage of time, the exigencies of communication came to play an ever-more important role in grammar. Human language today therefore reflects the influence of both types of factors.

References

Firbas, Jan 1987. "On the operation of communicative dynamism in functional sentence perspective." Leuvense Bijdragen 76: 289-304.
Greenberg, Joseph H. 1963. "Some universals of language with special reference to the order of meaningful elements." In Greenberg, Joseph, Ed. *Universals of language*, 73-113. Cambridge, MA: MIT Press.
Haiman, John, Ed. 1985. Iconicity in syntax. Typological Studies in Language. Amsterdam: John Benjamins.
Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.
Hurford, James R. 2002. "The roles of expression and representation in language evolution." In Wray, Alison, (ed.) *The transition to language*, 311-334. Oxford: Oxford University Press.
Levelt, Willem J. M. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
Newmeyer, Frederick J. 2003. "Grammar is grammar and usage is usage." *Language* 79: 682-707.
Traugott, Elizabeth C. and Richard B. Dasher 2002. *Regularity in semantic change.* Cambridge: Cambridge University Press.

Discussion

Three stages process in language evolution?
Gloria Origgi
17 mars 2004 9:58 UT

Frederick Newmeyer concludes his very stimulating paper with a three-stage picture of language evolution. First, the genetically determined conceptual structure. Second, the link between conceptual structure and vocal output. Third, the stabilization of the constraints on grammar due to the real-time use. I find his picture incomplete. Newmeyer argues through the paper for a massive underdetermination of language for communicative purposes. But his final evolutionary picture doesn't explain how language has become, in spite of that, a reliable and robust mean of communication.

And this leads us back to the main subject of this conference: How the communicative use of language has been influenced in evolution by other cognitive complex competences such as inferential pragmatic processes? What is the role of these processes in the externalization of grammar? Did the coevolution of language and other interpretive abilities influence the structure of public language?

In particular, how can you explain the transition from stage one to stage two : A mutation in the vocal tract that allows people to express their conceptual thoughts would be inert in a population lacking rich language-independent interpretive capacities.

A reply to Gloria Origgi
Frederick Newmeyer
18 mars 2004 10:51 UT

I must agree with Gloria Origgi that my account of the evolution of language is 'incomplete'. My (admittedly narrow) focus was on the transition from a language-ready brain to a speech-ready brain. A central conclusion was that grammatical changes supporting the communicative use of language seem to be 'historical', rather than 'evolutionary' developments. As a syntactician (rather than a specialist in pragmatics or discourse), I am more competent to speculate on the evolution of grammar rather than on that of inferential pragmatic processes. Origgi implies that in order for language to be useable, inferential-interpretive capacities would have to have been largely in place at the time of the mutation linking conceptual structures to the vocal tract. Since these capacities do not depend on the existence of spoken language, it does not seem problematic to conclude that they pre-existed the transition to full human language. But I leave it to others to discuss this issue in detail.

language design and communication principles
Jacques Moeschler
18 mars 2004 11:18 UT

I would like to begin my entry into this discussion by starting my complete agreement with one of the conclusion of Frederick's paper: "cognitive relations are at the heart of grammar in a way that comunicative aspect are not". Now if language is not designed as a perfect communicative tool, some non-linguistic factors should help speakers in communication to convey their intentions. Pragmatics traditionally describes such principles as cooperation, relevance, quantity or informativity. The question I would like to ask is the following: to what extent can we say that such principles in comunication describe any aspect of the social use of language? I recall one of Frederick's claims: "it has been shown that a sizeable percentage of our utterances are primarily social in character, rather than involving an exchange of (propositional) information". It seem to me that we meet a paradox here: language design can be (partially) motivated by cognitive factors, and if one adds parsing to other principles of language use, then the latter seem devoted to just a small part of communicative process, that is, information retrieval.

Does history improve communication ?
Viviane Deprez
19 mars 2004 17:27 UT

Although I am a priori sympathetic to Newmeyer's stimulating attempt to reconcile the functional and cognitive aspects of grammar, there are some emerging consequences of his suggestions that leave me unconvinced. In particular, underlying Newmeyer's suggestion that functional aspects of language emerge with history lurks a teleological perspective on diachrony that views language change as language 'improvement'. Studies of language change, however, have not so far supported such a view. To take a simple example, it is well known that in the course of the development of Romance languages, the word order of Latin, which was predominantly SOV changed to SVO. Now given that both grammars conform to homogeneous languages on Greenbergian criteria, the idea of historical improvement is surely not compelling here. Even more disturbing from the point of view of the language learner, who after all never comes to language with a diachronic lens, is the question of how at a given time when faced with ambiguous cues regarding word order, she came to chose the new SVO order rather than simply revert to the previous SOV one, both of them being a priori equally well adapted for efficient communication. Such examples seem to support a view of language change as 'a random walk through the parameters of UG' to borrow Hornstein's formula, not the view that history improves language. His discussion of the differences between discourse connectors and nouns and verbs, Newmeyer indirectly alludes to works such as that of Heine and Kuteva (2002) whose essential claim is that all words historically derive from either nouns or verbs through grammaticalization. On such views as well, history is teleological and assumed to strongly affect the nature of language since at the onset, language must have consisted solely of nouns and verbs and fail to manifest any of the functional words seeming so central to 'grammar'. But one central question that such works fail to raise is what kind of notions are in fact grammatical categories such as nouns and verbs? One possible view (Marantz's 1993) (Borer 2003) is that grammatical categories are nothing but one acceptable mapping between an encyclopedic notion (however defined) and a syntactic structure. That is, a root is not a priori a noun or a verb, it becomes that when it is mapped onto a syntactic structure. E.g [a [cook]] vs [to [cook]]. All words are grammaticalized in the sense that they are mapped onto a syntactic structure (here understood as

functional architecture) that contributes in defining their ultimate interpretation. On such a view, the observation that ‘all words derive from nouns and verbs’ is then a mere reflection of the tendency for word creation to be is rooted in world knowledge. Grammaticalization does not ‘create’ or 'evolve' grammar. It merely affects the mapping of a root onto an existing syntactic structure. And history again is not teleological but merely recycles already used material.

Borer, Haggit. 2003. Structuring Sense. Oxford University Press

Heine, Bernd and Tania Kuteva. 2002. On the evolution of grammatical forms. The transition to language, ed. by Alison Wray, 376-97. Oxford: Oxford University Press.

Marantz, Alec. (1997) No escape from syntax. Don’t try grammatical categories in the privacy of your own lexicon. University of Pennsylvania working papers in Linguistics. A. Dimitriadis.& I. Siegel.

A reply to Deprez on language change and language improvement
Frederick Newmeyer
22 mars 2004 9:50 UT

I certainly did not mean to imply that I view 'language change as language improvement' and have in fact rejected such an idea in several publications (e. g. Newmeyer 2001). If it were, as Deprez observes, we would note a (nonexistent) overall directionality to language change. While certain LOCAL changes may indeed represent improvements in user-efficiency (eg those that serve to harmonize branching direction), such changes are counterbalanced by the many counter-functional changes that result from borrowing and by the fact that a functionally-motivated change can have some incidentally dysfunctional consequences (for nice examples, see Lightfoot 1999). Furthermore, as noted originally by Andre Martinet, if not earlier, even functionally-motivated changes can conflict with each other – another fact that prevents an overall directionality to change.

We have records of human language going back about 4000 years. But presumably humans have had language for 100,000 years or more. If my idea that language originally served cognition and was later coopted for communication is correct, then at the outset language would have to have undergone a rapid transition to make it useable. These communicative pressures were at work long before the existence of any language for which we have records, but have nevertheless left their mark on those languages which are spoken today.

Hornstein's idea, however, that language change is 'a random walk through the parameters of UG' seems immediately falsified by the fact that the typological distribution of grammatical features is NOT random. Whatever parameters combine to yield OSV order are much less frequently chosen than those that combine to yield SOV order; if parameter settings were random then we would expect as many languages to have the parameter-setting combinations that yield preposition-stranding as those that do not; etc. Given that the walk through parameters is anything but random, we have two choices really: to attribute the lack of randomness to the (unexplained) quirks of an innate UG or to attribute the lack of randomness to (more-or-less) well understood facts about language users, such as their interest in processing sentences rapidly (see Newmeyer 1998).

Lightfoot, David. 1999. The development of language: Acquisition, change, and evolution. Oxford: Blackwell.

Newmeyer, Frederick. 1998. Language form and language function. Cambridge, MA: MIT Press.

Newmeyer, Frederick. 2001. Where is functional explanation? Proceedings from the parasessions of the Thirty-Seventh Meeting of the Chicago Linguistic Society (Volume 37-2), pp. 99-122.

Communicative pressure and UG
Viviane Deprez
22 mars 2004 21:54 UT

It appears I have misunderstood: history does not improve the communicative aspects of language. But what I now fail to understand is how can one have one’s cake and eat it too?

Quoting Newmeyer:

1.Language change is NOT random Evidence:Some types of languages are rare, other common, OSV vs SVO

2.Lack of randomness must be attributed either to Innate UG Communicative/parsing pressure shaping language

The former looks unlikely --> the latter must be the case

3. Question: What should the rare types be attributed to? Communicative / parsing pressure shaping language Innate UG

The former looks unlikely/contradictory --> the latter must be the case

Ergo: Universal Grammar is left to account for the exceptions???

Possible and probable languages
Frederick Newmeyer
23 mars 2004 12:28 UT

Déprez raises the question of the role of (innate) Universal Grammar (UG) in my scenario. In a nutshell, UG characterizes the class of 'possible languages', but not the class of 'probable languages'. That is, UG specifies the general form of grammars, the interrelationship of components, what a possible grammatical constraint, rule, etc. is. But why some UG-permitted grammars are typologically more common than others is explained by recourse to performance (mainly parsing) principles. So there is nothing in UG per se that explains why SOV order is more common than OSV order, why preposition-stranding is rare, etc.

About the adaptiveness of syntactic recursion (1)
Jean-Louis Dessalles
21 mars 2004 16:20 UT

In his stimulating paper, Frederick Newmeyer suggests that syntactic abilities evolved to serve cognitive purposes that are decorrelated from any communicative function. He identifies six features of grammar that he finds hard to explain if, as commonly supposed, syntactic abilities were designed to make communication optimally efficient. Among them, the ability to form recursive structures through language is claimed to be motivated by internal cognitive efficiency only.

The facts mentioned in Newmeyer's paper are disturbing and owe precise examination. The author's global position, if accepted, would undermine several efforts to reduce the paradox of language emergence during human evolution: since Monod (1970), several authors considered the hypothesis that human-specific thinking processes were a consequence of communication (e.g. Bickerton 1995). If Newmeyer is correct, we must account for two evolutionary unlikely events instead of one, and explain (1) why human ancestors did evolve enhanced cognitive abilities (while other primates did not); (2) why human beings did find a sudden advantage in communicating their enhanced thoughts to genetically unrelated conspecifics.

Since much is at stake here for those who want to understand what brought our species to existence, Newmeyer's arguments about the maladaptiveness of syntax for communication are worth studying. For the sake of concision, we will only examine the case of recursion.

A communicative function for recursion

The ability to form embedded phrases is a universal feature of human languages. Since there is always an alternative way to express the same content, Newmeyer concludes that recursion is a property of our mind rather than a requirement of communicative efficiency. We suggest the exact opposite. As analysed in (Dessalles 2000), syntactic embedding is used to connect predicates through variable sharing. The following sentence:

(1) Mary brought the book about Iran that her brother gave to her. can be glossed as:

(2) bring(Mary, x) & book(x) & subject(x,Iran) & give(y,x,Mary) & brother(y,Mary)

We may think of some extinct hominid species that evolved the ability to express thoughts that way, through simple unembedded clauses connected through variable sharing, as in (2) which computer scientists may regard as a Prolog program. This is not the evolutionary pathway that our lineage followed. We use a *semantic linking principle* that forces two syntactically connected phrases to share one variable. This avoids the burden of having to repeat variable names (Mary brought a thing; that thing is a book; someone gave that thing to Mary; etc.) and it makes the code more compact.

Recursion has a function: it gives a new role to predicates. The main predicate in a sentence expresses a thought for argumentative purposes. In (1), the good news is expressed through the primary predicate bring(Mary,x). The main predicate is what is really meant, what is offered to the addressees' critique (in the case of argumentation) or to their appraisal (in the case of event report). Thanks to recursion, other predicates can be introduced to *determine* arguments. These satellite predicates are expressed in (1) through the words book, about and give. They help addressees determine what x refers to in the scene. The satellite predicate expressed through brother helps them determine a supplementary argument introduced by give.

This function of recursion is purely communicative! There is no point for a human individual to determine predicate arguments for herself. When you have a thought, it is fully instantiated. Only when you want to make it public for other individuals, is the apparatus of recursion necessary. The author of utterance (2) wants to express an instantaneous thought, the good news that Mary brought that object he has in mind. He can't merely say "Mary brought it", since addressees won't be able to reconstruct what "it" refers to. In order to make his utterance intelligible to the audience, he strings satellite predicates in a recursive way until equation (2) is thought by him to be solvable by interlocutors.

If what precedes is correct, then a central feature of syntax, the ability to embed phrases in a recursive way, appears to be a communicative adaptation rather than a purely internal device.

[Continued on a second posting] (.../...)

On adaptiveness and recursion
Frederick Newmeyer
22 mars 2004 10:04 UT

While I don't deny that recursion serves a communicative role, I doubt that it is a very important one. I most certainly doubt that it was important enough to shape the use of language for communicative purposes. Sentences like Dessalles' 'Mary bought the book about Iran that her brother gave her' are rarely uttered in everyday conversation. Indeed, everyday communication makes use of surprisingly little recursion. Many linguists have suggested that the use of subordinate clauses increases dramatically with literacy. The major study along these lines is Kalmár 1985, which maintains that Samoyed, Bushman, Seneca, and various Australian languages rarely employ subordination. According to Kalmár:

It is quite likely that the number of subordinate clause types grew as narrative developed and accelerated with the advent of writing. Typical is the development of subordination in Greek, which hardly existed in Homer but was well developed in the classics (Goodwin 1912). (Kalmár 1985: 159)

Mithun 1984 carried out text counts on a number of languages with respect to the amount of subordination that one finds in discourses carried out in those language. All languages manifest some subordination, but there is a strong correlation between its rare use and the pre-literate status of their speakers. But the fact that every language allows the POSSIBILITY of recursion suggests that it is a genuine design feature of language, there from the beginning. Why would this be the case? The obvious answer is that human thought has recursive properties, even if the manifestation of the expression of that thought in communication does not necessarily draw on those properties.

Kalmar, Ivan. 1985. Are there really no primitive languages? In Literacy, language, and learning : The nature and consequences of reading and writing. Cambridge, Cambridge University Press.

Mithun, Marianne. 1984. How to avoid subordination. Berkeley Linguistics Society 10, 493-523.

About the adaptiveness of syntactic recursion (2)
Jean-Louis Dessalles
21 mars 2004 16:23 UT

(.../...)

Recursion and cognition

Some authors claim that the ability to form genuine (i.e. centrally embedded) recursive structures emerged by chance in our lineage, and that this new ability found a fortuitous use in the way we structure our linguistic utterances (Chomsky 1975; Hauser, Chomsky & Fitch 2002). Newmeyer's position undoubtedly opposes this extreme claim, but it still considers that the ability to embed linguistic structures is not motivated by communicative efficiency. If this were the case, then we could rightfully wonder why the tree-like branching of phrases happens to be a nice engineering solution to the problem of argument determination.

Our point is not to claim that recursion isn't a cognitive property. We do hold predicates in a mental stack until their arguments are considered sufficiently constrained for interlocutors. This cognitive device not only drives the generation of embedded phrases, but also generates larger portions of discourse, where sentences are connected to each other through anaphoric links. Our point is that even the very existence of this mental recursive ability, which is responsible for actual linguistic recursion, is only motivated by communicative requirements. The reason for this, again, is that predicate arguments are always internally instantiated when a thought is experienced. The various cognitive devices used to determine predicate arguments evolved for the purpose of making internal thoughts somewhat available to others. Linguistic recursion is one of these devices.

An unavoidable consequence is that human syntactic abilities are elements of an interface. This interface achieves the great task of establishing a link between separate minds. We suggested elsewhere that the human-specific part of our cognition, namely the ability to form genuine predicates (Dessalles & Ghadakpour 2003), was itself designed by natural selection to serve communication (Dessalles 2000).

Conclusion

The hydrodynamic efficiency of dolphins is a marvel of nature in the eye of engineers. Similarly, engineers dealing with automatic natural language processing acknowledge the complexity and efficiency achieved by human language. Despite the alleged maladaptiveness of grammar (Lightfoot 2000), no convincing suggestion was yet been made concerning syntactic universal aptitudes that could unequivocally improve their efficiency. Newmeyer's observation that syntactic structures are often ambiguous may be a problem, but as the author says, the vast majority of these ambiguities disappear in context. Moreover, their systematic suppression may have bad consequences on code compactness.

Previous papers in this conference considered the possibility that advanced cognitive abilities such as the ability to read others' mind could have granted conceptual powers to our ancestors, who subsequently began to talk to each other. Newmeyer's observations, by claiming that communication is of subordinate importance in determining syntactic abilities, comforts the idea that human cognition was shaped independently of language. By insisting on the essential role that syntactic recursion plays for making human communication possible, we want to suggest that the sequence of evolutionary events may have been quite different.

BIBLIOGRAPHY

Bickerton, D. (1995). Language and human behavior. London : UCL Press.

Chomsky, N. (1975). Réflexions sur le langage. Paris : Flammarion.

Dessalles, J-L. (2000). Aux origines du langage - Une histoire naturelle de la parole. Paris : Hermès. http://www.enst.fr/~jld/papiers/pap.evol/99111703.html

Dessalles, J-L. & Ghadakpour, L. (2003). "Object recognition is not predication - Commentary on Jim Hurford: 'The neural basis of predicate-argument structure'". Behavioral and Brain Sciences, 26(3), 290-291. http://www.ling.ed.ac.uk/~jim/BBSNEURO/dessalles.html

Hauser, M. D., Chomsky, N. & Fitch, W. T. (2002). "The faculty of language: what is it, who has it, and how did it evolve ?". Science, 298, 1569-1579.

Lightfoot, D. (2000). "The spandrels of the linguistic genotype". In C. Knight, M. Studdert-Kennedy & James Hurford (Eds), The evolutionary emergence of language: social function and the origins of linguistic form. Cambridge : Cambridge University Press.

Monod, J. (1970). Le hasard et la nécessité. Paris : Seuil.

Do sex differences in empathy account for sex differences in language acquisition?

Simon Baron-Cohen (Cambridge University)

(Date of publication: 29 March 2004)

Abstract: A new model of sex differences in the mind suggests that females, on average, show a stronger drive to empathize. Empathy is broader than 'theory of mind' because it not only involves identifying the mental states of the other person, but also responding to these with an appropriate emotion. In this paper, I discuss how the well-established finding that girls develop language earlier than boys might be a consequence of their better empathy. The argument hinges on the notion that picking up word meanings is facilitated if one can identify a speaker's intentions.

In this paper I want to go beyond 'theory of mind' (ToM) to talk about the broader capacity for empathy. I then review the evidence for sex differences in empathy (females developing faster, and showing a stronger drive in this, than males). Finally, I raise the possibility that since language acquisition requires not just decoding heard words in a look-up table but identifying the speaker's intended meanings (i.e., the speaker's mental states), a female advantage in empathy could have driven the female advantage in rate of language acquisition.

The term 'empathizing' encompasses the following earlier terms: 'theory of mind', 'mind-reading', and taking the 'intentional stance' (Dennett, 1987). Empathizing involves two major elements: (a) the ability to attribute mental states to oneself and others, as a natural way to understand agents (Baron-Cohen, 1994a; Leslie, 1995; Premack, 1990); (b) having an emotional reaction that is appropriate to the other person's mental state. In this sense, it includes what is normally meant by the term 'theory of mind' (the attributional component) but it goes beyond this, to also include having some affective reaction (such as sympathy).

The first of these, the mental state attribution component, has been widely discussed in terms of being an evolved ability, given that in the universe can be broadly divided into two kinds of entities: those that possess intentionality and those that do not (Brentano, 1970). The mental state attribution component is effectively judging if this is the sort of entity that might possess intentionality. Intentionality is defined as the capacity of something to refer or point to things other than itself. A rock cannot point to anything. It just is. In contrast, a mouse can 'look' at a piece of cheese, it can 'want' the piece of cheese, and it can 'think' that this is a piece of cheese, etc. Essentially, agents have intentionality, whereas non-agents do not.

This means that when we observe agents and non-agents move, we construe their motion as having different causes (Csibra, Gergely, Biro, Koos, & Brockbanck, 1999; Gelman & Hirschfield, 1994). Agents can move by self-propulsion, which we naturally interpret as driven by their goals and desires, whilst non-agents can reliably be expected not to move unless acted upon by another object (e.g., following a collision). Note that mental state attribution is quite broad, since it includes not just attribution of beliefs, desires, intentions, thoughts and knowledge, but also perceptual or attentional states, and all of the emotions (Baron-Cohen, Wheelwright, Hill, & Golan, submitted; Griffin & Baron-Cohen, 2002).

The second of these, the affective reaction component, is closer to what we ordinarily refer to with the English word 'empathy'. Thus, we not only attribute a mental state to the agent in front of us (e.g., the man 'thinks' the cake is made of soft, creamy chocolate'), but we also anticipate his or her emotional state (the man will be disappointed when he bites into it and discovers it is hard and stale), and we react to his or her emotional state with an appropriate emotion ourselves (we feel sorry for him). Empathizing thus essentially allows us to make sense of the behaviour of another agent we are observing, predict what they might do next, and how they might feel. And it allows us to feel connected to another agent's experience, and respond appropriately to them.

The normal development of empathizing

Empathizing develops from human infancy (Johnson, 2000). In the infancy period, it includes:

- being able to judge if something is an agent or not (Premack, 1990);
- being able to judge if another agent is looking at you or not (Baron-Cohen, 1994b);
- being able to judge if an agent is expressing a basic emotion (Ekman, 1992), and if so, what type;
- engaging in shared attention, for example by following gaze or pointing gestures (Mundy & Crowson, 1997; Scaife & Bruner, 1975; Tomasello, 1988);
- showing concern or basic empathy at another's distress, or responding appropriately to another's basic emotional state (Yirmiya, Sigman, Kasari, & Mundy, 1992);
- being able to judge an agent's goal or basic intention (Premack, 1990).

Empathizing can be identified and studied from at least 12 months of age (Baron-Cohen, 1994a; Premack, 1990). Thus, infants show dishabituation to actions of 'agents' who appear to violate goal-directedness (Gergely, Nadasdy, Gergely, & Biro, 1995; Rochat, Morgan, & Carpenter, 1997). They also expect agents to 'emote' (express emotion), and expect this to be consistent across modalities (between face and voice) (Walker, 1982). They are also highly sensitive to where another person is looking, and by 14 months will strive to establish joint attention (Butterworth, 1991; Hood, Willen, & Driver, 1997; Scaife & Bruner, 1975). By 14 months they also start to produce and understand pretence (Bates, Benigni, Bretherton, Camaioni, & Volterra, 1979; Leslie, 1987). By 18 months they begin to show concern at the distress of others (Yirmiya et al., 1992). By 2 years old they begin to use mental state words in their speech (Wellman & Bartsch, 1988).

Empathizing of course develops beyond early childhood, and continues to develop throughout the lifespan. These later developments include:

- attribution of the range of mental states to oneself and others, including pretence, deception, belief (Leslie & Keeble, 1987);
- recognizing and responding appropriately to complex emotions, not just basic ones (Harris, Johnson, Hutton, Andrews, & Cooke, 1989);
- linking mental states to action, including language, and therefore understanding and producing pragmatically appropriate language (Tager-Flusberg, 1993);
- making sense of others' behaviour, and predicting it, and even manipulating it (Whiten, 1991);
- judging what is appropriate in different social contexts, based on what others will think of our own behaviour;
- communicating an empathic understanding of another mind.

Thus, by 3 years old, children can understand relationships between mental states such as seeing leads to knowing (Pratt & Bryant, 1990). By 4 years old they can understand that people can hold false beliefs (Wimmer & Perner, 1983). By 5-6 years old they can understand that people can hold beliefs about beliefs (Perner & Wimmer, 1985). By 7 years old they begin to understand what not to say, to avoid offending others (Baron-Cohen, O'Riordan, Jones, Stone, & Plaisted, 1999). With age, mental state attribution becomes increasingly more complex (Baron-Cohen, Joliffe, Mortimore, & Robertson, 1997; Happe, 1993). The little cross-cultural evidence that exists suggests a similar picture in very different cultures (Avis & Harris, 1991).

These developmental data have been interpreted in terms of an innate module being part of the infant cognitive architecture. This has been dubbed a theory of mind mechanism (ToMM) (Leslie, 1995). But as we have suggested, empathizing also encompasses the skills that are needed for normal reciprocal social relationships (including intimate ones) and in sensitive communication. Empathizing is a narrowly defined domain, namely, understanding and responding to people's minds.

The female brain: empathizing

What is the evidence for female superiority in empathizing? In the studies summarised here, sex differences of a small but statistically significant magnitude have been found.

1. *Sharing and turn-taking*. On average, girls show more concern for fairness, whilst boys share less. In one study, boys showed fifty times more competition, whilst girls showed twenty times more turn-taking (Charlesworth & Dzur, 1987).
2. *Rough and tumble play or 'rough housing' (wrestling, mock fighting, etc)*. Boys show more of this than girls do. Although there's a playful component, it can hurt or be intrusive, so it needs lower empathizing to carry it out (Maccoby, 1999).
3. *Responding empathically to the distress of other people*. Girls from 1 yr old show greater concern through more sad looks, sympathetic vocalizations and comforting. More women than men also report frequently sharing the emotional distress of their friends. Women also show more comforting, even of strangers, than men do (Hoffman, 1977).
4. *Using a 'theory of mind'*. By 3 years old, little girls are already ahead of boys in their ability to infer what people might be thinking or intending (Happe, 1995). This sex difference appears in some but not all studies (Charman, Ruffman, & Clements, 2002).
5. *Sensitivity to facial expressions*. Women are better at decoding non-verbal communication, picking up subtle nuances from tone of voice or facial expression, or judging a person's character (Hall, 1978).
6. *Questionnaires measuring empathy*. Many of these find that women score higher than men (Davis, 1994).

7. *Values in relationships*. More women value the development of altruistic, reciprocal relationships, which by definition require empathizing. In contrast, more men value power, politics, and competition (Ahlgren & Johnson, 1979). Girls are more likely to endorse co-operative items on a questionnaire and to rate the establishment of intimacy as more important than the establishment of dominance. Boys are more likely than girls to endorse competitive items and to rate social status as more important than intimacy (Knight, Fabes, & Higgins, 1989).
8. Disorders of empathy (such as psychopathic personality disorder, or conduct disorder) are far more common among males (Blair, 1995; Dodge, 1980).
9. Aggression, even in normal quantities, can only occur with reduced empathizing. Here again, there is a clear sex difference. Males tend to show far more 'direct' aggression (pushing, hitting, punching, etc.,) whilst females tend to show more 'indirect' (or 'relational', covert) aggression (gossip, exclusion, bitchy remarks, etc.,). Direct aggression may require an even lower level of empathy than indirect aggression. Indirect aggression needs better mindreading skills than does direct aggression, because its impact is strategic (Crick & Grotpeter, 1995).
10. *Murder is the ultimate example of a lack of empathy*. Daly and Wilson (Daly & Wilson, 1988) analysed homicide records dating back over 700 years, from a range of different societies. They found that 'male-on-male' homicide was 30-40 times more frequent than 'female-on-female' homicide.
11. *Establishing a 'dominance hierarchy'*. Males are quicker to establish these. This in part may reflect their lower empathizing skills, because often a hierarchy is established by one person pushing others around, to become the leader (Strayer, 1980).
12. *Language style*. Girls' speech is more co-operative, reciprocal, and collaborative. In concrete terms, this is also reflected in girls being able to keep a conversational exchange with a partner going for longer. When girls disagree, they are more likely to express their different opinion sensitively, in the form of a question, rather than an assertion. Boys' talk is more 'single-voiced discourse' (the speaker presents their own perspective alone). The female speech style is more 'double voiced discourse' (girls spend more time negotiating with the other person, trying to take the other person's wishes into account) (Smith, 1985).
13. *Talk about emotions*. Women's conversation involves much more talk about feelings, whilst men's conversation with each other tends to be more object- or activity-focused (Tannen, 1991).
14. *Parenting style*. Fathers are less likely than mothers to hold their infant in a face-to-face position. Mothers are more likely to follow through the child's choice of topic in play, whilst fathers are more likely to impose their own topic. And mothers fine-tune their speech more often to match what the child can understand (Power, 1985).
15. *Face preference and eye contact.* From birth, females look longer at faces, and particularly at people's eyes, and males are more likely to look at inanimate objects (Connellan, Baron-Cohen, Wheelwright, Ba'tki, & Ahluwalia, 2001).

These sex differences in social interest appear very early: Little girls also show more eye contact that boys do by 1 yr of age (Lutchmaya & Baron-Cohen, 2002). Some argue that even by this age, socialization might have caused these sex differences. Although there is evidence for differential socialization contributing to sex differences, this is unlikely to be a sufficient explanation. This is because among one day old babies, girls look longer at faces than boys do (Connellan et al., 2001). This raises the possibility that, whilst culture and socialisation may partly determine sex differences in the mind, biology may also partly determine this. There is ample evidence for both cultural and biological influence (Eagly, 1987; Gouchie & Kimura, 1991). For example, the amount of eye contact a child makes at 1 yr old is inversely related to their level of prenatal testosterone (Lutchmaya, Baron-Cohen, & Raggatt, 2002; Lutchmaya, Baron-Cohen, & Raggett, 2002). The evidence for the biological basis of sex differences in the mind is reviewed elsewhere (Baron-Cohen, 2003).

Sex differences in language ability and language acquisition

Females have also been shown to have better language ability than males (Lutchmaya et al, 2002). We know from experimental studies that good empathizing promotes language development (Baron-Cohen, Baldwin, & Crowson, 1997) – for example, looking up at the speaker's face to take note of gaze direction allows normal toddlers of even 18 months old a quicker route into mapping novel words onto their correct referent. Treating language a string of lexical entries each of which has a one-to-one mapping with a referent is not the most effective route into language, since it raises the insoluble problem of reference: that any sound could refer to any object or state of affairs in the environment, or any aspect of an object, or even any

object or state of affairs in the past or future or in a fictional world. Using empathy helps narrow down the search space for reference.

The proof of this idea that good empathy causes faster language acquisition would be if it is shown that female toddlers engage in more joint attention than male toddlers, when hearing novel utterances. This is highly testable. Naturally, this is not to suggest that the direction of causality does not also go in the opposite direction, since a good facility for language would also provide access to another's mental states (people can tell you what's on their mind). The net result would be an upward spiral in development, with empathy promoting language and vice versa. Arguably this ontogenetic pattern might resemble what also occurred in the evolution of these crucial human abilities.

Bibliography

Ahlgren, A., & Johnson, D. W. (1979). "Sex differences in cooperative and competitive attitudes from the 2nd to the 12th grades". Developmental Psychology, 15, 45-49.

Avis, J., & Harris, P. (1991). "Belief-desire reasoning among Baka children: evidence for a universal conception of mind". Child Development, 62, 460-467.

Baron-Cohen, Joliffe, T., Mortimore, C., & Robertson, M. (1997). "Another advanced test of theory of mind: evidence from very high functioning adults with autism or Asperger Syndrome". Journal of Child Psychology and Psychiatry, 38, 813-822.

Baron-Cohen, S. (1994a). "How to build a baby that can read minds: Cognitive mechanisms in mindreading". Cahiers de Psychologie Cognitive/ Current Psychology of Cognition, 13, 513-552.

Baron-Cohen, S. (1994b). "The Mindreading System: new directions for research". Current Psychology of Cognition, 13, 724-750.

Baron-Cohen, S. (2003). The Essential Difference: Men, Women and the Extreme Male Brain. London: Penguin.

Baron-Cohen, S., Baldwin, D., & Crowson, M. (1997). "Do children with autism use the Speaker's Direction of Gaze (SDG) strategy to crack the code of language?" Child Development, 68, 48-57.

Baron-Cohen, S., O'Riordan, M., Jones, R., Stone, V., & Plaisted, K. (1999). "A new test of social sensitivity: Detection of faux pas in normal children and children with Asperger syndrome". Journal of Autism and Developmental Disorders, 29, 407-418

Baron-Cohen, S., Wheelwright, S., Hill, J., & Golan, O. (submitted). "A Taxonomy of Human Emotions".

Bates, E., Benigni, L., Bretherton, I., Camaioni, L., &

Volterra, V. (1979). "Cognition and communication from 9-13 months: correlational findings". In E. Bates (Ed.), The emergence of symbols: cognition and communication in infancy. New York: Academic Press.

Blair, R. J. (1995). "A cognitive developmental approach to morality: investigating the psychopath". Cognition, 57, 1-29.

Brentano, F. (1970). Psychology from an empirical standpoint. London: Routledge, and Kegan Paul.

Butterworth, G. (1991). "The ontogeny and phylogeny of joint visual attention". In A. Whiten (Ed.), Natural theories of mind. Oxford University Press.

Charlesworth, W. R., & Dzur, C. (1987). "Gender comparisons of preschoolers' behavior and resource utilization in group problem-solving". Child Development, 58, 191-200.

Charman, T., Ruffman, T., & Clements, W. (2002). "Is there a gender difference in false belief development?" Social Development, 11, 1-10.

Connellan, J., Baron-Cohen, S., Wheelwright, S., Ba'tki, A., & Ahluwalia, J. (2001). "Sex differences in human neonatal social perception". Infant Behavior and Development, 23, 113-118.

Crick, N. R., & Grotpeter, J. K. (1995). "Relational aggression, gender, and social-psychological adjustment". Child Development, 66, 710-722.

Csibra, G., Gergely, G., Biro, S., Koos, O., & Brockbanck, M. (1999). "Goal attribution without agency cues: the perception of 'pure reason' in infancy". Cognition, 72, 253-284.

Daly, M., & Wilson, M. (1988). Homicide. New York: Aldine de Gruyter.

Davis, M. H. (1994). Empathy: A social psychological approach. Colorado: Westview Press.

Dennett, D. (1987). The Intentional Stance. Cambridge, Mass: MIT Press/Bradford Books.

Dodge, K. (1980). "Social cognition and children's aggressive behaviour". Child Development, 51, 162-170.

Eagly, A. H. (1987). Sex differences in social behavior: A social-role interpretation. Hillsdale, NJ: Erlbaum.

Ekman, P. (1992). "Facial expression of emotion: an old controversy and new findings". In V. Bruce (Ed.), The Face, Royal Society.

Gelman, S., & Hirschfield, L. (1994). Mapping the Mind. Cambridge: Press Syndicate, University of Cambridge.

Gergely, G., Nadasdy, Z., Gergely, C., & Biro, S. (1995). "Taking the intentional stance at 12 months of age". Cognition, 56, 165-193.

Gouchie, C., & Kimura, D. (1991). "The relationship between testosterone levels and cognitive ability patterns". Psychoneuroendocrinology, 16, 323-334.

Griffin, R., & Baron-Cohen, S. (2002). "The intentional stance: developmental and neurocognitive perspectives". In D. Ross (Ed.), Dennett Beyond Philosophy, Cambridge University Press.

Hall, J. A. (1978). "Gender effects in decoding nonverbal cues". Psychological Bulletin, 85(845-858).

Happe, F. (1993). "Communicative competence and theory of mind in autism: A test of Relevance Theory". Cognition, 48, 101-119.

Happe, F. (1995). "The role of age and verbal ability in the theory of mind task performance of subjects with autism". Child Development, 66, 843-855.

Harris, P., Johnson, C. N., Hutton, D., Andrews, G., & Cooke, T. (1989). "Young children's theory of mind and emotion". Cognition and Emotion, 3, 379-400.

Hoffman, M. L. (1977). "Sex differences in empathy and related behaviors". Psychological Bulletin, 84, 712-722.

Hood, B., Willen, J., & Driver, J. (1997). "An eye-direction detector triggers shifts of visual attention in human infants". Unpublished ms., Harvard University.

Johnson, S. (2000). "The recognition of mentalistic agents in infancy". Trends in Cognitive Sciences, 4, 22-28.

Knight, G. P., Fabes, R. A., & Higgins, D. A. (1989). "Gender differences in the cooperative, competitive, and individualistic social values of children". Motivation and Emotion, 13, 125-141.

Leslie, A. (1995). "ToMM, ToBy, and Agency: core architecture and domain specificity." In S. Gelman (Ed.), Domain specificity in cognition and culture. New York: Cambridge University Press.

Leslie, A., & Keeble, S. (1987). "Do six-month old infants perceive causality?" Cognition, 25, 265-288.

Leslie, A. M. (1987). "Pretence and representation: the origins of "theory of mind"". Psychological Review, 94, 412-426.

Lutchmaya, S., & Baron-Cohen, S. (2002). "Human sex differences in social and non-social looking preferences at 12 months of age". Infant Behaviour and Development, 25, 319-325.

Lutchmaya, S., Baron-Cohen, S., & Raggatt, P. (2002). "Foetal testosterone and vocabulary size in 18- and 24-month-old infants". Infant Behavior and Development, 24(4), 418-424.

Lutchmaya, S., Baron-Cohen, S., & Raggett, P. (2002). "Foetal testosterone and eye contact in 12 month old infants". Infant Behavior and Development, 25, 327-335.

Maccoby, E. (1999). The two sexes: growing up apart, coming together. Harvard University Press.

Mundy, P., & Crowson, M. (1997). "Joint attention and early social communication". Journal of Autism and Developmental Disorders, 27, 653-676.

Perner, J., & Wimmer, H. (1985). ""John thinks that Mary thinks that..." Attribution of second-order beliefs by 5-10 year old children." Journal of Experimental Child Psychology, 39, 437-471.

Power, T. G. (1985). "Mother- and father-infant play: A developmental analysis". Child Development, 56, 1514-1524.

Pratt, C., & Bryant, P. (1990). "Young children understand that looking leads to knowing (so long as they are looking into a single barrel)". Child Development, 61, 973-983.

Premack, D. (1990). "The infant's theory of self-propelled objects". Cognition, 36, 1-16.

Rochat, P., Morgan, R., & Carpenter, M. (1997). "Young infants' sensitivity to movement information specifying social causality". Cognitive Development, 12, 537-561.

Scaife, M., & Bruner, J. (1975). "The capacity for joint visual attention in the infant". Nature, 253, 265-266.

Smith, P. M. (1985). Language, the sexes and society. Oxford: Basil Blackwell.

Strayer, F. F. (1980). "Child ethology and the study of preschool soical relations". In J. R. Smith (Ed.), Friendship and Social Relations in Children. New York: John Wiley.

Tager-Flusberg, H. (1993). "What language reveals about the understanding of minds in children with autism". In D. Cohen, J, (Ed.), Understanding other minds : perspectives from autism. Oxford: Oxford University Press.

Tannen, D. (1991). You Just Don't Understand: Women and Men in Conversation. London: Virago.

Tomasello, M. (1988). "The role of joint-attentional processes in early language acquisition". Language Sciences, 10, 69-88".

Walker, A. S. (1982). "Intermodal perception of exptessive behaviours by human infants". Journal of Experimental Child Psychology, 33, 514-535.

Wellman, H., & Bartsch, K. (1988). "Young children's reasoning about beliefs". Cognition, 30, 239-277.

Whiten, A. (1991). Natural theories of mind. Oxford: Basil Blackwell.

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

Yirmiya, N., Sigman, M., Kasari, C., & Mundy, P. (1992). "Empathy and cognition in high functioning children with autism". Child Development, 63, 150-160.

Discussion

Theory of Mind need not lead to Sympathetic Empathy

Robert Stonjek
Mar 30, 2004 0:36 UT

The definition of empathy given by Baron-Cohen here is incomplete and one sided in its nature, possibly leading to a wrong conclusion.

By associating empathy with sympathy and related emotional predispositions, Baron-Cohen is associating ‘Theory of Mind’ only with the observed female bias toward those emotions.

‘Theory-of Mind’ occurs when a child understands that other people (and even animals) have the same experiences as they do, including their emotions and thoughts. That this should lead to sympathy requires some consideration. Is it a natural response achieved by all individuals who tweak to Theory-of-Mind (ToM), or is it one kind of response that can occur, say, if a sympathetic predisposition existing in the observer’s mind is reflected onto the observed individual’s mind via ToM?

Logically, one projects the self onto other people and assume that they respond ‘as I do’. What if I am predisposed to distrust others, to take what I can get and to steal if I can get away with it? Will realising that others have a similar mind to mine produce feelings of sympathetic empathy?

It is likely that some form of innate competition is more predominant in males than females. Wouldn’t this competitiveness lead to greater caution, less openness, greater defensiveness and other related emotional predispositions for the ToM aware male? Wouldn’t the signs of ToM in males be less sympathy, less openness, greater aggression, more defensiveness and a more guarded response to knowledge acquisition?

If acquisition of ToM was the same for males and females, but the response to knowledge of another’s mental state resulted in very different responses to that knowledge, then the differences in language acquisition could be explained quite outside the timing of acquisition of ToM and instead be seen as a consequence of the differences between male and female emotional predispositions.

empathy and different types of language use
Douglas Galbi
Mar 31, 2004 19:39 UT

Rather than considering empathy in relation to "language acquisition," it might be helpful to consider different models of communication involving language. One model is information transfer. Language acquisition in this model is learning the system of codes. Such learning seems like, in Dr. Baron-Cohen's terminology, systemizing or "male brain" activity.

Another model of communication is story-telling. Most stories have a small set of conventional schema, they can imply a coherent alternative physical world, and they can include characters across a wide range of depth and complexity. Language acquisition might be the capacity to enter into stories as systems of coherent relations, or to encounter the character in the stories as engaging persons who have a life beyond the story. Language acquisition in this model might involve different mixtures of empathy and systemizing, depending on the story and the persons.

A third model of communication is making sense of presence of another like oneself. Making sense of presence can be understood as a bodily function contributing to inclusive fitness in strategic interaction among organisms with highly developed neuro-cognitive systems. It is an alternative to a bounded set of representations and decision rules. Language acquisition in this model is a tool for maintaining the state of persons together making sense of the presence of each other. It's like grooming in primates or the sort of social conversation that is the main driver of communication services demand. For further development of this model, see my work, "Sense in Communication," available at www.galbithink.org

Recognizing these different models of communication might contribute to a better understanding of empathy. As Baron-Cohen has documented, some persons don't naturally engage in making sense of others, they see little reason for doing so, and they can't understand how to. Persons who are disadvantaged in this way might be helped by being taught to recognize cues for different mental states, to generate appropriate emotional responses to others' mental states, and to recognize the value of this social competency. This sort of skill might be called functional empathy: the ability to determine the mental state of other persons, "predict what they might do next, and how they might feel." (from Baron-Cohen's text). But that's a systematic substitute for making sense of presence. It's essentially different from open engagement in making sense of another.

As I emphasize in my work, making sense of presence involves a person's whole living body. Baron-Cohen states, "Empathizing is a narrowly defined domain, namely, understanding and responding to people's minds." That seems to me too close to a functional skill that might be helpfully taught to "male brains." Persons seem to often apply social rules and expectations to computers and other inanimate objects (see, e.g. Nass and Moon, "Machines and Mindlessness: Social Responses to Computers," at http://www.findarticles.com/cf_dls/m0341/1_56/63716503/p1/article.jhtml). Yet these persons understand that a computer is not a person. I think that means that they would have a different pattern of activity in their nervous system across all different sensory modes (see Section I of "Sense in Communication"). In other words, what they are doing is not making sense of presence of another like themselves.

Empathy and genetics
Anne Reboul
Apr 1, 2004 11:55 UT

Anyone who has had the priviledge of raising both boys and girls knows that they are indeed very different from birth on, not only for the obvious physiological reason but in terms of behavior. Thus Simon Baron-Cohen's careful review of sex differences in empathy may not come as a surprise. It raises however interesting questions regarding the possible genetic basis not only for greater or lesser empathy, but also for mental pathologies such as autism and (possibly) schizophrenia given that problems with ToM (which might stem in part from an impairment in empathy) have been hypothesized in both of these pathologies. What is more, the sex ratio in autism clearly indicates that males are at greater risk than females and there have been claims of genetic factors given the distribution of the pathology among the population. Thus, I just wanted to know what Simon Baron-Cohen's position is regarding genetics and autism. There have been quite a lot of recent research in genetics about autism, though there seems (unsurprisingly) not to be a single genomic locus. However, Simon's position regarding empathy, as well as the much greater number of males versus females in the affected population might lead one to expect the responsible genes to be sex-linked. The same kind of prediction has been made by Crow and colleagues regarding schizophrenia, not on the basis of empathy, but on the basis of the cerebral peculiarites in schizophrenics, whose brains are less

asymetrical than those of unaffected people, and who seem to acquire language later than the general population. Though there are considerable controversies on this issue, this does raise interesting questions, presumably more general than those Simon talks about in the last alinea of his paper regarding empathy and language acquisition. Here are some of them: are there any anomalies regarding asymetry of brain structures in autistic people, as has been claimed for schizophrenic people? If one or several genes are identified as the genetic basis of autism, should one expect them to be found on a sex-linked rather than autosomal chromosome? And finally, if there are anomalies in brain asymetry, should this be viewed as explaining the difficulties in language acquisition which seem general among autistic (though not Asperger) subjects, or should it be viewed as explaining the difficulties in empathizing? and this, of course, leads us back to the general topic of conference, i.e. what is the relation (and I have no doubt that there is a strong one) between language and ToM (including of course empathy)?

Did women invent language?
Dan Sperber
Apr 1, 2004 23:08 UT

Let us (and Simon in particular) speculate: assuming that these differences in empathy between males and females also existed among our ancestors before they developed a language faculty, might man and women have played a different role in the emergence of language? Could language have been more a female contribution to our becoming modern humans? Such an imbalance could have resulted either from a difference in selective pressures on males and females, or as a by-product. Any ideas?

Empathy or intention-attribution?
Robyn Carston
Apr 11, 2004 21:39 UT

Three fairly well-supported findings are brought together in Simon Baron-Cohen's paper to support his hypothesis that sex differences in language acquisition (specifically, the acquisition of word meanings) might be accounted for by sex differences in empathy. These are: (a) The acquisition of word meanings appears to be greatly facilitated by a capacity to figure out a speaker's referential intentions; (b) Girls have a 'better language ability' than boys, which, for it to be relevant here, must include better vocabulary at particular stages of development; (c) Girls and women are more empathic than boys and men.

While I find this thesis very interesting, it is not clear to me that it is 'empathy' that plays a crucial role here rather than (just) the narrower capacities of joint attention and contentful (propositional) mental state attribution. What role is played by the additional component, which I think is intrinsic to what we think of as empathy, of 'having an emotional reaction that is appropriate to the other person's mental state'?

I guess there is a wider issue here too which is whether or not the two conceptually distinct elements in Simon's description of empathy are cognitively dissociable. We can imagine a person who has a more or less normal, or even enhanced, ability to attribute beliefs, intentions, desires, etc, to another person while not having an 'appropriate' (=sympathetic?), or indeed any, emotional response to that other person. Could certain socio/psychopathological conditions perhaps be characterised in this way? (I don't know.) The standard false belief task doesn't seem to call for any emotional response. We can also envisage the reverse, that is, someone who is sensitive to other people's emotional states and responds with appropriate expressions of feeling while being rather poor at attributing contentful mental states such as beliefs and intentions. I believe there is some (very tentative) speculation along these lines about children with Williams Syndrome.

Returning to the issue of sex differences in vocabulary acquisition, I think that for the thesis to be shown to hold we need to be given some reason fo believe that what is specific to empathy (i.e. the emotional response element) plays a role in acquiring the meaning of words. Is there any evidence of this? An equally interesting sex-based thesis would be that girls' better vocab can be accounted for by their having better theory-of-mind in the narrower sense (i.e. minus the specifically empathic quality), but, from what Simon tells us, the results of the existing studies on this are equivocal.

A Pragmatic Perspective on the Evolution of Langage and Languages.

Gloria Origgi (CNRS, Institut Jean-Nicod) and Dan Sperber (CNRS, Institut Jean-Nicod)

(Date of publication: 12 April 2004)

Abstract: Language is both a biological and a cultural phenomenon. Our aim here is to discuss, in an evolutionary perspective, the articulation of these two aspects of language. For this, we draw on a pragmatic view of language development and language evolution.

(Translated from French by Marcel Lieberman)

Two Views of Linguistic Communication

Simplifying somewhat, one can contrast two models of linguistic communication: the classical model of communication, or the "code model", and the inferential model. According to the code model, the communicator encodes her message by means of a signal that the hearer then decodes. Sentences of a language are just complex signals that encode messages.

Against this model, pragmatic linguistics objects that the same sentence can be used to communicate an indefinite number of different messages that cannot be retrieved by simple decoding. Take, for example, the sentence: "It's too slow." This very ordinary sentence does not present any particular linguistic difficulty. Yet, it can be used to convey an indefinite number of meanings, for example: The mouse is too slow in solving the maze; The chemical reaction is too slow compared to what we expected; The decrease in unemployment is too slow to avoid social unrest; Jacques' car is too slow (and so I'd suggest we take Pierre's) etc. In order to discover what meaning is transmitted by the uttering of this sentence, one needs contextual information. Contextual information will help one understand what "it" refers to, what kind of slowness is at issue (slowness of movement, of thought, of a process, etc.), and what criteria determine the value of "too". It will help one recognize the more or less literal, hyperbolic or ironic character of the utterance. Lastly, it will help one retrieve its possible implicatures.

According to the inferential model, different versions of which have been developed in contemporary pragmatics, an utterance is a piece of evidence of the speaker's meaning. Decoding the linguistic sentence meaning is seen as just one part of the process of comprehension—a process that relies on both this linguistic meaning and on the context in order to identify the speaker's meaning. It was the philosopher Paul Grice (1957) who first developed this point of view. He approached the relationship between sentence meaning and speaker's meaning in an original way. Speaker's meaning, in Grice's analysis, is a complex communicative intention that must be recognized by the hearer in order to be fulfilled. It is an intention to achieve a certain effect upon the mind of the hearer by means of the hearer's recognition of the very intention to achieve this effect.

Seen this way, communication depends upon the ability of human beings to attribute mental states to others; that is, it depends upon their naïve psychology. This ability has been the subject of considerable work in developmental psychology (Baron-Cohen et al., 2000) and in the study of the evolution of social behavior (Byrne and Whyten, 1988). Humans spontaneously interpret one another's behavior, not as simple body movements, but as the belief-guided fulfillment of intentions. Living in a world inhabited not only by physical objects and living bodies, but also by mental states, humans may want to act upon these mental states. They may seek to change the desires and beliefs of others. Such action can be carried out unbeknownst to the person one seeks to influence. It can also be performed overtly—one makes it manifest that one is traying to cause one's audience to believe or desire something—and then it is a communication proper.

Communication is achieved by giving the hearer evidence of the meaning one intends to communicate. This evidence can be of any sort—gestures, mimicry, showings; and they can be coded or not, provided that they allow the hearer to infer the speaker's meaning.

In inferential communication the communicator seeks to fulfill her intention by making it manifest to the hearer. Such a procedure carries a clear risk: the hearer, recognizing that one seeks to influence him, can easily foil this intention. On the other hand, inferential communication, because of the very fact that it is overt, has two advantages that make it generally much more powerful than all the other ways of acting upon people's mental states. While a mistrustful hearer may refuse to be influenced, a hearer who trusts the communicator's competence and honesty will make an effort to understand a message that he'll be willing to accept. More importantly still, whereas the manipulation of the mental states of others by non-communicational means is relatively cumbersome to enact and always imprecise, overt communication makes it possible to transmit at very little cost contents as rich and precise as one wants.

The role of language in inferential communication is to provide the communicator with evidence, as exact and complex as she wishes, of the content she wants the hearer to accept. For this, it is not necessary that the utterance encode this content in extenso and unambiguously. Quite commonly, a fragmentary, ambiguous and loose coding is sufficient, in the context, for indicating a complete and unequivocal meaning. In this respect, inferential comprehension is not different from any other cognitive process of non-demonstrative inference that draws relatively reliable conclusions from fragmentary evidence open to multiple interpretations by relying upon both empirical regularities and context.

The main task of pragmatics is to explain how such a process of inference is carried out in the particular case of linguistic communication: what empirical regularities guide the process? How are the linguistic properties of the utterance, on the one hand, and contextual information on the other, put to use? Although different pragmatic theories (e.g. Ducrot, Grice, Levinson, Sperber & Wilson) give different answers to these questions, they agree on the two basic considerations: comprehension is inferential, and, by drawing on both the sentence meaning and the context, it aims at discovering the meaning intended by the speaker.

One sees that the classical code model and the inferential model developed by pragmatics assign different functions to language in linguistic communication. To different functions there should correspond, in the species' history, different selective pressures and hence different hypotheses regarding the biological evolution of language. Yet, whether it is because, in practice, they accept the code model (e.g. Pinker, 1994, ch. 7), or because they consider the evolution of language without worrying about its specifically linguistic properties (e.g. Dunbar 1996), theorists of the evolution of language have not given much of a role to the pragmatic dimension of language (Dessalles 2000 is an exception), and have considered even less the precise role of communicational processes in linguistic communication.

The Evolution of Language and Two Models of Linguistic Communication

Coded communication functions best when interlocutors share exactly the same code. Any difference between the communicator's code and that of the hearer is, on the other hand, a source of possible error in the communication process. Under these conditions, a mutation affecting an individual's language faculty places her at the risk of internalizing a code that is different from that of her conspecifics on the basis of the same linguistic data. This mismatch of codes would be detrimental to the individual's ability to communicate. It would be counter-adaptive.

More generally, since a code must be shared by a population in order to be advantageous, evolution cannot easily "experiment" with modifications whose anyhow low chance of being advantageous could not verified until the modification was sufficiently widespread. The most plausible modifications are additions of new signals to the code (for example, a signal of alarm for a new species of predators in the environment)—additions that do not modify the structure of the pre-existing code. The very modest size of codes in animal communication suggests that these additions are themselves quite rare. Indeed, animal communication codes which, unlike human languages, really function according to the code model, are typically small and highly stable within a given species. The great majority of them involve no learning, and when learning is involved, as in the case of songbirds, it usually concerns only a single signal that must be learned since it serves for distinguishing between local populations of the same species.

In the case of inferential communication, the situation is quite different. The success of inferential communication does not require that the communicator and the audience have the same semantic representation of the utterance. It suffices that the utterance, however they may represent it, be seen as evidence for the same conclusion. Take, for example, the following trivial dialogue:

Pierre: *I'm beat!*

Marie: *Ok, let's go back home.*

It is of little importance whether the meaning that Pierre and Marie associate with the word "beat" is the same. It may be that, for Pierre, "beat" means an extreme fatigue, while for Marie, "beat" is simply a synonym of "tired". In any event, Pierre says, "I'm beat," not in order to indicate a degree of fatigue that this term might encode, but in order to indicate contextually both his wish to return home and the reason for it, namely his fatigue. The level of fatigue that may justify one's desire to return home depends on the situation: it is not the same at a party among friends, while taking a stroll, or at work. In Pierre's utterance, then, "beat" indicates the level of fatigue which, in the situation of the utterance, is relevant in that it justifies Pierre's wish. It is not necessary that the codes between interlocutors be identical; nor is it sufficient. Consider the following dialogue:

Pierre: *Can you fix my watch?*

Watchmaker: *That will take some time.*

The semantics of “will take some time” is trivial (or in any case, let us suppose that it is, and in the same manner, for Pierre and the watchmaker): everything that has non-zero duration takes time. Yet, in uttering this truism, the watchmaker sets Pierre along the way to a relevant interpretation. It is indeed a matter of time, and repairing the watch will take a certain amount of time to which it is relevant to draw Pierre’s attention. If Pierre expects that the time for repair will be at least one week, he will understand “It will take some time” as meaning that the repair will take several weeks. If the watchmaker, for his part, thinks that Pierre expects the repair to be done the same day, he will express himself as he did in order to say that the repair will be a matter of days rather than hours. That the words “to take”, “some” and “time” have the same meaning in their lexicon does not protect them from the possibility of misunderstanding. According to the inferential model, the near identity of the interlocutors’ codes is not necessary in order for them to best communicate. In these conditions, a mutation affecting the language faculty and causing the mutant’s grammar to diverge from that of her interlocutors is not necessarily detrimental to her ability to communicate. As we will now show, such a mutation may even be advantageous.

In particular, a language faculty leading to the internalization of a grammar that attributes more structure to utterances than they superficially realize (one that, for example, projects onto them unexpressed constituents) could facilitate inferential comprehension. Imagine a proto-language having only word-size sound-meaning pairs, without any syntactic structure. The word “drink” in this proto-language designates the action of drinking and nothing else (it is not, unlike the word “drink” in English, a two-place predicate); the word “water” designates the substance and nothing else, and so on. With such a limited code, a hearer’s decoding of the meaning associated with the word pronounced by the speaker would not suffice to assure communication between them. The hearer who associates with the utterance “water” the concept of water is not thereby informed of anything whatsoever. Even a concatenation of expressions in such a language, for example, “drink water”, would not be decoded as we spontaneously tend to do on the basis of our comprehension of English. “Drink water” does not denote, in this proto-language, the action of drinking water. One only has two concepts, that of drinking and that of water, which are activated without being semantically linked. The mental activation of one or several concepts having no semantic linkage between them does not denote a state of affairs nor an action associating these two concepts; and it expresses even less a belief or desire.

In these conditions, such a proto-language could be of use only to beings capable of inferential communication. For such individuals, activation by decoding, even of only a single concept, could easily provide them with evidence sufficient for reconstructing a full-fledge meaning, the speaker’s meaning. Imagine two speakers of this proto-language, let us call them Pierre and Marie, walking in the desert. Pierre points to the horizon and utters, “water”. Marie correctly infers from this that he means something like, There is water over there. Just when they reach the water hole, Pierre, exhausted, collapses and mutters, “water”. Mary correctly infers that he means something like, give me some water. With the signals of animal communication—communication that is fully coded—such a range of interpretive constructions in not possible.

Imagine now that Marie was infact a mutant whose language faculty, more complex than that of her fellow creatures, had allowed her as a child to analyze the words of the proto-language that she was in the process of acquiring, either as arguments or as one- or two-place predicates. She had thus categorized “drink” as a two-place predicate, “water” as an argument, and so on. When Marie the mutant hears parched Pierre mutter, “water,” what is activated in her mind is not only the concept of water, but a syntactic structure with an unexpressed predicate capable of taking water as an argument. Her decoding thus goes beyond what was in fact encoded by Pierre. He is not a mutant and therefore expresses himself in the rudimentary language of their community, without mentally adding to it an underlying syntactic structure. This mismatch between Pierre and Marie’s representation of the utterance is not, however, detrimental to communication. Even if she weren’t a mutant, Marie would have had to mentally (but not linguistically) represent, in order to interpret what Pierre meant, not only water but also the action that had water as an object. Marie the mutant is immediately set along the right path thanks to the syntactic structure she falsely, though usefully, attributes to Pierre’s utterance. When she speaks, Marie the mutant encodes, by means of signals that are homonymous to those of her community, not only atomic concepts but also predicate-argument structures. When she says, “water”, her utterance also encodes the unexpressed place-holder of a predicate for which “water” would be the argument. When she says “drink”, her utterance encodes the unexpressed place-holder of the two arguments of “drink”. When she says “drink water”, her utterance encodes not only the two concepts drink and water, but also the complex concept drink some water (plus the unexpressed place-holder of the argument-subject of “drink”). Marie’s interlocutors do not recognize these underlying structures in her utterances, but they arrive at the intended interpretations all the same by a linguistically less prepared inferential path.

Now, if Marie is a second-generation mutant, having among her interlocutors brothers and sisters who are also mutants who therefore speak and comprehend as she does, then she and her co-mutants communicate more effectively than the other members of their community. They communicate, in fact, by means of a

language whose utterances, phonologically identical to those of the non-mutants' language, are syntactically and semantically more complex and hence easier to deal with pragmatically. In the language of these mutants, new linguistic signs may emerge and stabilize by a process of grammaticalization that is inaccessible to non-mutants. For example, pronouns could come to take the place of unspecified arguments.

This imaginary example illustrates the way in which a more advanced language faculty, which leads individuals possessing it to internalize a code that is richer than that of their community, may emerge and evolve. It only occurs this way in a system of inferential communication.

In a system of code-based communication, every departure from the common grammar will be disadvantageous or at best neutral, but it will never be advantageous. These considerations apply to all possible stages of the evolution of the language faculty as well as to its initial emergence. Being disposed to treating uncoded communicational behavior as a coded signal may facilitate inferential comprehension of the communicator's intentions and lead to the stabilization of this kind of behavior as a signal.

Conclusions

The human mind is characterized by two cognitive abilities having no real equivalent in other species on Earth: language and naïve psychology, that is, the ability to represent the mental states of others. We have suggested here that it is because of the interaction of these two abilities that human communication was able to develop and acquire its incomparable power (cf. Origgi 2001, Origgi & Sperber 2000, Sperber 2000). From a pragmatics perspective, it is quite clear that the language faculty and human languages, with their richness and flaws, are only adaptive in a species that is already capable of naïve psychology and inferential communication. The relatively rapid evolution of languages themselves and their lack of homogeneity within one and the same linguistic community—these two aspects being associated—can only be adequately explained if the function of language in communication is to provide evidence of the speaker's meaning and not to encode it. In these conditions, the study of the evolution of language must be closely associated to that of the evolution of naïve psychology. Likewise, the study of the evolution of languages must systematically take into account their pragmatic dimension.

French version

Qu'est-ce que la pragmatique peut apporter à l'étude de l'évolution du langage ?

Résumé : L'expression "évolution du langage" se réfère à deux phénomenes distincts: l'évolution biologique d'une faculté du langage et l'évolution historique et culturelle des langues humaines. Notre propos c'est d'articuler, dans une perspective évolutionniste, ces deux aspects du langage.

Deux conceptions de la communication linguistique

On peut, en simplifiant, opposer deux modèles de la communication linguistique, le modèle classique de la communication ou «modèle du code» et le modèle inférentiel. Selon le modèle du code, tout communicateur encode son message au moyen d'un signal que le destinataire décode. Les phrases de la langue sont précisément des signaux complexes qui encodent des messages.

A ce modèle du code, la pragmatique linguistique objecte que la même phrase peut servir à communiquer un nombre indéfini de messages différents qui ne peuvent donc pas être reconstitués par simple décodage. Soit par exemple la phrase: «Elle est trop lente». Cette phrase tout à fait ordinaire ne présente aucune difficulté linguistique particulière. Elle peut cependant servir à véhiculer un nombre indéfini de sens, par exemple Marie calcule trop lentement pour finir les exercices à temps, la réaction chimique est trop lente par rapport à ce qui avait été prévu, la baisse du taux de chômage est trop lente pour éviter des mouvements sociaux, la voiture de Jacques est trop lente (et donc je propose de prendre celle de Pierre), etc. Pour découvrir quel sens une énonciation de cette phrase véhicule, il faut disposer d'informations contextuelles. Ces informations aideront à comprendre à qui ou à quoi «elle» se réfère, de quelle lenteur il est question (dans le mouvement, dans la pensée, dans le déroulement d'un processus, etc.), quels critères déterminent la valeur de «trop». Elles aideront à reconnaître le caractère plus ou moins littéral, hyperbolique ou ironique de l'énoncé. Elles aideront enfin à en reconnaître les éventuelles implicitations.

Selon le modèle inférentiel, dont différentes versions sont développées dans la pragmatique contemporaine, un énoncé est un indice du sens voulu par le locuteur. Le décodage du sens linguistique y est vu comme un

élément seulement du processus de compréhension, processus qui s'appuie à la fois sur ce sens linguistique et sur le contexte pour aboutir à une identification du vouloir-dire du locuteur.

C'est, le philosophe Paul Grice (1957) qui a le premier développé cette perspective. Il a analysé en effet de façon nouvelle les rapports entre sens linguistique (sentence meaning) et sens voulu (speaker's meaning). Le sens voulu, dans l'analyse que Grice en propose, est une intention complexe du communicateur qui a besoin d'être reconnue par le destinataire pour être accomplie. Il s'agit en effet d'une intention d'avoir un certain effet sur l'esprit du destinataire en faisant reconnaître l'intention qu'on a d'avoir cet effet.

La communication ainsi conçue relève de la capacité des êtres humains d'attribuer des états mentaux à autrui, c'est-à-dire de leur psychologie naïve Cette capacité a fait l'objet de nombreux travaux en psychologie du dévéloppement (Baron-Cohen et al., 2000) et dans l'étude de l'évolution des comportements sociaux (Byrne et Whyten, 1988). Les humains interprètent spontanément le comportement les uns des autres non pas comme des simples mouvements corporels, mais comme la réalisation d'intentions guidées par des croyances. Vivant dans un monde habité non seulement par des objets physiques et des corps vivants mais aussi par des états mentaux, les humains peuvent vouloir agir sur ces états mentaux. Ils peuvent chercher à modifier les désirs et les croyances d'autrui. Cette action peut être menée à l'insu de celui qu'on essaye d'influencer. Elle peut aussi être menée ouvertement – on montre à autrui qu'on l'incite à croire ou à vouloir quelque chose –, et il s'agit alors de communication proprement dite. La communication s'effectue en donnant au destinataire des indices du sens qu'on entend lui communiquer. Ces indices peuvent être de tout ordre, gestes, mimiques, monstrations, ils peuvent être codés ou non codés, pourvu qu'ils permettent au destinataire d'inférer le sens voulu.

Dans la communication inférentielle le communicateur cherche à réaliser son intention en la rendant manifeste au destinataire. Une telle procédure comporte un risque évident : le destinataire, comprenant qu'on cherche à l'influencer, peut facilement déjouer cette intention. En revanche, la communication inférentielle, du fait même qu'elle est ouverte, comporte deux avantages qui la rendent, dans la plupart des cas, bien plus puissante que toutes les autres façons d'agir sur les états mentaux. En effet, si un destinataire méfiant refusera de se laisser influencer, un destinataire qui a confiance en la compétence et l'honnêteté du communicateur fera de lui-même un effort pour comprendre un message qu'il sera disposé à accepter. Plus important encore, tandis que la manipulation des états mentaux d'autrui par des moyens non communicationnels est relativement lourde à mettre en œuvre et reste toujours imprécise, la communication ouverte permet de véhiculer à très peu de frais des contenus aussi riches et précis qu'on le souhaite. Le rôle du langage, dans la communication inférentielle est justement de fournir au communicateur des indices aussi précis et complexes qu'il le souhaite du contenu qu'il veut faire accepter par le destinataire. Il n'est pas besoin pour cela que l'énoncé encode ce contenu in extenso et sans ambiguïté. De façon tout à fait ordinaire, un encodage fragmentaire, ambigu et imprécis suffit, dans le contexte, à indiquer un sens complet et univoque. La compréhension inférentielle n'est pas, à cet égard, différente de tous les autres processus cognitifs d'inférence non-démonstrative, qui tirent des conclusions assez fiables d'indices fragmentaires et ouverts à plusieurs interprétations en s'appuyant sur des régularités empiriques et sur le contexte. La tâche principale de la pragmatique est d'expliquer comment s'effectue un tel processus d'inférence dans le cas particulier de la communication linguistique : quelles régularités empiriques guide le processus ? Comment les propriétés linguistiques de l'énoncé d'une part et les informations contextuelles d'autre part sont-elles exploitées ? Si différentes théories pragmatiques (e.g. Ducrot, Grice, Levinson, Sperber et Wilson) donnent des réponses différentes à ces questions, elles s'accordent en revanche sur les deux considérations qui fondent toute leur démarche : la compréhension est inférentielle, et elle vise, en s'appuyant sur le sens linguistique de la phrase et sur le contexte, à découvrir le sens voulu par le locuteur.

Le modèle classique du code et le modèle inférentiel développé par la pragmatique assignent, on le voit, des fonctions différentes au langage dans la communication linguistique. A des fonctions différentes devraient correspondre, dans l'histoire de l'espèce, des pressions sélectives elles aussi différentes et donc des hypothèses différentes sur l'évolution biologique du langage. Cependant, soit qu'en pratique ils acceptent le modèle du code (comme par exemple Pinker, 1994, ch.7), soit qu'ils envisagent l'évolution du langage sans se soucier de ses propriétés proprement linguistiques (comme par exemple Dunbar 1996), les théoriciens de l'évolution du langage n'ont pas donné de place significative à la dimension pragmatique du langage (Dessalles 2000 faisant à cet égard exception) et ont encore moins pris en considération le rôle précis des processus communicationnels dans la communication linguistique.

L'évolution du langage et les deux modèles de la communication linguistique

La communication codée fonctionne au mieux quand les interlocuteurs partagent exactement le même code. Toute différence entre le code du communicateur et celui du destinataire est en revanche une source d'erreur possible dans le processus de la communication. Dans ces conditions, une mutation affectant la faculté de langage d'un individu risquera de lui faire intérioriser un code différent de celui d'autrui à partir des mêmes données linguistiques Cette non-correspondance des codes nuira à la capacité de l'individu de communiquer. Elle sera anti-adaptative.

Plus généralement, puisqu'un code doit être partagé par une population pour être avantageux, l'évolution ne peut pas facilement «expérimenter» avec des modifications dont les chances, de toute façon faibles, d'être avantageuses ne se vérifieront que lorsque la modification sera suffisamment répandue. Les modifications les plus plausibles sont des ajouts au code de nouveaux signaux (par exemple d'un signal d'alarme pour une nouvelle espèce de prédateurs dans l'environnement), ajouts qui ne modifient pas la structure du code préexistant. La taille très modeste des codes de la communication animale suggère que ces ajouts eux-mêmes sont rares. De fait, les codes de la communication animale, qui, à la différence des langues humaines, fonctionnent vraiment selon le modèle du code, sont typiquement petits et d'une grande stabilité à l'intérieur d'une espèce donnée. La grande majorité d'entre eux n'implique aucun apprentissage, et lorsque apprentissage il y a, comme dans le cas des oiseaux chanteurs, il ne porte typiquement que sur un seul signal dont il faut bien qu'il soit appris puisqu'il a pour fonction de distinguer des populations locales de la même espèce.

Dans le cas de la communication inférentielle, les choses se présentent tout autrement. En effet, le succès de la communication inférentielle ne requiert pas que le communicateur et le destinataire aient la même représentation sémantique de l'énoncé. Il suffit qu'ils voient dans l'énoncé, quelle que soit la façon dont ils se le représentent, un indice indiquant la même conclusion. Soit par exemple le dialogue banal suivant:

Pierre: *je suis crevé !*

Marie: *Eh bien, rentrons à la maison*

Peu importe que le sens auquel Pierre et Marie associent le mot «crevé» soit le même. Il se peut que, pour Pierre, «crevé» signifie une fatigue extrême, tandis que, pour Marie, «crevé» est un simple synonyme de «fatigué». De toute façon, Pierre dit «je suis crevé» non pour indiquer un degré de fatigue que ce terme encoderait, mais pour indiquer contextuellement à la fois sont souhait de rentrer à la maison et la raison de ce souhait, à savoir sa fatigue. Le degré de fatigue qui justifie qu'on désire rentrer chez soi dépend des situations : il n'est pas le même à une soirée entre amis, en promenade, ou au travail. Dans l'énoncé de Pierre, donc, «crevé» indique le degré de fatigue qui, dans la situation d'énonciation, est pertinent en ceci qu'il justifie le souhait de Pierre.

L'identité de code entre les interlocuteurs n'est pas nécessaire; elle n'est pas suffisante non plus. Soit le dialogue suivant:

Pierre: *Pouvez vous réparer ma montre ?*

L'horloger: *Cela prendra du temps.*

La sémantique de «prendra du temps» est triviale (ou en tout cas admettons qu'elle le soit, et de la même manière, pour Pierre et l'horloger) : prend du temps tout ce qui a une durée non nulle. Si on s'en tient au sens littéral de son énoncé, l'horloger énonce donc un truisme. Cependant, en énonçant ce truisme, l'horloger met Pierre sur la voie d'une interprétation qui, elle, est pertinente. C'est bien de temps qu'il s'agit et la réparation de la montre prendra un temps sur lequel il est pertinent d'attirer l'attention de Pierre. Si Pierre s'attend à une réparation d'au moins une semaine, il comprendra «Cela prendra du temps» comme voulant dire que la réparation prendra plusieurs semaines. Si l'horloger pour sa part pense que Pierre s'attend à une réparation le jour même, il se sera exprimé comme il l'a fait pour dire que la réparation sera une affaire non d'heures mais de jours. Que dans leur lexique les mots «prendre» «du» et «temps» aient le même sens ne les protège donc pas du malentendu.

Selon le modèle inférentiel, donc, la quasi-identité de code entre interlocuteurs n'est pas nécessaire pour qu'ils puissent communiquer au mieux. Dans ces conditions, le fait qu'une mutation affectant la faculté de langage peut faire diverger la grammaire du mutant de celle de ses congénères ne nuit pas forcément à sa capacité de communiquer. Comme nous allons le montrer maintenant, une telle mutation peut même être avantageuse.

En particulier, une faculté de langage qui mènerait à intérioriser une grammaire qui attribue plus de structure aux énoncés qu'ils n'en réalisent superficiellement (qui, par exemple, y projetterait des constituants non-exprimés) pourrait faciliter la compréhension inférentielle.

Imaginons une proto-langue ne comportant que des paires son-sens de l'échelle du mot, sans aucune structure syntaxique. Le mot «boire» dans cette proto-langue désigne l'action de boire et ne fait rien de plus (ce n'est pas, contrairement au «boire» du français, un prédicat à deux places), le mot «eau» dans cette proto-langue désigne la substance et ne fait rien de plus, etc. Avec un code aussi limité, le seul décodage par l'auditeur du sens associé au mot prononcé par le locuteur ne suffirait pas assurer de communication entre eux. L'auditeur qui associe à l'énoncé «eau» le concept d'eau n'est pas pour autant informé de quoi que ce soit. Même une concaténation d'expressions d'un tel langage, comme par exemple «boire eau» ne serait pas décodée comme nous avons tendance à le faire spontanément sur la base de notre compréhension du français. «Boire eau» ne désigne pas, dans cette proto-langue, l'action de boire de l'eau. On a seulement deux concepts, celui de boire et celui d'eau qui sont activés sans être sémantiquement liés.

L'activation mentale d'un ou plusieurs concepts sans lien syntaxique entre eux ne désigne pas un état de chose ou une action associant ces deux concepts; encore moins exprime-t-elle une croyance ou un désir.

Dans ces conditions, une telle proto-langue n'aurait d'utilité que pour des êtres capables de communication inférentielle. Pour de tels individus, l'activation par décodage ne serait-ce que d'un seul concept pourrait facilement leur fournir un indice suffisant pour reconstruire un sens complet, le vouloir-dire d'un locuteur. Imaginez deux locuteurs de cette proto-langue, appellons les Pierre et Marie, marchant dans le désert. Pierre montre du doigt l'horizon et dit «eau». Marie en infère correctement qu'il veut dire quelque chose comme il y a de l'eau là-bas. Juste quand ils parviennent au point d'eau, Pierre s'effondre épuisé et dit «eau». Marie en infère correctement qu'il veut dire quelque chose comme donne moi de l'eau. Juste quand ils parviennent au point d'eau, Pierre s'effondre épuisé et dit «eau». Marie en infère correctement qu'il veut dire quelque chose comme donne moi de l'eau. Les signaux de la communication animale, communication qui elle est pleinement codée, ne permettent jamais un tel éventail de constructions interprétatives.

Imaginons maintenant que Marie ait été une mutante dont la faculté de langage, plus complexe que celle de ses congénères, l'avait disposée, enfant, à analyser les mots de la proto-langue qu'elle était en train d'acquérir soit comme des arguments, soit comme des prédicats à une ou deux places. Elle avait ainsi catégorisé «boire» comme un prédicat à deux places, «eau» comme un argument, etc.

Quand Marie la mutante entend Pierre assoiffé murmurer «eau», ce qui est activé dans son esprit, ce n'est pas seulement le concept d'eau, mais une structure syntaxique avec un prédicat non-exprimé capable de prendre eau comme argument. Son décodage va donc au-delà de ce qui a, en fait, été encodé par Pierre. Lui n'est pas un mutant et il s'exprime donc dans la langue rudimentaire de leur communauté, sans y ajouter mentalement de structure syntaxique sous-jacente. Cette non-correspondance entre la représentation que Pierre et Marie se font de l'énoncé, ne nuit cependant pas à la communication. Même si elle n'avait pas été une mutante, Marie aurait du, pour interpréter ce que voulait dire Pierre, se représenter mentalement (mais pas linguistiquement) non seulement l'eau mais l'action qui avait l'eau pour objet. Marie la mutante est d'emblée mise sur la voie grâce à la structure syntaxique qu'elle attribue, faussement mais utilement, à l'énoncé de Pierre.

Quand elle parle, Marie la mutante encode au moyen de signaux homonymes avec ceux de sa communauté, non seulement des concepts atomiques, mais des structures prédicat-arguments. Quand elle dit «eau», son énoncé encode aussi la position non-exprimé d'un prédicat dont «eau» serait l'argument. Quand elle dit «boire», son énoncé encode la position non-exprimée des deux arguments de «boire». Quand elle dit «boire eau», son énoncé encode non seulement les deux concepts boire et eau mais aussi le concept complexe boire de l'eau (plus la position non-exprimée de l'argument-sujet de «boire»). Les interlocuteurs de Marie ne reconnaissent pas ces structures sous-jacentes dans ses énoncés, mais ils aboutissent tout de même aux interprétations voulues par un chemin inférentiel moins bien préparé linguistiquement. Maintenant, si Marie est une mutante de la deuxième génération et a parmi ses interlocuteurs des frères et sœurs eux aussi mutants qui parlent et comprennent comme elle, alors elle et ses co-mutants communiquent plus efficacement que les autres membres de leur communauté. Il communiquent en effet au moyen d'une langue dont les énoncés, phonologiquement identiques à ceux de la langue des non-mutants, sont syntaxiquement et sémantiquement plus complexes et ainsi plus faciles à traiter pragmatiquement. Dans la langue de ces mutants, de nouveaux signes linguistiques peuvent émerger et se stabiliser par un processus de grammaticalisation inaccessible aux non-mutants. Par exemple des pronoms pourraient venir prendre la place des arguments non-spécifiés.

Cet exemple imaginaire illustre la façon dont une faculté de langage plus avancée, qui amène les individus qui en sont dotés à intérioriser un code plus riche que celui de leur communauté, peut émerger et évoluer. Il n'en va ainsi que dans un système de communication inférentielle. Dans un système de communication codique, tout écart par rapport à la grammaire commune sera désavantageux ou au mieux neutre, mais il ne pourra guère être avantageux.

Ces considérations s'appliquent à toutes les étapes éventuelles de l'évolution de la faculté de langage et aussi à son émergence initiale. Etre disposé à traiter un comportement communicationnel non-codé comme un signal codé peut faciliter la compréhension inférentielle des intentions du communicateur et mener à la conventionnalisation de ce type de comportement en tant que signal.

Conclusions

L'esprit humain se caractérise par deux capacités cognitives sans véritable équivalent dans d'autres espèces terrestres : le langage et la psychologie naïve, c'est-à-dire la capacité de tout un chacun de se représenter les états mentaux d'autrui. Nous avons suggéré que c'est grâce à l'interaction de ces deux capacités que la communication humaine a pu se développer et acquérir une puissance incomparable (cf. aussi Origgi 2001, Origgi & Sperber 2000, Sperber 2000). Dans une perspective pragmatique, il est clair en effet que la faculté de langage et les langues humaines, avec leur richesse et leur imperfections, ne sont adaptatives que dans une espèce déjà capable de psychologie naïve et de communication inférentielle.

L'évolution relativement rapide des langues elles-mêmes et leur manque d'homogénéité à l'intérieur même d'une communauté linguistique – ces deux traits étant associés – ne s'expliquent bien, eux aussi, que si la fonction du langage dans la communication est de fournir des indices sur le sens voulu et non de l'encoder. Dans ces conditions, l'étude de l'évolution du langage doit être étroitement associée à celle de l'évolution de la psychologie naïve. De même l'étude de l'évolution des langues doit prendre systématiquement en considération leur dimension pragmatique.

References

Baron-Cohen, S. ; Tager-Flusberg, H. ; Cohen, D. J. (2000) *Understanding Other Minds*, Second edition, New York, Oxford University Press.
Byrne, R. Whiten, A. (eds.) (1988) *Machiavellian Intelligence : Social Expertise and Evolution of Intellect in Monkeys, Apes and Humans*, Oxford : Oxford University Press.
Dessalles, J-L. (2000) *Aux origines du langage. Une histoire naturelle de la parole*, Paris, Hermès.
Ducrot, O. (1972) *Dire et ne pas dire*, Paris, Hermann.
Dunbar, R. I. M. (1996) *Grooming, Gossip and the Evolution of Language*, London, Faber & Faber.
Grice, P. (1957) "Meaning", *Philosophical Review*, 66 : 377-88.
Grice, P. (1992) *Studies in the Ways of Words*, Cambridge, Mass., Harvard University Press.
Levinson, S.C. (1983) *Pragmatics*, Cambridge : Cambridge University Press.
Origgi, G. (2001) "Interpretare il linguaggio e interpretare gli altri : una o due teorie?" *Sistemi Intelligenti*, XIII : 171-188.
Origgi, G. ; Sperber, D. (2000) " Evolution, communication and the proper function of language" in P. Carruthers, A. Chamberlain (eds.) *Evolution and the Human Mind*, Cambridge, Cambridge University Press : 140-169.
Pinker, S. (1994) *The Language Instinct*, New York, Penguin.
Sperber, D. (2000) "Metarepresentations in an Evolutionary Perspective" in D. Sperber (ed.) *Metarepresentations*, Basil Blackwell, Oxford.
Sperber, D. ; Wilson, D. (1986), *Relevance, Communication et Cognition* Basil Blackwell.

Discussion

More syntax, less inference
Jean-Louis Dessalles
Apr 14, 2004 12:48 UT

In their paper, Gloria Origgi and Dan Sperber suggest that various aspects of the language faculty, like the ability to form predicates and to express them through syntax, developed to influence others' mental states, thus creating an efficient communication system that relies on inferential capacities. I would like to question the plausibility of this scenario.

The authors assume that inferential communication presupposes the ability to represent the mental states of others, and that syntactic language is an improvement because it allows for more inferencing, as when arguments are left unspecified in a predicate. These two statements aren't obvious.

Contrary to what the authors seem to suggest, protolanguage, as defined by Bickerton (1990), heavily relies on inference to be intelligible to hearers. Proto-utterances like 'hole--water' or 'house--fire' may mean that there is a leak in the water tank or that there is a fireplace in that house. Their interpretation requires that the context strongly constrain the universe of possible inferences. In this respect, the advent of cognitive predicates and of linguistic syntax considerably limits the need for inferential interpretation.

The authors are right to say that human communication is not a word for word translation between ostensible signs and internal meanings. But I doubt that we are the first species to enjoy inferential communication. Even a mere alarm call is context-sensitive and requires interpretation. If protolanguage has ever been used by some homo species, it must have been to refer to proximal situations, where the context could channel the diversity of possible inferences that were available to these ancestors.

I strongly support the authors' concern to look at the evolution of language at the pragmatic level. From that perspective, the advent of the ability to form predicates, and the correlative advent of syntactic abilities,

should indeed be seen as a qualitative change. However, rather than seeing there some kind of evolutionary "progress" due to increased inferential power, I proposed elsewhere (Dessalles 1998, 2000a, 2000b) that they fulfilled a new function, argumentation, that emerged as a way to detect liars.

References

Bickerton, D. (1990). Language and species. Chicago : University of Chicago Press. Dessalles, J-L. (1998). "Altruism, status, and the origin of relevance". In J. R. Hurford, M. Studdert-Kennedy & C. Knight (Eds), Approaches to the Evolution of Language - Social and Cognitive Bases. Cambridge : Cambridge University Press, 130-147. http://www.enst.fr/~jld/papiers/pap.evol/96122602.ps.gz Dessalles, J-L. (2000). Aux origines du langage - Une histoire naturelle de la parole. Paris : Hermès. http://www.enst.fr/~jld/papiers/pap.evol/99111703.html Dessalles, J-L. (2000). "Two stages in the evolution of language use". In J-L. Dessalles & L. Ghadakpour (Eds), Proceedings of the Third Conference on the Evolution of Language. Paris : E.N.S.T. 2000-S-002. http://www.enst.fr/~jld/papiers/pap.evol/00012401.html

Reply to Jean-Louis
Dan Sperber
Apr 15, 2004 16:46 UT

The title of Jean-Louis' comment is justified by a remark he makes in passing: "the advent of cognitive predicates and of linguistic syntax considerably limits the need for inferential interpretation." This may be true if you compare the amount of inference involved in interpreting the same message (e.g. that the hearer should pour some water on the head the speaker) expressed in two version, one, say, proto-linguistically ("water!"), the other with the full power of modern-human languages ("Please, could you pour some water on my head"). Both involve inferences, but not the same ones and the linguistic sentential expression constrains more its interpretation. However, thanks to the power of their languages, modern humans communicate much more information and much more complex information than they did their Bickertonian ancestors. Therefore the inferential load involved in human communication is likely to be much higher, just because communication, which is still inferential, is much, much richer.

Apart from this particular point, Jean-Louis expresses two disagreements, one regarding the role of inference in communication in general, the other regarding the role of syntax in inferential communication. Only the second is a genuine disagreement. The first one is more of a misunderstanding.

We borrowed the phrase "inferential communication" from Sperber & Wilson Relevance: Communication and Cognition 1986/1995. Inferential communication--or to be more technical "ostensive-inferential communication"--refers to a form of communication where the communicator presents evidence of her intention to inform her addressee of something, and where the addressee infers this intention on the basis of this evidence and the context. This inference could be seen as a form of "inference to the best explanation," where the best explanation is a psychological one: a communicative action is explained, as are actions generally, by attributing an intention to the agent.

Of course, there are inferences drawing on the context involved in some forms of communication other than 'inferential communication' proper. To give but one example, the content conveyed by a bee dance is contextually dependent on the position of the Sun in the sky. We agree with Jean-Louis that utterances in a Bickertonian proto-language spoken by our ancestors had to be inferentially understood. We also agree that the interpretation of these utterances "require[d] that the context strongly constrain the universe of possible inferences." However, this kind or contextual restriction is always relative to a given interpretation algorithm or heuristic. The position of the Sun restricts the interpretation of a given bee dance to a single possibility, but only because bees are endowed with the appropriate algorithm to make use of this contextual datum. Our claim is that the inference heuristic involved in the case of human ancestors was a form of mindreading heuristic about the intention of the communicator, i.e. that proto-linguistic communication was a case of inferential communication proper. I don't know whether Jean-Louis agrees with this, but if he does not, it would be useful if he indicated what alternative he favours. Again, saying that interpretation was contextually restricted is describing the problem, not the solution.

There seems to be a clearer disagreement regarding the emergence of a language faculty and of a rich and productive syntax. Yes, we claimed that it evolved because it gave a communicative

advantage (or a series of such advantages along the way, if the evolution was, as seems plausible, in steps) to people already capable of inferential communication, making their communication easier and incomparably richer. (Jean-Louis speaks of “some kind of evolutionary ‘progress’” with “progress” in scare quotes to remind us, rightly, that evolution is not necessarily towards progress in any sense, but surely it often is, and for obvious reasons, no?). Jean-Louis favours what he sees as an alternative explanation, namely that these new features “fulfilled a new function, argumentation, that emerged as a way to detect liars.” Jean-Louis and myself (Sperber 2000, 2001) have independently developed comparable views on the evolution of argumentation, but whereas he sees it as the basis of the emergence of syntactically rich languages, I see the argumentative function as driving only some aspects of language and language use.

Sperber, D. (2000) Metarepresentations in an evolutionary perspective . In Dan Sperber (ed.) Metarepresentations: A Multidisciplinary Perspective. (New York: Oxford University Press). 117-137.
Sperber, D. (2001) An Evolutionary perspective on testimony and argumentation. Philosophical Topics. 29. 401-413

(Both texts are available at www.dan.sperber.com)

Animal communication and human communication
Anne Reboul
Apr 14, 2004 16:04 UT

I find myself in more or less complete agreement with Sperber and Origgi and thus this comment is more an open question than anything else. In the second section of their paper, Sperber and Origgi note that most animal communication is codic in nature, that the systems of signals used by animals "are typically small and highly stable within a given species". This is entirely right, but I want to add something to this description. In an elegant study, Seyfarth & Cheney (1999) have distinguished three types of vocalizations in vervet monkeys: alarm calls, social grunts and a "wrr" signal indicating the presence of another vervet group. The alarm calls are the most clearly symbolic vocalisations, in the sense that they refer to different types of predators. They are also clearly and absolutely innate (they don't change with time and they are used from the time of birth, though infants tend to overgeneralize, producing leopard calls when any fair sized ground mamal is in the vincinity). By contrast, social grunts and the wrr signal are partly innate and partly acquired. Additionally, the production of alarm calls seem more or less automatic and is done whenever a predator of one of five types is identified. The question I want to ask is the following: granted, there is the big question of the evolution of grammar (which Sperber and Origgi have approached through the notion of the evolution of the language acquisition device), but there is also the question — which may as central to the evolution of language — of the production of symbolic signals in the absence of what they refer to. Though no monkey or ape species seem to have developped them, language trained apes have demonstrated an ability to use symbols to refer to absent (though not to non-existent) objects. The question is on what this ability (which is much more developped in humans who use symbols not only to refer to absent objects but also to non-existent objects) is based? Should one link it to a mindreading mechanism of some sort, given that though all alarm calls elicit more or less the same reaction, the reaction to even a single word used flexibly will depend on what the hearer supposes the speaker's informative intention to be? For instance, "leopard" would elicit flight if the hearer takes the speaker to indicate the presence of a leopard here and now, but would merely be an incitation to use another path if the speaker is taken to indicate the presence of a leopard on the intended one. So that's my question. Should we link that flexibility in meaning to the ability of using a signal in the absence of its referent and, if this is the case, what is the link? And what is the link with mindreading?

References Seyfarth & Cheney (1999), "Production, usage and response in nonhuman primate vocal development", in Hauser & Konishi (eds), The design of animal communication, Cambridge, MA, The MIT Press.

referring in absentia
Dan Sperber
Apr 15, 2004 15:09 UT

Thanks, Anne, for the comment and the question. Here is a tentative answer. One should distinguish a general ability to use any symbol in one’s repertoire to refer to an item denoted by this symbol, whether the item is present or not, from the much narrower ability to use specific symbols in specific circumstances to refer to absent objects. The latter ability is found in non-human animals, e.g. in bees who communicate information about a source of pollen that is out of sight. The former, more typically

human communicative ability is presumably based on the prior human cognitive ability to think about absent objects, to imagine them, to remember them, and so on. Memory power seems to me the main basis for such a cognitive ability. Possibly, other species have the ability to entertain thoughts about absent items, but they are not able to communicate these thoughts. Humans can think, whether *in praesentia* or *in absentia*, about anything that they can categorise, and moreover, they can communicate about anything they can categorize and think about. I am not convinced that mindreading should be involved at all.

Language and naive psychology in ontogeny
Gil Diesendruck
Apr 18, 2004 13:40 UT

Sperber and Origgi make a convincing argument that human communication was able to develop and aquire its power due to the interaction between language and naive psychology. What I would like to raise in this commentary is how does this interaction unfold in development? Putting it bluntly, how does the infant brain know -- or come to know -- that it should use its naive psychological mechanisms in order to interpret language? Is this routing instinctive and thus automatic? Or is there something about early communicative interactions that prompt the brain to take this route?

It seems to me that infants indeed need to "discover" that the way into language understanding is through mindreading. They have to understand that what they employ to interpret human behavior in general, as shown by the work of Baldwin, Meltzoff, Woodward and others, they need to extend to this other form of human expression -- namely, speech. Note, however, that even the first mapping -- between mindreading and human behavior -- is not trivial (e.g., the work of Gergely & Csibra). The extension onto human language might be even less so.

For instance, recent studies have shown that while 13- to 18-month olds are quite flexible in terms of the symbols (e.g., gestures or sounds vs. words)they accept and remember as referring to objects, by 20- to 26-months of age they begin to priviledge words over other symbolic means (Namy & Waxman, 1998; Woodward & Hoyne, 1999).

The implication for the current discussion about the evolution of language and naive psychology, is that given the non-trivial coordination of these capacities, we might need to look for some mechanisms capable of explaining how the mapping occurs so rapidly, efficiently, and universally.

References:

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naive theory of rational action. Trends in Cognitive Sciences, 7, 287-292.

Namy, L.L., & Waxman, S.R. (1998). Words and gestures: Infants' interpretations of different forms of symbolic reference. Child Development, 69, 295-308.

Woodward, A.L., & Hoyne, R. (1999). Infants' learning about words and sounds in relation to objects. Child Development, 70, 65-77.

...and ostensive-inferential communication
Gloria Origgi
Apr 22, 2004 0:08 UT

Gil Diesendruck raises the interesting and difficult question of the interaction between language and mindreading in development. Although we do not have a detailed answer, let us try to sketch the picture we have in mind (see also our answer to Peter Dominey). Let us distinguish three types of cognitive achievements typical of children:

(1) Attributing mental states to others (in fact, depending on the type of mental state --desire, belief --, its complexity, its epistemic properties -- truth and falsehood --, a series of distinct cognitive achievements is involved).

(2) Attributing to the producer of an ‘ostensive stimulus’ a communicative intention. (An ostensive stimulus, as described in relevance theory, is a piece of behaviour that is overtly aimed at attracting

the attention of others in order to influence their mental states by conveying some content. Typically, ostensive behaviours are recognisably intentional behaviours that have no obvious physical goal and that are best explained by seeing them as intended to influence others.)

(3) Recognising a linguistic utterance as an 'ostensive stimulus'.

We speculate that each of these three achievements has a distinct innate basis. In particular we assume that there is an evolved pragmatic module specialised in the recognition of ostensive stimuli and in the inference of the communicative intentions that caused the production of such stimuli. So, even if communicative intentions are fairly complex mental states, children don't have to reach a stage in their mindreading abilities where they can generally interpret all mental states of such a complexity before they can interpret communicative intentions. They have a special readiness for communicative intentions.

We also assume that infants are (or soon mature into being) disposed to pay special attention to linguistic utterances and to recognise that these utterances are intended to grab their attention, i.e. are ostensive-inferential stimuli. So, they don't have to discover that linguistic utterances call not just for decoding but also for a mindreading kind of inferential interpretation.

We are suggesting, in other terms, that it need not be the case that children have to "discover" (at least not in the 'little scientist model' sense) how to extend mindreading to the case of language. In any case, the extension is not in one but in two steps: mindreading to inferential comprehension of ostensive stimuli, and recognition of speech as ostensive stimulus. These two extensions may have occurred, or at least may have been prepared, over evolutionary time, leaving to appropriate cognitive inputs little more than the role of triggering in due time (here maturational factors come into play) the activation of several distinct modular abilities.

Gil concludes; "*The implication for the current discussion about the evolution of language and naive psychology, is that given the non-trivial coordination of these capacities, we might need to look for some mechanisms capable of explaining how the mapping occurs so rapidly, efficiently, and universally*." We agree, with the qualification that not two, but three abilities are involved: inferential comprehension (not limited to linguistic stimuli) has to be added to general 'naïve psychology' or mindreading, and language. We would also argue that such 'rapid, efficient and universal mapping,' as Gil describes it, strongly suggest an explanation that is, in good part, evolutionary and nativist.

Gloria and Dan

Proto inferential communication?
Peter Ford Dominey
Apr 20, 2004 15:39 UT

If I infer correctly, then it appears that the inferential communication mechanism described by Sperber and Origgi is of sufficient power to eliminate (or significantly reduce) Quine's problem of referential ambiguity. Indeed, Mutant Marie appears to be able to divine from Parched Pierre's utterance (in context) more than the intended meaning, and right on target. Via this inferential communication mechanism the hearer can know what the speaker means almost without a need for decoding. Can we thus consider that the problems of referential ambiguity will be significantly reduced by these inference mechanisms? If so, this has striking implications with respect to language learnability.

As suggested in the conclusion, then, language evolution and development should take the power of these naive psychology and associated inference mechanisms seriously. This is quite fitting with the topic of the conference "Coevolution of language and theory of mind". In this context, I wonder if the authors could provide a bit of (perhaps speculative) filling in of the gaps of the evolutionary (and developmental) trajectory that leads to this pre-linguistic inferential communication mechanism. For example, would they image a proto- inferential communication mechanism in which only physical action goals could be inferred and attributed to others, with the progressive introduction of more advanced aspects (intentions, beliefs, etc.); or rather do the authors imagine that there is a more radical jump or transition with a full blown inferential communication (and mind-reading) capability appearing from the start? Clearly a full response to this question may be beyond the space limitations of the reply text, but a schematic outline of progressive steps would be most useful.

Reply to Peter Dominey

Dan Sperber

Apr 22, 2004 0:35 UT

Peter writes: "*If I infer correctly, then it appears that the inferential communication mechanism described by Sperber and Origgi is of sufficient power to eliminate (or significantly reduce) Quine's problem of referential ambiguity*." Yes indeed. The inferential communication mechanism we assume is the one described within relevance theory. With this mechanism, in a great variety of cases, "*the hearer can know what the speaker means almost without a need for decoding*" and "*this has striking implications with respect to language learnability*." This was not the place to recapitulate the basics of relevance theory (see Wilson & Sperber 2004), but let us speculate in answer to to Peter's more specific question about the evolutionary trajectory involved (see also Origgi & Sperber 2002, Sperber 2000, Sperber & Wilson 2000). (Regarding the developmental trajectory, see our answer to Gil Diesendruck)

We accept the idea that mindreading evolved as a form of social intelligence in a species whose social life was gaining in complexity (more or less the 'Machiavelian hypothesis'). It evolved with functions other than communication (e.g. prediction of others' behavior, exploitation of others, protection from exploitation by others, tacit coordination of action). We assume that this ability developed beyond the mere attribution of physical action goals to the attribution of mental states (true 'mindreading'). This made it possible

1) for individuals to intend to influence others' mental states,

2) for individuals to realise that others were attempting to influence their mental states,

3) for these attempts at influencing others to be performed overtly as a form of (at least apparently) helping behaviour,

4) for such attempts at benevolent influence to be recognised and accepted.

In other words, mindreading made possible what in relevance theory is known as ostensive-inferential communication ('ostensive' because the communicator is making manifest her intention to influence her audience, 'inferential' because the audience must infer from the communicator's behaviour and the context what content the communicator is trying to convey). Communication is a secondary function of mindreading. Of course, as long as it was based on a general mindreading ability, intending and understanding ostensive-inferential communication involved fairly elaborate inferences, and only rudimentary contents could be reliably communicated in this way. However (and that is one of the main points of relevance theory) communicative intentions exhibit specific regularities (linked to the fact that they convey a presumption of their own relevance) that allow for a quite effective communication-specific inferential procedure (as opposed to a more general mindreading ability). We speculate that a pragmatic module realising this specific inferential procedure evolved, and in fact co-evolved with the language faculty. The latter was adaptive only in a species capable of fairly fluid ostensive inferential communication.

If something deserves to be called "proto-inferential communication" is is early inferential communication, based just on general mindreading abilities and benefitting neither from a dedicated module nor from a genuine language faculty. On the other hand, before genuine mindreading, you cannot have inferential communication, not even "proto".

Origgi, G. ; Sperber, D. (2000) " Evolution, communication and the proper function of language" in P. Carruthers, A. Chamberlain (eds.) Evolution and the Human Mind, Cambridge, Cambridge University Press : 140-169.

Sperber, Dan (2000) Metarepresentations in an evolutionary perspective . In Dan Sperber (ed.) Metarepresentations: A Multidisciplinary Perspective. (New York: Oxford University Press). 117-137.

Sperber, Dan & Deirdre Wilson (2002) Pragmatics, Modularity and Mind-reading. In Mind and Language. 17. 3-23.

Wilson, Deirdre & Dan Sperber (2004) Relevance Theory. In (Horn, L.R. & Ward, G. eds.) The Handbook of Pragmatics. Oxford: Blackwell, 607-632.

(all these texts are available at www.dan.sperber.com)

Dan & Gloria

Not quite satisfied yet
Anne Reboul
Apr 23, 2004 13:49 UT

Of course, memory is a crucial factor in being able to think about absent objects, and that may be an ability present in both human and non-human animals, with the only difference that the first can and the second cannot communicate these thoughts about absent objects. However, the question remains: maybe such species as for instance vervet monkeys can communicate in the rigid way common to most (if not all) animal and non-human communication system, but, supposing that they are able to think about absent objects (to remember them for instance), or more precisely, about absent predators, why aren't they able to communicate, even in a very simple way, about these thoughts? According to you, the basis for the human ability to use symbols to to refer flexibly in absentia is mainly based on memory. So let us suppose that vervet monkeys have the kind of memory that enables them to think about things in absentia (let me add that it's not clear whether pure semantic memory would or would not be enough for that) and that they have a symbolic communicative system enabling them to refer to different types of predators (as we know they have). Why shouldn't they use their symbolic system to refer to predators even when they are not present? One possible answer is that using the symbol for, e.g. leopard, if neither the communicating monkey nor its conspecifics have a theory of mind, will leave it entirely open whether the monkey is merely performing an alarm call or communicating a (possibly fairly complex predicative) thought about an (absent) predator. In other words, to be able to use the symbolic signs of an animal communication system in this "detached" way, i.e. when the referent is not present, a (possibly primitive) form of mindreading ability should be present. Let me add that though the presence of episodic memory in non-human animals is the subject of intense controversy right now, the possession of semantic memory by animals is not: everyone agrees that non-human animals in a wide range of species from birds to primates have it. So this leaves open the question of why animals that do possess a symbolic communication system and who have semantic memory should not use their communication system in a flexible way. Note that I'm not claiming that theory of mind is the only possible answer — remember that I'm doubtful about mindreading abilities in non-human animals, though open to persuasion, given some convincing evidence — but I'm merely pointing out that this is just as much of a mystery in the evolution of language as is syntax.

Let's try harder then
Dan Sperber
Apr 23, 2004 15:35 UT

Suppose it were generally advantageous for vervet monkeys to communicate about, say, a past leopard alarm (a speculation not so easy to flesh out in a realistic manner by the way). Then there would have been selective pressure for the evolution of such an ability. What would it have taken, at the cognitive level, for such an ability to evolve? Here we can contrast two views.

According to a minimalist view I would defend, all that would be needed is the emergence of an ad hoc signal-type meaning leopard-alarm-in-the-past (or something a bit more specific but of roughly this kind of tenor). When monkey A would emit this signal, receptor monkey B would be put in the same cognitive state as the one she would be in if she were herself remembering a leopard alarm. Just as a regular alarm call provides a kind a perception by proxy, a memory-of-alarm call would provide a kind of memory by proxy (as does for instance, in another species and context, the bee dance).

According to the view Anne is considering, the ability to communicate about a past event would involve something much richer, possibly mindreading or something equally complex. Why assume anything of the sort? It may seem that communicating about the past must involve greater flexibility than communicating about the present. As the bee's example shows, this need not be the case. If, on the other hand, vervet monkeys could combine symbols to express a variety of complex thoughts, then indeed they would need another kind of mind. Plausibly too, if they were able to express such complex thoughts, these would be not just about the present, but also about the past. But the crucial factor explaining why their mind would have to be different is the richness and flexibility of communication, not the ability to refer *in absentia* (or so I surmise).

Two Aspects of Linguistic Communication
Michael Arbib
Apr 24, 2004 19:08 UT

I agree with Gloria Origgi and Dan Sperber with the need for an inferential model that explains how the same sentence can be used to communicate an indefinite number of different messages dependent on context (including the mental state of both speaker and hearer). However, I think they are mistaken in rejecting the "code model", rather than seeing it as a necessary ingredient of the inferential model. They state "Communication is achieved by giving the hearer evidence of the meaning one intends to communicate. This evidence can be of any sort—gestures, mimicry, showings; and they can be coded or not, provided that they allow the hearer to infer the speaker's meaning." I agree. But why then do the authors invest so much effort in downplaying the role of shared meaning. They give the example

Pierre: I'm beat!

Marie: Ok, let's go back home.

and then comment "It is of little importance whether the meaning that Pierre and Marie associate with the word 'beat' is the same. It may be that, for Pierre, 'beat' means an extreme fatigue, while for Marie, 'beat' is simply a synonym of 'tired'." This makes the point that the meaning in the heads of speaker and hearer need not be identical if they are to ground successful inference by the hearer. But if Pierre had said "I'm a potato", Marie would not infer that as a sign of fatigue but perhaps of excessive drinking – a comment which makes the point that, yes, meaninglessness within a context can itself carry a message, but also reinforces the crucial part that "codes" share in grounding communication. Indeed, Origgi & Sperber say "Quite commonly, a fragmentary, ambiguous and loose coding is sufficient, in the context, for indicating a complete and unequivocal meaning." But again, they feel the need and have considered even less the precise role of communicational processes in linguistic communication, as in this example:

Pierre: Can you fix my watch?

Watchmaker: That will take some time.

They comment that "If Pierre expects that the time for repair will be at least one week, he will understand 'It will take some time' as meaning that the repair will take several weeks. If the watchmaker, for his part, thinks that Pierre expects the repair to be done the same day, he will express himself as he did in order to say that the repair will be a matter of days rather than hours. According to the inferential model, the near identity of the interlocutors' codes is not necessary in order for them to best communicate. Certainly, but the fact is that speaker and hearer have a shared meaning for the construction "That will take some time" - without denying that it is relative to context. If Pierre did misheard the word "time" as "dimes" (a US coin) then he would take the Watchmaker's expression as synonymous for "That will take some money" - meaning that the repairs might be more expensive than Pierre would have expected.

To summarize: the idea that "inference" and "coding" are alternatives in the study of communication is mistaken. Whether in the evolution of language, or in the child's acquisition of a language, the acquisition of the various "codes" associated with each word and the inferential strategies for using them, go hand in hand.

One final comment. The last sentence before the Conclusions of the article reads "Being disposed to treating uncoded communicational behavior as a coded signal may facilitate inferential comprehension of the communicator's intentions and lead to the stabilization of this kind of behavior as a signal." Does this mean that, despite all the arguments marshaled against the coding theory, Origgi & Sperber do concede that in a stable communication system, coding of more or less shared meanings and the inference of intended meanings must indeed go hand in hand?

Articulating the two aspects
Dan Sperber
Apr 24, 2004 23:56 UT

We thank Michael Arbib for his comment, and, as he will see, we hardly disagree. In particular, we agree with the points he makes regarding our examples. The role of these examples was not to show that linguistic meaning does not matter. It was to show that identity of meaning in the idiolects of the speaker and hearer is neither necessary nor sufficient to secure understanding (and this in quite

ordinary conditions – we are not suggesting fancy Gedankenexperiments). So where does the disagreement come from? At least in good part it comes from the fact that we presumed wrongly that the contrast between the code model and the inferential model would be understood as described in *Relevance: Communication and Cognition* (Sperber and Wilson 1986; Revised edition 1995).

The general point is this. Everybody working today on language use agrees that verbal communication involves both coding-decoding and inferential processes. The issue is how are these two types of processes articulated.

According to the "code model", the basic mechanism of human communication is a coding-decoding one, and inference plays at best an ancillary role. Basically, what inferential processes do, according to this model, is permit short-cuts: it is unnecessary for the speaker to fully encode her meaning (although, if needed, she could always do so). She can use pronouns rather than description, she can be elliptical, she can be vague or metaphorical and expect nevertheless to be understood, because she can trust that the hearer will be able inferentially to reconstruct the parts and features that she did not encode.

According to the "inferential model", communication is a special form of mindreading done with the help of the communicator who makes it manifest that she intends to have her mind read (or to be more precise, to have her meaning read, where meaning is a kind of intention). For this, the communicator must provide evidence of her meaning. This does not necessarily require the use of coded signals. But of course coded signals, and in particular the richly structured signals permitted by human languages, provide the best kind of evidence. So we do not deny that human languages are codes or that human communication crucially depends on these codes to convey the extraordinarily rich and diverse meanings humans manage to convey. The claim of relevance theory is that the function of the linguistic code in verbal communication is not to encode the speaker's meaning but to give evidence of this meaning. What we tried to do is look at some of the evolutionary implications of such a claim. More on this in our reply to Michael Arbib's second message.

Dan & Gloria

Examining the Assumptions of an Evolutionary "Just So" Story.
Michael Arbib
Apr 24, 2004 19:13 UT

The evolutionary "Just So" story offered by Gloria Origgi and Dan Sperber may be a useful addition to our collection. However, it makes a number of unexamined assumptions and fails to take certain evolutionary considerations into account.

They assert "According to the inferential model, the near identity of the interlocutors' codes is not necessary in order for them to best communicate." but (see my other commentary, Two Aspects of Linguistic Communication) successful communication needs not only inference but a set of sufficiently shared meanings to ground that inference. They infer from the previous assertion that "In these conditions, a mutation affecting the language faculty and causing the mutant's grammar to diverge from that of her interlocutors is not necessarily detrimental to her ability to communicate."

We already have an unwarranted assumption:

1. That inference needs no grounding in a set of (more or less) shared meanings. I would add that "inference" (in the informal sense of drawing conclusions from partial data rather than applying a logical argument) is an essential part of animal cognition long antedating the evolution of language - consider the decision as to whether that waving of the grass is caused by wind, predator, or prey - and so can be assumed as part of the grounding of any theory of the evolution of human language and cognition.

But there is also an unexamined assumption:

2. Key changes in grammar are due to mutations

The alternative is to consider that biological changes that made language possible were no more related to the structure of grammar than to, say, the ability to use Arabic numerals. We need to analyze carefully the

tradeoff between brain mechanisms that evolved to support aspects of grammar and those that, e.g., make learning certain grammatical innovations (due, e.g., to grammaticization) possible without requiring their encoding in the genome.

Origgi and Sperber ask us to imagine "a proto-language having only word-size sound-meaning pairs, without any syntactic structure." I accept their conclusion that "such a proto-language could be of use only to beings capable of inferential communication." - but have already offered reasons why this adds nothing to the utility of their particular story. Moreover, their story rests on shared "codes" or "meanings," rather than denying their importance. In any case, we see here another unexamined assumption:

3. Proto-language has only word-size sound-meaning pairs

This may be true (as argued, e.g., by Bickerton, 1995), but Wray (2000) and I (Arbib, 2002) I have offered reasons why protolanguage may have been composed of protowords which may have been closer in meaning to modern-day sentences than to words as we know them. On this account, the predicate is already in the meaning of the protoword - and thus the evolutionary challenge is not to add predicates to protolanguage so much as to "fractionate" them, giving explicit verbal expression to a predicate hitherto implicit in a set of protowords. My point here is not to argue for my theory, but to stress the importance of making our assumptions explicit rather than acting as if their plausibility needed no discussion.

In any case, when the authors ask us to imagine "that Marie was in fact a mutant whose language faculty … had allowed her as a child to analyze the words of the proto-language that she was in the process of acquiring, either as arguments or as one- or two-place predicates" it would be helpful to augment the story to render it plausible that such a mutation should occur. This raises a general problem:

4. When we provide an evolutionary account, we may postulate a series of changes, with each change occurring as a result either of biological evolution or of cultural innovation. For each change we posit, then, we need to carefully weight the two alternatives, and assess whether such a change is indeed plausible.

One final comment: Origgi & Sperber conclude that "the language faculty and human languages, with their richness and flaws, are only adaptive in a species that is already capable of naïve psychology and inferential communication." At some level I agree, but I wonder how useful it is to talk of being "already capable of naïve psychology" rather than teasing apart what aspects of such a psychology are relevant at each stage of evolution of inferential communication, and entertaining the notion that neither language nor naïve psychology are indivisible, and that they too may have co-evolved in complexity.

Assumptions and misunderstandings

Dan Sperber
Apr 25, 2004 0:11 UT

1) Michael Arbib is obviously right in saying that inference "*is an essential part of animal cognition long antedating the evolution of language*". Nothing we said implies otherwise. However, there are many different forms of inference. According to relevance theory, which we follow here, the inferential procedures involved in comprehension are quite domain-specific. Our point is not the trivial one that inference in general must have played a role in the evolution of language, it is to argue that language evolved because linguistic utterances provide an optimal kind of evidence for inferential comprehension.

2) Yes we do assume that language learning is based on a domain-specific biologically evolved language faculty, that the range of possible human languages is both opened and constrained by this faculty, and that steps in the evolution of the language faculty (if there were any such steps) were likely to change the range of possible human languages. (Of course standard historical changes in grammars are not due to changes in the language faculty.) This nativist assumption is "*unexamined*" only in the sense that we did not examine it here. Otherwise, it is one of the most carefully examined assumption in cognitive science, having been and remaining the object of intense controversies. We find the arguments in favour of this assumption compelling, but have no new arguments to offer. For people who are unconvinced, our proposal may be seen as an exploration of a possible way in which a language faculty, if there is one, might have become adaptive.

3) We are not claiming that there has been a Bickertonian proto-language. What we are doing is explaining how, such a proto-language, if there was one, could (A) have been used to communicate,

and (B) have evolved in a syntactically richer system. Our point in doing so is to illustrate the kind of explanatory power that may be found in a pragmatic approach. Suppose Michael Arbib is right and that human protolanguage expressions were more like single-word sentences (a very interesting hypothesis, by the way). Then, at first blush, the pragmatic story would be even easier to tell.

4) Michael Arbib writes: "*When we provide an evolutionary account, we may postulate a series of changes, with each change occurring as a result either of biological evolution or of cultural innovation. For each change we posit, then, we need to carefully weight the two alternatives, and assess whether such a change is indeed plausible.*" Sounds excellent, but isn't the game, at present, way too speculative to impose such demands? If we want to discuss methodology, shouldn't we impose *realistic* demands on speculation, rather that irrealistic demands that are all too likely to result in mere window-dressing, paired with description of alternative accounts as "Just-so Stories"?

5) Michael Arbib writes: "*Origgi & Sperber conclude that "the language faculty and human languages, with their richness and flaws, are only adaptive in a species that is already capable of naïve psychology and inferential communication." At some level I agree, but I wonder how useful it is to talk of being "already capable of naïve psychology" rather than teasing apart what aspects of such a psychology are relevant at each stage of evolution of inferential communication, and entertaining the notion that neither language nor naïve psychology are indivisible, and that they too may have co-evolved in complexity*." Within the limits -- length constraint in particular (but see Origgi & Sperber 2000, Wilson & Sperber 2002) -- of this web conference, we were trying to contribute to teasing apart various aspects of naïve psychology, and their different role in the coevolution of language and mindreading.

Origgi, Gloria & Dan Sperber (2000) " Evolution, communication and the proper function of language" in P. Carruthers, A. Chamberlain (eds.) *Evolution and the Human Mind*, Cambridge, Cambridge University Press : 140-169.

Sperber, Dan & Deirdre Wilson (2002) Pragmatics, Modularity and Mind-reading. In *Mind and Language*. 17. 3-23.

Comment on Origgi and Sperber: Context and inference are ubiquitous

Petr Kotatko
Apr 25, 2004 15:25 UT

Although I agree with many of the points emphasized in this paper, I think it sets up a false dichotomy between two models of communication - "code" and "pragmatic" - models, and compounds the problem by thinking that animal systems are of the former, and only language the latter, type. The problem with this dichotomy is that context and inference play a key role in ANY adequate model of communication, regardless of whether this is stressed by the model (as in pragmatics) or left implicit (as in "classical" models).

To see why, take a classic "coded" example: a digital photographic image encoded in a computer file. To be unpacked (e.g. to project or print these bits as an image), the binary digit code must be interpreted in a specific way, determined by its file type and the program decoding it. Thus, a JPEG file and a TIFF file of "the same" image are quite differently organized, have different headers and bit encodings, and thus have a very low overlap in the actual bits stored on disk. It is only when they are interpreted relative to some independent criteria, stored in the program, that they can be projected as "the same" image. Thus even in a prototypical "code" situation, interpretation is context dependent. We do not need pragmatic implicatures or human theory of mind to already run into this problem (though they certainly highlight its importance).

True, the role of such contextual information is not highlighted by contemporary information theory, so this type of context-dependence has not been properly mathematically formalized (at least to my knowledge). This paper, along with other work by Sperber and colleagues, properly highlights the need for this aspect of communication to be more rigorously explored. But in a final, adequate theory of information and communication the pragmatic and classical approaches are complementary, not contrasting as presented in this paper.

Another false dichotomy is between human and animal communication, with the latter seen as fitting the "code" model. Take for instance an extremely simple example: the honeybee dance system. Honeybees signal the location of various resources (flowers, water, nest sites) with a spatially- and physically-encoded

movement pattern that reveals the distance and angle (relative to the sun) to the resource being indicated. If the honeybee is outside (an infrequent but possible case) it will use the actual sun's current direction as the referential context for its dance vector. More typically, in the darkeness of the hive, the dancer will use gravity (downward vector) as the reference, completely adjusting its dancing to this new reference frame. Recipients of the signal also obviously adjust their interpretation to this change in context. Such points could be multiplied indefinitely: most animal communication systems are quite indeterminate, with the same vocalization occurring in many different contexts and appropriately interpreted in each (baboon grunts are a well-researched example). Thus it is simply not true that all animal systems are "fully coded". The same point might be made about non-linguistic signals such as laughter or sighs in humans. I believe the data suggest quite strongly that "inferential communication" is the norm among animal communication systems, not a recent human innovation.

An important distinction left implicit in Origgi and Sperber's discussion is that between co-operative and competitive communication. Gricean maxims (and indeed most of pragmatics) assumes the former. Some animal communication systems (e.g. the honeybee example, or parent-offspring communication) are cooperative (usually arising from kin selection); but many of the more impressive animal signals are not cooperative (territorial and mating calls usually are made in a competitive context that does not encourage honesty, broadcasting one's true quality or intentions, etc.). Given this distinction the evolutionary discussion could be thought through more carefully. For example, a mistmatch of codes might be quite adaptive in certain situations (especially competitive situations where callers are attempting song-matching, and failure to do so indicates defeat). More importantly, it is not always the case that a code must be shared by the whole population to be advantageous: a code shared just by a closely-knit group of kin and unintelligible to the rest of the group might be more useful than a shared code, in certain situations. Finally, the statement in the penultimate paragraph: that in a code-based system "every departure from the common grammar will be disadvantageous or neutral" is clearly overlooking a variety of imagineable "mutations" that would help a lot in even a strictly code-based system: anything allowing more rapid parsing, better memory access or more efficient encoding of the same message set could be selected for.

In summary, I am wholly sympathetic to an approach to communication that stresses the importance of context and inference in both human language AND at least some animal communication systems. The above criticisms are intended to constructively aid in placing this argument in its proper comparative, evolutionary and conceptual framework.

Yes but inferential communication proper is rare
Dan Sperber
Apr 25, 2004 19:07 UT

We thank Tecumseh Fitch for his comments. However we feel that they don't have the import he attributes to them.

1) As we said in reply to Arbib, we borrowed the distinction between the code and the inferential models of human communication from relevance theory. The distinction is not between a model that would deny any role to context-sensitive inference in communication and another one that would acknowledge it. It is between a view of linguistic communication where the function of linguistic utterances is to encode the speaker's meaning (allowing some bits and aspects to be inferred rather than decoded when convenient), and another view where the function of linguistic utterances is to provide evidence of the speaker's meaning. According to the code model, inference is an ancillary and, many argue, optional addendum to what is basically a coding-decoding mechanism. According to the inferential model, linguistic encoding-decoding is just the most powerful way of providing evidence of one's meaning in what is basically an inferential process where the inference is about a mental state of the communicator. These different approaches yield divergent analyses for a wide variety of pragmatic phenomena, from the interpretation of quantifiers to that of illocutionary forces. To show that it is a "*false dichotomy*" would take much more than invoking the case of bees that Sperber & Wilson and we have used precisely to illustrate the distinction between context-sensitivity that remains well within the code model, and context-sensitivity that does not.

2) Our contrast is not between animal and human communication per se (if only because humans are animals). It is between mindreading-based communication and other forms. To the best of our knowledge, there is no uncontroversial case of non-human mindreading-based communication, but there is some suggestive data collected among other primates, so the issue is still open. Also, we don't know of any plausible explanation of non-mindreading-based communication that is not in terms of the code model.

3) Relevance theory does not assume that communicators must be cooperative in the Gricean sense. However even antagonistic communicators, if they are at all involved in intentional communication (which need not be the case), with the communicator intending to be understood, and the addressee intending to understand (which need not be the same thing as accepting), are sharing at least one goal: communication itself. If communication itself is not a shared goal of communicators (because at least one of the parties is not intending to be involved in communication, either as a communicator or as an addressee), or if it is not in the interest of one or the other party that communication should succeed, then, indeed, the matching of codes need not be necessary or even advantageous to at least one of the parties. Our point however was that, in inferential communication, the matching of codes is unnecessary (and insufficient) in any case, including cases of cooperation.

4) Fitch writes: "*the statement in the penultimate paragraph: that in a code-based system "every departure from the common grammar will be disadvantageous or neutral" is clearly overlooking a variety of imaginable "mutations" that would help a lot in even a strictly code-based system: anything allowing more rapid parsing, better memory access or more efficient encoding of the same message set could be selected for*". Thanks for the observation. It is indeed correct, providing that the mutation affects the language faculty itself, rather than memory or other non-language-specific capacities that have an impact on the efficiency of communication. Still, the important point, from the point of view of the evolution of the language faculty, is that, in the code model, mutations that would result in a different pairing of sounds and meanings would be detrimental to communication. Not necessarily so in the inferential model.

5) Again, the point of distinguishing the code and the inferential model in the context of this conference is that they attribute different functions to language, and this should be of obvious evolutionary relevance.

Towards an explanation of the evolution of language

Ingar Brinck
Apr 25, 2004 21:30 UT

I agree with Sperber and Origgi that the inferential model better explains the flexibility of natural language than the code model, and that it captures important properties of linguistic communication. However, it seems to focus on the wrong kinds of communication for an explanation of the evolution of language. I think that such an explanation should focus on the most basic acts of symbolic communication, and be guided by a similar principle of parsimony as the one used in explanations of animal communication. That means that appeals to mental states and mind-reading should not be made unless necessary. The primary mode of symbolic communication is, I submit, the non-linguistic act of declarative pointing (DP). DP is intimately related to language acquisition and only occurs in humans. It emerges in children by 12 months, and aims at shared reference by a process that leads to joint attention to the indicated object. Indication is an illocutionary act that may have a variety of perlocutionary effects, depending on the actual properties of the context and speakers. Engaging in the process that in the case of DP leads to joint attention requires the capacity for attention reading, i.e., for having intentional states directed at states of attention, paired with an emotionally based kind of intersubjectivity. Intention reading is not required for either producing or understanding the illocutionary force and the perlocutionary effect of an act of DP. The evidence for the content of an act of DP is manifest in the physical context and the behaviour of the speaker, which reflects her intentions to communicate and indicate. To me it seems that a great deal of contemporary communication relies on similar resources as DP does. The mechanisms that underlie basic, symbolic communication and present-day linguistic communication are similar. Once they were in place, they made possible an expansion of the human mind, and allowed for not only representing external states and one's own internal states, but also the internal states of others. By being scaffolded by language, the human mind grew increasingly bigger. Our present capacity for irony, metaphor, and the like, are extensions of basic, symbolic communication, arising from the prior existence of a meagre version of today's augmented natural language. Natural language emerged as a way of better managing co-operation, and creating new ways of co-operating that would answer to the demands of the environment. It was recruited as a tool for mind-reading, and as such has promoted the mind's progression towards more and more elaborate forms. To conclude, an explanation of the evolution of language should aim at disclosing the fundamental mechanisms of basic communication, and consider more sophisticated forms of linguistic communication as varieties of the language that once evolved.

References.

Brinck, I. (in press). “The pragmatics of imperative and declarative pointing”, Cognitive Science Quarterly.
Brinck, I. (to appear) “Joint attention, triangulation and radical interpretation: A problem and its solution”, Dialectica.
Brinck, I. & Gärdenfors, P. 2003. “Co-operation and communication in apes and humans”, Mind & Language, 18(5), 484-501.

Reply to Ingar

Gloria Origgi
Apr 25, 2004 22:36 UT

Thanks to Ingar for her insightful comment. Let me just post some brief reactions given the little time left before the closing of the debate. I hope that we will have other occasions in the future debates to come back to these issues.

1. We do not claim that an inferential model of communication implies an attribution of mental states at any step of the process. Rather, as we have stressed in our reply to Gil Diesendruck, the interpretation of an ostensive act of communication amounts to the attribution of a speaker’s meaning as a result of an inferential process of maximizing relevance that isn’t necessarily metarepresentational.

2. Although it is very useful to investigate in detail a particular mechanism of “basic symbolic communication” such as declarative pointing, it seems to amount to a case of “ostensive-inferential communication”, in which the recognition of the overt communicative intention is part of the success of the act of communication. This again doesn’t imply a complex attribution of mental states. Rather, we may suppose a readiness in recognizing communicative acts that prompts us to maximize their relevance.

Sperber, Dan (2000) Metarepresentations in an evolutionary perspective . In Dan Sperber (ed.) Metarepresentations: A Multidisciplinary Perspective. (New York: Oxford University Press). 117-137.

Sperber, Dan & Deirdre Wilson (2002) Pragmatics, Modularity and Mind-reading. In Mind and Language. 17. 3-23.

Imitation, Quoting and Theory of Mind

Tecumseh Fitch *(*University of St. Andrews)

(Date of publication: 26 April 2004)

Abstract: Despite ongoing debate concerning the presence of theory of mind capabilities in nonhuman primates, it is clear that 1) human theory of mind capacities are very highly developed, and 2) these capacities are a crucial component underlying pragmatically well-formed communication. Human language clearly benefits from these enhanced theory of mind capacities. I suggest that language (and in particular the capacity for vocal imitation) symmetrically aids theory of mind capacities by providing a "scaffolding" for the representation of others' minds, via quoting. By providing a relatively durable sensorimotor representation of another individual's actions, quoting allows off-line processing of their thoughts and intentions.

Introduction

A key desideratum in the study of the biology and evolution of language ("biolinguistics") is to understand what allows humans to acquire full-blown language, but not other animals, especially our nearest relatives the chimpanzees. Studies of chimpanzees have revealed an impressively broad set of cognitive and communicative abilities, and these empirically-determined similarities are a key source of data for reconstructing the capacities of our last common ancestor with chimpanzees (hereafter, the "LCA") which lived around 6 million years ago. However, a chimp, raised in diapers in a human home, will develop no speech production capabilities, and only limited comprehension of speech. Even with intensive training using gestures or lexigrams, chimps acquire a productive vocabulary of only a few hundred signs, unlike the child who under identical circumstances acquires a lexicon of tens of thousands of words, and a complex syntax that allows generation of an unlimited number of sentences. Taking the reconstructed cognitive capabilities of the LCA as the starting point, then, the goal is to understand what changes in preexisting capabilities, and addition of new capacities, have occurred in the hominid line since our divergence from chimps.

Detailed comparative study has revealed at least three capacities that play a key role in language and differentiate humans from chimps: speech (especially vocal imitation), syntax (especially recursive syntax) and semantics (especially intentional communication). A recent survey of hypotheses about language evolution (Christensen & Kirby, 2003) reveals broad agreement about these three components (although authors differ in their emphasis and terminology), despite ongoing debate concerning many other issues. This consensus list highlights the minimum that biolinguistics should be able to explain in an adequate theory of language evolution, which starts with the LCA and ends with all key components of the contemporary language capacity in place. In principle, each of these capacities could have evolved at different times, and under different selective pressures. Further, the selective value of some capacity (say vocal imitation) might have varied during different stages of hominid evolution. Indeed, most contemporary theorists, have offered "dual-stage theories" positing different selective forces at different phylogenetic stages (e.g.,Arbib, 2003; Bickerton, 1998; Corballis, 2003; Donald, 1993). Current consensus thus isolates three key components to be explained, and posits a multistage phylogenetic process to attain them.

Although speech, syntax and semantics are profitably separated for conceptual clarity, they obviously represent closely-linked capacities, and their interactions are as crucial as their individual function. Here I explore some of these interactions from a comparative, evolutionary viewpoint. To the extent that synergies can be identified between components, coevolutionary accounts of language evolution gain in plausibility. After a brief comparative review of speech and syntax, I focus on semantics, especially intentional communication in relation to theory of mind. Intentionality in the human sense does not appear to be present in the communication systems of nonhuman primates (Seyfarth & Cheney, 2003; Seyfarth & Cheney, in press), despite a widespread current assumption (based on misinterpretation of vervet alarm calls) that alarm calls are homologous to words. Second, while theory of mind abilities in chimpanzees remains an area of intensive investigation and debate, it is clear that human children go far beyond chimps in this arena by the age of three or four, and thus that this capacity also has either appeared, or hypertrophied, since our evolution from the LCA. I then suggest 1) that vocal imitative skills provide a "scaffolding" for theory of mind abilities, via the mechanism of quoting, and 2) that the capacity for recursive embedding of minds within minds (high-order theory of mind) may be closely related to recursive embedding in syntax, itself related to quoting.

Speech and Syntax

Vocal imitation, the ability to flexibly recreate novel utterances after hearing them produced by another

individual, is a well-developed skill in human children by the age of three, and provides a crucial prerequisite for spoken language. Without this imitative ability, an indefinitely-extensible lexicon would be impossible. Although well-developed vocal imitation is observed in many nonhuman species, including birds, cetaceans (whales and dolphins), seals and perhaps bats (Janik & Slater, 1997), it has not been observed in any nonhuman primate, including chimps, despite intensive investigation. While some birds can imitate arbitrary sounds, including human speech, chimpanzees fail to do so even with intensive training. Indeed, primate vocalizations appear to be relatively involuntary (though audience effects exist, showing that primate calls are not entirely stimulus-driven as sometimes stated (Cheney & Seyfarth, 1990b)). In addition, primate calls are highly genetically canalized: many are present at birth, and will reliably develop in the absence of any relevant input (Winter et al., 1973; Hauser, 1996). Thus primate calls appear to resemble human laughter or crying (also present at birth, even in deaf children) more than human words. This is as true for vervet alarm calls as for any other primate call: the structure and even some aspects of the meaning of these calls appears to be genetically-determined (Seyfarth & Cheney, 1980). This is one reason that the facile equation of vervet alarm calls with human words must be avoided: the only enlargement of the vervet call "lexicon" must be accomplished by natural selection over millennia, not during individual ontogeny as in humans. Thus vocal imitation, and the lexicon that goes with it, is a crucial component of language that has evolved since the LCA (note that the existence of non-vocal imitation, equally important for signed languages, does not change this necessity).

Syntax in modern human languages is highly complex, and entails several potentially independent innovations. At the simplest level, the ability to construct and process phrase structure, independent of meaning, is a crucial component of our ability to make "infinite use of finite means" in language: embedding phonemes within syllables within words within sentences, we can use a small "alphabet" to generate an unlimited set of signals. Production of correspondingly flexible phrase structures is not observed in nonhuman primate calls, nor is it available at the perceptual level by those primates tested so far (Fitch & Hauser, 2004); however, phrase structure may be present in the "songs" of birds and humpback whales.. A second aspect of syntax, recursion, is more restricted and more powerful. Recursion entails phrase structure, but goes beyond it in that structures of the same type may be embedded within one another. Recursion is not present at the phonological level (we cannot embed syllables within syllables) but is clearly present at the sentential level (we can embed phrases within phrases). We currently have no evidence for such recursive embedding within animal communication systems (Hauser et al., 2002). Finally, long distance dependencies (connections of elements between sentences, or between multiple phrases in a sentence), as in anaphora, has no known parallel in animal communication. However, detecting such semantically-defined dependencies in non-semantic vocalizations like birdsong may be impossible. Below I focus on recursive syntax, because it has not yet been observed in animal communication systems, but could be in principle (a mockingbird could "quote" the phrase-structured utterance of another bird within its own phrases, or a humpback whale the song of another male).

Semantics and Theory of Mind

A virtually universal aspect of human languages is symmetry between signaler and recipient: if you can make a signal with a certain meaning, then you can perceive the same correspondence, and vice versa. This symmetry or "parity" is central to both word meaning and Gricean maxims in pragmatics. Speaker/listener symmetry seems so natural to us that it has taken many years to recognize that it is not necessarily present in other communication systems. A key example comes from experimental work on monkey calls (Cheney & Seyfarth, 1990a). Japanese macaques have food and predator alarm calls which are reliably produced in the presence of food or predators, and which appear to be appropriately perceived by others. In these experiments, a mother monkey was shown food or a predator in a separate test enclosure, while her infant was absent. Then the infant was introduced into the test enclosure. If the mother is aware of her offspring's ignorance, we predict an increase in her food or alarm calling when her infant enters. No such increases were observed. Furthermore, mothers produced virtually no alarm calls in the "infant with predator" context. Given the strong inclination of primate mothers to protect their offspring, and the strong evolutionary incentive to do so, these results powerfully suggest that monkeys are unable to represent ignorance in their own offspring, or to vocalize in a way that remedies such ignorance. In contrast, monkeys have a flexible and well-developed interpretive ability (on the part of receivers), easily learning proper interpretation of novel calls including those of other species (Seyfarth & Cheney, 1990). The fact that the infant clearly would respond appropriately to alarm or food calls, if they had been made, makes the absence of calling even more striking.

The strong conclusion to be drawn from this and similar research is that there is an asymmetry in monkeys' perception and production of calls. In particular, signalers do not appear sensitive to the knowledge or ignorance of intended listeners, and thus appear unable to intentionally inform others. The signaler reacts vocally (modulo some audience effects) to its perceptual world, irrespective of its listener's knowledge, and

most of the communicative work is done, inferentially, by the perceiver (Seyfarth & Cheney, in press; Seyfarth & Cheney, 2003). Thus, functional referentiality of alarm calls (shown by listeners' proper reaction to played-back calls) appears different in kind from the referentiality of human language. The producer's half of the Gricean bargain is not upheld.

This is perhaps less surprising given the paucity of evidence for theory of mind in primates (including chimpanzees). In the classic "knower-guesser" paradigm, where a subject desiring food must use cues from individuals, who either could or could not know where the food is hidden, primates consistently perform at chance (Povinelli et al., 1990; Povinelli et al., 1991). In a striking example, chimps fail to choose correctly when one of the two "informants" has a bucket over their head. This and other evidence led (Tomasello & Call, 1997) to conclude that absence of even rudimentary theory of mind capacities was a key (perhaps the key) difference between humans and chimps. However, new data has moderated this view. In particular, when placed in a competitive situation, chimps do behave as if they understand that "seeing is knowing", and are able to adjust their competitive behaviour accordingly (Hare et al., 2000). Thus, the problem may not be a complete inability to represent other minds, but a missing propensity to interpret them cooperatively. Again, the kind of collaborative exchange of information that is at the heart of human linguistic communication appears absent in other primates (interestingly, domestic dogs DO collaboratively interpret seeing as knowing (Hare et al., 2002)).

Summarizing, the referential nature of nonhuman primates has a peculiar asymmetry in comparison with human language. The poorly developed capacity to represent the contents of others' minds leads to a situation in which most of the communicative work is done by the receiver of a signal, based on inference, rather than intentionally by the producer. Well-developed theory of mind thus seems to be a crucial component of human cognition that is both necessary for fully-formed language, and absent in our nearest cousins. In the rest of this essay will argue that the connection between theory of mind and language is not one way, and propose two independent co-evolutionary links. First, the simple ability to produce a novel signal with the same structure as some other signaller's, which I will call "quoting" provides a potentially invaluable scaffolding for theory of mind. And second, the ability to represent minds within minds (high-order theory of mind) may have provided a crucial stepping stone towards, and selective pressure for, recursive syntax.

Quoting and Theory of Mind

Children hear their parents say many things they do not understand. By the age of three or four, they can reproduce these utterances effortlessly (and by rote, often with humorous effect). Put another way, children exhibit a sensorimotor ability (vocal imitation) that allows them to recreate the sensorimotor component of a meaningful utterance, while comprehending the meaning dimly if at all. However, the ability to retrieve the utterance (silently or out loud) along with its context, long after the communicative act is over, perhaps in concert with new information, provides a very important advantage in trying to make sense of others' minds. At the very least, the ability to encode part of a communicative event in the concrete realm of sensorimotor memory aids the encoding and retrieval of the event. More generally, the phonological and syntactic structure of the remembered utterance can act as a scaffold for a slow but steady increase in the understanding of the intended meaning. Anchored by the sensorimotor representation, the semantic and pragmatic interpretation process can continue days or even years after the utterance itself has faded. We all remember instances where we suddenly realized "ah, that's what she meant by that", months or years post-utterance. Thus, I suggest, the durable non-semantic sensorimotor representation (the "quote") can act as a scaffold upon which a rich interpretation of meaning can slowly build.

It is important to note that this advantage to listeners would accrue even in the absence of intentional signaling by the speaker. Someone muttering to himself, speaking to a third party, or even producing inchoate grunts associated with some other task, might produce signals that afford quoting. Nor must the call quoted be structured syntactically (into words and morphemes) for quoting to be advantageous. A "frozen" holphrastic utterance like "how do you do" can still be subjected to offline rumination and interpretation to discover its appropriate context ("aha - they always do that when two unacquainted people first meet!"). Given the repeated exposures and variegated data necessary to infer context, even simple holophrastic quoting could prove advantageous. Thus the coevolutionary relationship between quoting and theory of mind might have long predated language in its fully modern form, and provides a plausible, persistent selective force in the evolutionary path to language.

Second, and finally, I want to consider a related but considerably more sophisticated potential relationship between theory of mind and syntactic recursion. Basic phonological phrase structure entails combination of one type of entity into larger wholes (syllables into words, for example) without blending of their properties. This kind of structured compositional recombination seems quite ubiquitous in nature (Studdert-Kennedy,

1998). An example is the combination of small motor actions into larger complexes, whether chews within a feeding complex or prey capture within a locomotion complex. While the existence of compositional ability in one domain does not entail it in another (monkeys have it in their motor control, but not their vocal communication), we need not look far for possible precursor abilities of such basic phonological-level recombination, motor control being the most obvious possibility (Lieberman, 1984; MacNeilage, 1998). Another possibility is spatial cognition: we readily represent furniture within rooms within houses within towns, and so on (or substitute fruit, branches, trees and forests). In recursive syntax, in contrast, entities of the same type are embedded ("self-embedding"). This provides more of an evolutionary challenge. We string motor acts together sequentially, but we do not embed chews within chews, or syllables within syllables. It is not clear that motor behaviour could provide a preadaptation for recursion. Are there other cognitive capacities that might?

I join many others (Bergman et al., 2003; Byrne & Whiten, 1988; Dunbar, 1998) in suggesting that social intelligence has the requisite properties to both provide precursors for, and drive the evolution of, aspects of language.. The reasons are simple: in a complex social world there is always a premium on representing the minds of others. Furthermore, the potential rewards are great for doing so even slightly better than the competition. Such a Machiavellian "arms race" is the perfect evolutionary context to drive the rapid evolution of intelligence (Humphrey, 1976). In particular, few cognitive domains seem more suited to utilize full syntactic recursion than social cognition. The ability to represent minds within minds ("John knows that I know") is a huge step beyond simply knowing what John knows (or doesn't). But the ability to perform such embedding at progressively higher levels may be equally valuable strategically. The sentence "Mary was surprised that Jacques didn't know that Judith lied to Pierre" is both easy for humans to understand, and of great relevance in planning a coalition with Mary or Jacques. But it requires a level of self-embedding that is hard to envision being possible (or important) in other nonlinguistic cognitive domains. It should require no further argument to see that the quoting abilities discussed above are also relevant to this ability (both simply storing and processing them, and communicating them). Thus the hypothesis that high-order theory of mind is closely related to recursive syntactic self-embedding should be given serious consideration. More generally, detailed comparative consideration of the interactions between cognitive components of the language faculty may provide useful insights for hypotheses about language evolution.

References:

Arbib, M. A. 2003. The evolving mirror system: A neural basis for language readiness. In: *Language Evolution* (Ed. by Christensen, M. & Kirby, S.), pp. 182-200. Oxford: Oxford University Press.

Bergman, T. J., Beehner, J. C., Cheney, D. L. & Seyfarth, R. M. 2003. Hierarchical classification by rank and kinship in baboons. *Science*, **302,** 1234-1236.

Bickerton, D. 1998. Catastrophic evolution: the case for a single step from protolanguage to full human language. In: *Approaches to the Evolution of Language* (Ed. by Hurford, J. R., Studdert-Kennedy, M. & Knight, C.), pp. 341-358. New York: Cambridge University Press.

Byrne, R. W. & Whiten, A. 1988. *Machiavellian Intelligence: Social expertise and the evolution of intellect in monkeys, apes and humans*. Oxford: Clarendon Press.

Cheney, D. L. & Seyfarth, R. M. 1990a. Attending to behaviour versus attending to knowledge: Examining monkeys' attribution of mental states. *Animal Behaviour*, **40,** 742-753.

Cheney, D. L. & Seyfarth, R. M. 1990b. *How monkeys see the world: Inside the mind of another species.* Chicago, IL: Chicago University Press.

Christensen, M. & Kirby, S. 2003. Language Evolution. Oxford: Oxford University Press.

Corballis, M. C. 2003. From hand to mouth: The gestural origins of language. In: *Language Evolution* (Ed. by Christensen, M. & Kirby, S.), pp. 201-218. Oxford: Oxford University Press.

Donald, M. 1993. *Origins of the Modern Mind*. Cambridge, Massachusetts: Harvard University Press.

Dunbar, R. 1998. Theory of mind and the evolution of language. In: *Approaches to the Evolution of Language* (Ed. by Hurford, J. R., Studdert-Kennedy, M. & Knight, C.), pp. 92-110. New York: Cambridge University Press.

Fitch, W. T. & Hauser, M. D. 2004. Computational constraints on syntactic processing in a nonhuman primate. *Science*, **303,** 377-380.

Hare, B., Brown, M., Williamson, C. & Tomasello, M. 2002. The domestication of social cognition in dogs. *Science*, **298,** 1634-6.

Hare, B., Call, J., Agnetta, B. & Tomasello, M. 2000. Chimpanzees know what conspecifics do and do not see. *Anim Behav*, **59,** 771-785.

Hauser, M., Chomsky, N. & Fitch, W. T. 2002. The Language Faculty: What is it, who has it, and how did it evolve? *Science*, **298,** 1569-1579.

Hauser, M. D. 1996. *The evolution of communication*. Cambridge, MA: MIT Press.

Humphrey, N. K. 1976. The social function of intellect. In: *Growing points in ethology* (Ed. by Bateson, P. P. G. & Hinde, R. A.), pp. 303-317. Cambridge: Cambridge University Press.

Janik, V. M. & Slater, P. B. 1997. Vocal learning in mammals. *Advances in the study of behavior*, **26,** 59-99.
Lieberman, P. 1984. *The Biology and Evolution of Language*. Cambridge, MA: Harvard University Press.
MacNeilage, P. F. 1998. The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, **21,** 499-546.
Povinelli, D. J., Nelson, K. E. & Boysen, S. T. 1990. Inferences about guessing and knowing by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, **104,** 203-210.
Povinelli, D. J., Parks, K. A. & Novak, M. A. 1991. Do rhesus monkeys (*Macaca mulatta*) attribute knowledge and ignorance to others? *Journal of Comparative Psychology*, **105,** 318-325.
Seyfarth, R. M. & Cheney, D. in press. Constraints and preadaptations in the earliest stages of language evolution. *Linguistic Review*.
Seyfarth, R. M. & Cheney, D. L. 1980. The ontogeny of vervet monkey alarm-calling behavior: A preliminary report. *Zeitschrift fur Tierpsychologie*, **54,** 37-56.
Seyfarth, R. M. & Cheney, D. L. 1990. The assessment by vervet monkeys of their own and another species' alarm calls. *Animal Behaviour*, **40,** 754-764.
Seyfarth, R. M. & Cheney, D. L. 2003. Signalers and Receivers in Animal Communication. *Annual Review of Psychology*, **54,** 145-173.
Studdert-Kennedy, M. 1998. The particulate origins of language generativity: from syllable to gesture. In: *Approaches to the Evolution of Language* (Ed. by Hurford, J. R., Studdert-Kennedy, M. & Knight, C.), pp. 202-221. New York: Cambridge University Press.
Tomasello, M. & Call, J. 1997. *Primate Cognition*. Oxford: Oxford University Press.
Winter, P., Handley, P., Ploog, W. & Schott, D. 1973. Ontogeny of squirrel monkey calls under normal conditions and under aco.

Discussion

Precursory Cognitive Plasticity
Robert Stonjek
Apr 26, 2004 13:17 UT

'Theory of mind' may well be an indicator of some underlying cognitive ability that also results in a language ability as is known in humans, but it obviously isn't the causative agent per se. The question is "what underlying change in the brain leads to 'theory of mind' and a human-like language ability?"

On page 98 of 'The Singing Gorilla: Understanding Animal Intelligence' by George Page we read: "Like all animals, including us, pigeons and terns give evidence of 'behavioural prejudices', to use a phrase from 'The Animal Mind'. However, I was fascinated to learn that pigeons cannot be trained to peck in order to avoid a shock. They can be trained to hop on a treadle in order to avoid this shock, but they cannot be trained to hop on a treadle in order to obtain food. Do these results make sense? Yes, definitely. Pigeons eat by pecking, and they flee danger by jumping or flying away. Ask them to switch these two instinctive behaviours - to flee by pecking, so to speak, or obtain food by hopping - and they are in trouble. The same behavioural prejudices hold for lab rats, who can be taught to jump in order to avoid a shock, but cannot learn to jump in order to obtain food. Jumping for food is not their innate repertoire, and no amount of learning can compensate for that lack."

I am also reminded of the constrictor snakes that must 'discover' the prey that they have just killed. It seems that the hunting and eating part of their brains are unable to communicate with each other.

This indicates a lack of the essential plasticity required for language ie that words, like math, can be applied universally. That is, we don't have one language for feeding, one for defence, one for mating and so on. Language is very much an indicator of a gateway between various context specific behaviours.

Chimps do not show universal plasticity in this regard. They do not apply their language skills outside of the context in which they acquire them and do not seem to be able to transfer skills between types of tasks eg foraging for food to foraging for knowledge, a cognitive transposition that humans can perform seamlessly.

The physical indicator may well be bilateral asymmetry, though I stress that bilateral asymmetry is an indicator of the presence of this kind of neural plasticity and not the cause of the resulting cognitive plasticity. With this kind of plasticity the cerebral cortex becomes what I like to call "the scratch-pad of consciousness", that is, it can become what the cognitive system needs, that is, can respond to the environment during maturation to assign larger areas of the cerebral cortex to cognitive tasks stimulated during maturation, such

as language acquisition and the development of Broca's and Wernicke's area. Thus the brain can effectively mould itself to the (cognitive) environment.

For this to be possible, there needs to be a liberal sharing of incoming information across the cerebral cortex so that no confining of information or specialisation innately occurs, thus the cerebral cortex can 'mould itself' during maturation. A degree of informational spillage across modules (physical and cognitive) must continue to occur after the cerebral cortex is fully formed with all specialisations set in place. This must occur at around 7 years of age, the age beyond which a child will not fully recover from a hemispherectomy.

Apart from humans, only birds, especially the African Grey Parrots that Irene Pepperberg has worked with, seem to have a high degree of this kind of plasticity. Pepperberg's parrot 'Alex' can use words contextually including counting and classifying objects, making requests for food and toys etc (see 'The Alex Studies:Cognitive and Communicative Abilities of Grey Parrots' by Irene Maxine Pepperberg).

The informational traffic between cognitive modules may well make up the bulk of the experience of consciousness as known to humans. Though consciousness exists in non-human animals, the form known to them may be analogous to the many persons of a siphonophores, that is, in fight or flight your cat doesn't know you because the conscious platform has switched from the one that recognises you as a friend and master.

If information can be transferred between cognitive modules (eg from vision to language to another individual to plan of action to locomotion) then it can be transferred to and from the *same* cognitive module. This is, in essence, cognitive recursion (eg language to vision to language truncating to a language-language loop).

Thus I suggest that plasticity of the cerebral cortex indicated by bilateral asymmetry forms the physical base, cognitive module information sharing results from that plasticity, and 'theory of mind', recursive abilities and language follow on from that.

Kind regards, Robert Karl Stonjek

Recursion in mindreading and language
Dan Sperber
Apr 27, 2004 11:32 UT

Thank you for an excellent contribution to this conference. Here is a question about a specific point: recursion.

The thought that John knows that I doubt that Mary is coming can be formulated as in (1) with two levels of embedding, or as in (2) with two levels of demonstation:

(1) [John knows [that I doubt [that Mary is coming]]]

(2) John knows this: I doubt this: Mary is coming

In many languages (e.g. Dorze, an Omotic languages of Ethiopia I studied a long time ago), demonstrative formulation of attributions of thoughts or utterances is the only formulation available. Also, it has been argued by Davidson (the "paratactic theory") that something like (2) provides the proper semantic analysis of sentences like (1). Other forms of embedding in language, restricitive relative clauses for instance, don't lend themselves to such an alternative formulation or analysis.

Hence my question: how are we exactly to take the idea that social intelligence provides a precursor for recursion in language? Is, before the emergence of language, the format of attributions of thoughts about thoughts on the model of (1) or of (2), and does it matter? Is this format, whatever it is, recycled ('exapted') and modified in the evolutionary emergence of a language faculty?

Mechanisms of Recursion
Petr Kotatko
May 6, 2004 16:49 UT

Dan Sperber's question about recursion is a very deep one, and I can't really do it justice in this forum. But here's a few ideas.

The possibility of a paratactic analysis for belief embedding is a specific case, I think, of a more general problem confronting any discussion of recursion. Take the Chomsky hierarchy and finite-state vs. phrase-structure grammars as an example. Chomsky 1957 pointed out that a simple recursive phrase structure grammar can generate "sentences" of the form a^n b^n, while no finite-state grammar can do so. However, this is only literally the case if n can go to infinity; a finite state grammar can easily be constructed to generate such sentences for any finite n. While it gets more and more complex and unwieldy as n increases, it is possible. So, as is generally the case in mathematics, infinity becomse a very useful tool for abstraction and generalization: the differences between FSGs and PSGs can be made clear and explicit only in this limiting case.

This is an example of a general problem for discussing recursion because there is always a non-recursive way to implement a noninfinite recursive system. And any real system (like the brain, or a computer for that matter) is noninfinite. So the types of mathematical certainties that are so satisfyingly possible in computational linguistics are simply not available for the emprical scientist seeking to answer such questions: the troops on the ground will have to be satisfied with making the recursive option seem ever more plausible, and the alternative ever less so.

So, let us ask what the formal notion of "recursion" might mean when applied to the brain. In computer science, a recursive procedure is one that calls itself. The sequence of calls are stored on a "stack", of limited memory, and this puts a finite limit on the depth of recursion. The interesting thing is that each new function call (each level of embedding in linguistic terms) costs resources, so that the complexity of the whole data structure grows more complex with increasing recursion. This is not the case with a simple iterative system (xt+1 = xt+1), because here all you need to store is a single number x, even as the system counts to infinity (well, there will be overflow eventually, but this is a different problem).

Clearly we don't expect neurons to be embedded within other neurons in the brain. So that's certainly not where to look.

Is a recurrent feedback system (of the sort that is ubiquitous in mammalian neocortex) recursion? No, because a neuron back-wired onto itself can keep firing itself till the cows come home without ever increasing the complexity of the stored information.

So what is needed is a system of many neurons that can implement the two key parts of the system, the embedder and the hierarchical structure (the function and the stack, in CS terms). And the load on the latter part will increase as embedding increases to some finite limit (only three for center emedding, lots more for right-branching but still obviously finite). I think its fair to say that no one has the slightest idea how this is implemented in the brain.

So, to finally address Dan's question, I think it DOES matter whether the embedded or paratactic analysis is the right one, if we are asking questions about how the recursive system is implemented in the brain. I think finding out will require a back and forth between linguists (who after all framed the problem in the first place) and neuroscientists (who will someday, hopefully, give us the answer).

BUT, as far as the functional selective value is concerned I think these are notational variants that don't make much difference. Understanding the sentence requires us to keep in mind the same number of propositions ("on the stack") in either case, and it is the abilitie to do this that was of value to our ancestors.

Finally, a critical question, if we accept the mind-reading as preadaptation idea, is whether the exapted linguistic use has added its owns new tweaks to the system (e.g. because of the additional "hooks" provided by the speech output system. It seems plausible that it did, but by no means is this a logical necessity.

On Quoting and Theory of Mind

Gloria Origgi
May 4, 2004 21:22 UT

Tecumseh Fitch puts forward the interesting hypothesis that the vocal imitative ability to repeat utterances even with a very poor understanding of them is an ingredient to develop an inferential capacity to grasp speaker's meaning.

Children entertain in their minds some of their parents' utterances without understanding them, and the very fact of making them available in their minds is, according to him, a starting point to make inferences about their meaning.

Fitch grounds this ability in the linguistic resource of "quoting". But is it quoting a specific linguistic mechanism? Doesn't it require already a "metarepresentational" ability to represent in our own mind a representation?

W.W. Quine used to define quoting as an "essentially dramatic act". I can quote any aspect of an utterance, not only its phonological form, but its prosody or intonation. Also, very often quotation is not "literal", but a mix of direct quotation and indirect reconstruction of what has been said.

All these uses of quotation seem to suggest a capacity of mentioning language rather than using it that implies already some metarepresentational abilities or even some abilities to mentalize. Hence my question: Is quotation a linguistic device or is it an example of a metarepresentational use of language that requires rich inferential capacities?

References:

WW. Quine (1960) Word and Object, Harvard.

D. Wilson (2000) "Metarepresentation in Linguistic Communication", in D. Sperber (ed) Metarepresentations, Oxford, pp. 411-448.

Bare-Bones Quoting

Petr Kotatko
May 6, 2004 16:55 UT

Gloria Origgi's question is a good one, and certainly if Quine's very high-level definition of "quoting" were intended, she would be right about it entailing metarepresentational skills.

I was using the term in a much more stripped-down way: to mean parroting or mimicking without understanding or even full structure (with apologies to Alex the parrot); so the "indirect reconstruction" part (which we of course do) could be lacking in my bare-bones quoting. That's part of the point, since if it weren't based on something simpler it couldn't help explain metarepresentation!

So I think there is a reading of quoting that entails ONLY the ability to recreate a motor sequence, and that's what I meant throughout.

Parrot's Quoting

Gloria Origgi
May 7, 2004 12:59 UT

Yes, but if quoting must do the job that you are saying, that is, facilitating inferences about speaker's meaning, then it must be something more than the mere ability ro recreate a motor sequence.

Parrots are able to recreate words and sentences but apparently this doesn't have any consequence on their theory of mind!

Recursion is parallel, but which comes first?

Jill de Villiers

May 5, 2004 18:16 UT

Fitch's paper is a wonderful contribution to this taxing exercise of linking language, evolution and theory of mind. Fitch joins "many others in suggesting that social intelligence has the requisite properties to drive the evolution of aspects of language." I completely agree that there is an advantage to representing the minds of others in complex social worlds.

Fitch points to the striking parallel in the recursion of beliefs and of sentences, and suggests it is the recursion of beliefs that drives evolution of the recursion of sentences. But in my own paper (next on the list!) I ask the question the other way round: were the representations necessary for embedding beliefs enabled by developing recursive language first? But then the obvious question is, linguistic recursion for what, if not for talking about beliefs?

I share with Fitch the intuition that quotation has everything to do with it, but in my argument imitation isn't the key, it is recursive embedding of statements of what someone said. I speculate about the advantage that would have, irrespective of representing beliefs.

I am delighted to learn from Dan Sperber's comment about the interesting case of the language Dorze with "this" and embedded quotes. As long as there is variation in how the formal markers of recursion are expressed across different languages, the child has something to learn even if the capacity for linguistic recursion is innate.

Chickens and Eggs
Petr Kotatko
May 6, 2004 16:04 UT

Jill de Villiers excellent question is impossible to answer definitively, like so many others in phylogenetic reconstruction. In any complex system like language, there will be interactions over evolutionary time that tax our everyday intuitions about causality. For example, the first lungs evolved in things that were clearly still fish (they spent all their time in the water), probably to help them get oxygen when the water around them was anoxic. This enabled them (our crossopterygian ancestors), to make the first tentative lunges out of the water, and began the evolutionary process leading to tetrapods. Now, the more time spent out of the water, the better the lungs and ventilation system had to work, beginning a circular feedback process that continues even today. Once the ball starts rolling, questions about which came first are truly like asking about the primacy of the chicken or the egg.

In the case of lungs we have fossils to answer the question (lungs came first, before land living) but with language (particularly with abstractions like recursion) we will never have fossils to go on. So what basis have we to answer de Villiers' question? As she points out, once linguistic recursive embedding is established it certainly is useful in mind-reading and Machiavellian scheming. But since we it looks like some mind-reading is already going on in chimps, we can at least infer some such abilities in our LCA w/ chimps, and it is clear that more such abilities would be quite useful even for today's chimp lifestyle. I thus think that the mind-reading first, linguistic recursion second hypothesis is better supported by the comparative data. However, once linguistic recursion was established, it set off the feedback loop, and the upward spiral of complexity. Again, this selection force is probably still active and powerful today.

But with all this said, we still can never really know. Alas.

Recursion, Quotation, and Representation
Tom Roeper
May 8, 2004 21:13 UT

The dicussion of recursion in human and animal, language and thought is interesting. It seems to me that it calls for refinement in order to make real progress. In particular we need to distinguish recursion from iteration and to distinguish implicit notions of other minds from actual cogitation.

William Snyder and I, in a series of papers, have traced the phenomenon of recursion in language acquisition. Languages differ in the kinds and domains where recursion can occur. English allows recursive compounds, possessives, and adjectives, but disallows direct recursive verbs. Crucial is not two, but three instances, to establish recursion for children. That is, English has student film committee" or continuing:

“student film committee program office”. French allows two nouns, but not more (homme grenouille). English allows multiple possessives (John’s friend’s car’s motor) while German allows only one (Marias Haus). English has multiple adjectives, but French only one. Swahili has multiple serial verbs (buy read love a book) but English has only one (come play with me, but not *come play roll down the hill).

Children with crucial (but very infrequent) input realize whether their grammar has recursion in particular areas. They must be programmed to carry out this form of recognition. That is, one has: “big, red, strange sweater” in English, but only one pronominal adjective in French. Hearing one or two adjectives will not be enough.

One dimension of variation is whether the grammar favors recursion to the left or right. Another is whether center-embedded recursion can occur. Iteration (very, very, very big) seems to be something else. The semantics have to be recursive too: John’s friend’s car is not John’s car (so it is not John and friend’s car). In order to distinguish paratactic connections from genuine recursion, this kind of analysis should be undertaken.

In an informal experiment, Sarah Gentile introduced John and his sister. Then she showed them three pictures: John, John and his sister, and just his sister. Then she asked children 3-5yrs to point to: John’s sister’s picture. Only 4yr olds reliably knew that only the picture of the sister (not John and his sister) was correct. This is the kind of test one would need to do with chimps to show real recursion it seems to me.

Quotation is not really recursive I believe. I have not found anyone who can easily understand: “”Nothing” said Bill” said John. With enough from context one can understand, but it is nothing like the sixfold embeddings we find with verbs of thinking. It is nothing like the kind of saying-embeding (not quotation) a 5yr-old might say: “Mom said you said I did it, but I didn’t “ which involves four embedded sentences plus ellipsis in ten words.

Some kind of recursive cognitive structure is arguable involved in thoughts about other people’s minds. But what is the line between mind and body? When a boxer feints, and his opponent moves to counter a punch that does not happen, but then gets socked in the jaw, are we dealing with a physically embedded notion of deliberately provoked false belief? If so, can animals do the same? It seems quite different from conscious thought about someone else’s thought? When animals exhibit fears of humans, what is the mental state that is involved.

Why does one cat sat in my lap and another dart away? Is it afraid that I will hit her? There is “implicit knowledge” here not completely unlike the unconscious knowledge used in language. It is not what we mean by “mental state”, but this sphere of knowledge is difficult to isolate.

As Churchland puts it : “We make each one of them [representational spaces] plastic and multipotent in its semantic content and its conceptual organization. And we reach out to include motor cognition and practical skills, along with perceptual apprehensions and theoretical judgment, as equal partners in our account of human knowledge”

Words
Anne Reboul
May 9, 2004 8:33 UT

The present paper is extremely interesting as is the discussion that it triggered. I just want to come back to a passing remark in Fitch's paper. He (rightly) says that vervet monkeys' alarm calls are not homologous to words. In the part of his paper that he devotes to "Semantics and theory of mind", he comes back to the issue and gives a detailed and highly interesting discussion on how alarm calls fall short of words, suggesting that in monkeys, the symetry between signaler and recipient which "is central to both word meaning and Gricean maxims" is absent. Though "monkeys have a flexible and well-developed interpretive ability" as interpreters — Fitch reminds us that they can learn novel calls even when they're produced by other species —, they do not seem to be able to intentionally signal and are not sensitive to putative recipients' knowledge or ignorance.

There are a few points to be made here: first of all, the Gricean view of meaning (which is at least partly independent from Gricean maxims) distinguishes between natural meaning and non-natural meaning. Non-natural meaning, which is the kind of meaning found in language and in words, is dependent on two intentions, a first-level intention to communicate such-and-such and a second-level intention that the first

intention succeeds in virtue of the recipient's recognition of the first-level one. This would seem to entail a fair-developed mindreading ability. Though the Gricean view of semantics has been criticized, I think that the distinction between natural and non-natural meaning stands and that it may enlighten the discussion. For instance, the description of alarm calls suggest that such calls ressort to the domain of natural meaning rather than non-natural meaning. In the same way, monkeys' ability to learn novel calls even when they are produced by other species may be seen as an ability to learn natural meaning, given that natural meaning can generally be learnt through a simple association between a non-intentional clue (here the calls) and what it non-intentionally signals (here the presence of a predator). There is no doubt that animals have to be able to learn to read natural signs of all kinds (from footprints to odors) if they are to survive. So monkeys' ability to do so is quite expected. The question is whether this indicates an asymetry, as Fitch suggests. If alarm calls are natural signs, then one should not expect them to be intentional and there is no reason to think that they would be interpreted as non-natural signs. This goes for same species calls but it does also go for other species calls. In this sense, monkeys as interpreters would not be significantly different from monkeys as producers. In neither case is any intentional data involved. So maybe monkeys, and more generally non-human primates, whether or not they do have some kind of ToM, just do not need it in their communicative systems. But on this account, there is no special reason to think that communication is asymetrical in monkeys, in that interpreters would be cooperative while signalers would not. Both are non-cooperative, symetrically so.

Why language first?

Jill De Villiers (Smith College)

(Date of publication: 10 May 2004)

Abstract: For some philosophers, thinking is necessarily propositional and symbolic even in animals and infants, and logically prior to natural language but equally complex in its capacity for representing meanings. If so, a nonverbal creature could make inferences over such representations for predicting how others act. Alternatively, not only are propositional attitudes the appropriate way to describe our internal representations, but these are scaffolded in human development by having language of the appropriate degree of complexity. On this view, an individual with less language would not be able to formulate the appropriate representation of another person holding a false belief, and hence have no basis for reasoning about their actions. But how could that individual gain the necessary linguistic structure without acquiring first the conceptual underpinnings of false beliefs? The possible solution I provide for ontogenesis might be plausible for the species too.

Meme or Module?

Two strikingly different opinions exist about the origins and ontogenesis of theory of Mind. One position posits it as a culturally transmitted meme, a theory of naïve psychology that has proved sufficiently useful so as to occur to most if not all the human groups. The second posits a module genetically specialized for the understanding of mind-reading, sprung forth anew in every infant by virtue of being human (or maybe just a social primate). The meme view explains the four-year delay for *false belief reasoning*, though perhaps not the surprising invariance in the timetable. The module view explains the uniformity, but only at the cost of positing a moderating influence of "processing capacity" or its kin, executive control, which must mature.

Not surprisingly, two parallel opinions also exist about the origins and ontogenesis of human language. One position posits a cultural transmission, the other a genetic module. On the first view language learning will take time and experience, on the second it won't, except for the vexing issues of the time needed to learn a arbitrary lexicon, and to sort out from primary data which set of parameters is appropriate.

I venture that a popular opinion is a mixed model: a module for language, and a meme for theory of mind. That is, genetic transmission for the first, cultural transmission for the second, using language as the medium. On the basis of poverty of the stimulus for language learning, the abstractness of what is learned, and the lack of direct instruction, linguists have concluded that language learning is too complex for it to be a meme. Gold's mathematical proof of the unlearnability of a human grammar within a finite time (without UG to guide it) has held firm after 35 years of debate (8). But the idea of cultural transmission of theories about the mind not only seem plausible, but is apparently supported by data showing the impact of different rich or impoverished linguistic circumstances for learning the theory.

The literature on Theory of Mind in animals and humans leads me to the conclusion that whereas animals such as chimpanzees may share many of the mind-reading skills of humans, humans uniquely can do *false belief reasoning*. The literature on natural communication systems and on artificial language training with animals and humans convinces me that whereas several properties once held unique, such as referentiality and combinatorial ability, might be shared, humans uniquely have *recursive* syntax (10).

Might there be a plausible evolutionary story connecting the two findings? A cultural meme account of the transmission of theory of mind entails a complex enough language in which to couch talk about false beliefs, and that might in turn entail recursive syntax. This is a sensible account, but let us not stop at sensible accounts. It so happens that I have a theory about this that makes a connection between syntax and false belief reasoning in ontogenesis, that is not due to either
a) both being products of cultural transmission or
b) mature Theory of Mind arising by cultural transmission via language.
I use this essay to ask: Can that theory be extended to the case of language evolution?

Language for thinking?

Several philosophers of mind have argued that we take seriously the notion of a language of thought, a language rich enough to express the contents of our propositional attitudes and therefore as rich as natural languages. We can make inferences across the propositions of this language of thought and deduce

consequences e.g.:

If George thinks his keys are in the hall, George will search for them in the hall.

Sometimes we engage in this reasoning overtly, often it does not reach 'consciousness', and may be in some form prior to full linguistic formulation. The normal condition under which we test 3 and 4 year olds' understanding of another's false beliefs are simple: we set up a condition under which someone would act one way if she knew a fact, and another way if she believed something contrary to that fact. We ask: can the child represent the contents of that other person's mind? So on the classic task can the child retain:

Maxi believes the chocolate is in the basket

even when it has now been moved from the basket to the cupboard, unseen by Maxi. If the child can retain that, he can deduce:

Maxi will look in the basket

Or explain:

Maxi looked in the basket because he left his chocolate there

The question is, how is that belief about Maxi's belief represented? I have argued that the representation must be as rich as the statement in our natural language, and have the crucial property of a recursive sentence, namely one sentence embedded in another, in which the embedded sentence is false while the overall sentence remains true. This type of structure is uniquely associated with verbs of mental state and communication and this fact provides an important key to its learnability on my account. I have argued that the child can recognize the appropriate recursive structure using communication verbs:

Maxi said the chocolate was in the basket

for which the child can experience overt events, speech acts, and recognize how the truth conditions hold against the speech. Only when a child has this linguistic structure established, will he have the representational capacity to reason about false beliefs, by analogy with attributed false statements. That is, the verbs of communication and mental state share a crucial *recursive* syntactic/semantic structure, and identifying it in the case of overt speech provides a bootstrap to the understanding and generation of statements about false beliefs. The predictions have been borne out in several ways empirically, though there is still much theoretical work to be pursued. In brief, normally-developing children (both English-speaking and native ASL) show evidence of mastery of complement syntax via understanding wh-question-movement:

The girl bought apples, but she said she bought pears.
What did she say she bought? (pears, not apples)

before they pass the false belief tasks, even non-verbal false belief tasks (3,4,5). So do deaf children (both non-native born ASL and orally-trained) who are language delayed (6), and children with SLI (7), and high-functioning children with autism (20), though this often means on a timetable several years beyond normal. Training studies to "teach" children complements with communication verbs show transfer to false belief reasoning, just as if the language mastery had enabled their representational skills(9,12).

Is it trivial?

There are several ways that the relationship between language skills and theory of mind could be relatively trivial, and critics have been quick to point them out. First and foremost, it is not just that passing the verbal FB task itself requires certain linguistic skills, e.g. the ability to follow a narrative or answer a question. Instead, several different nonverbal (or at least low verbal) FB tests show the same dependency on the syntax of complementation. Second, it is not that that the language task and the FB task share in common executive function or meta-representational skills that provide the resource underpinning: language-delayed deaf children succeed on those at a younger age. Thirdly, it is not that the language task entails false belief understanding, because it precedes success on the FB task, and because the task does not require the child to make an attribution of belief. Furthermore, the older deaf children should pass the nonverbal FB task before the language in the case that FB is conceptually prerequisite, and they do not (6). (For a full review of these and answers to claimed counter-evidence, see (2)).

Why children take four years

It is important to recognize the peculiarity in this story, that children take almost four years to 'recognize' the right structure as a recursive complement. Why, if they have recursion as part of the evolutionary bio-program for language? I argue that it is critical that they avoid false analogues such as adjunct clauses (which don't have the right truth properties):

Maxi spoke and the chocolate was in the basket

and irrealis clauses (which don't have truth conditions of the same sort)

Maxi wanted the chocolate to be in the basket

To get this far, the child has to make a multitude of subtle lexical and morphological distinctions that will take time and experience to learn. Recursion may be there from the start but *recognizing* it cannot be instantaneous (see also (17)).

Evolution and language

Hauser, Chomsky and Fitch (2002) identify recursion as the fundamental property of human languages that distinguishes them from other animal communication systems. The authors conclude that recursion may be present in other systems not related to language, such as navigation through space, but not in the communication systems of animals.
If true, the natural question arises, why not? Is all such speculation tainted with the shadow of Dr. Pangloss? Bickerton (1) argues for instantaneous change or punctuated equilibrium deriving syntax *in toto* from a previous stage of "proto-language"- essentially just a lexicon and non-syntactic strings. Although some thinkers admit the possibility of gradual change, Bickerton and others scorn the notion of e.g. generative syntax minus Binding rules, or without Subjacency, because they are adamant that these pecularities of human syntax are not *functional* and therefore could never have been favorable adaptations. Furthermore, Bickerton claims these "partly evolved" languages should show up still in remote cultures, and that they don't. However, new data on early forms of Nicaraguan Creole Sign hints at the possibility of "intermediate stage" languages (16). While highly sympathetic to Bickerton's objections, I am tempted into speculation, not just what use is *syntax*, but more particularly, what use is *recursion*? Note, I am NOT suggesting that recursion was "invented" for a particular semantic or pragmatic need. I can still believe that a sudden mutation allowed for a formal property that did not exist in language before, which then conferred advantages on its speakers when it was used for a certain purpose. Bickerton's own argument takes the same form.

The mathematician Nowak and colleagues (13,14,15) have begun asking questions about the adaptive value and learnability of simple protolanguages such as a primitive lexicon linked to referents. Not surprisingly, they find that the demands of memory, discriminability and time to learn constrain such a system to a fairly small finite set of signals, just what is seen in animal communication repertoires such as monkey calls. Their analysis of the evolution of grammar considers the conditions under which a combinatorial system would be more advantageous than a finite lexicon. In their modeling, if the number of distinct pairings of objects with actions is above a certain threshold, then there is an advantage to a combinatorial syntax that allows creativity for novel events over a large lexicon. They specifically compare two kinds of transmission: cultural, versus genetic, so this work is getting closer to the kinds of questions linguists have asked. That is, are we talking about the evolution of languages here, or the evolution of speakers of those languages? Can we talk about the adaptive value of the first without the second? One must assume that for something to be coded in the genes, it has to confer reproductive advantage on the carrier of those genes, yet languages change (on the surface) at a much more rapid rate than populations. It is important to consider both possibilities of transmission.

Adaptive value of recursion

A simple combinatorial syntax is a long way from the kind of recursive syntax that linguists find uniquely human. Nowak and colleagues have so far steered clear of claims about particular linguistic properties, so Dr. Pangloss hoves into view. The only way to proceed it to consider carefully what recursion might gain us, and then ask if this is enough to give languages with it an advantage over languages that do not have it, and their speakers too! Hauser et al do not specify which variety of recursion they regard as critical, and there may be several forms (17).

First, recursion often provides economy, that is, it avoids redundancy across multiple independent sentences. I do not find his kind of recursion, such as in relative clauses, particularly convincing in terms of evolutionary adaptation:

The boy was stuck in a tree. John helped the boy.

→John helped the boy who was stuck in a tree.
The boy who was stuck in a tree? John helped him.

Not only are there alternative non-embedded structures that reduce redundancy, but human languages also show variation in which types of relative clauses they permit (3). For example, someone might argue that it is economical to say,

"I met the man that my brother is taller than"

but you can't even say that in French! Since forms of recursion vary across languages, this variation is probably a function of cultural transmission. Consider cases also like:

"Fred's mother's dog's bowl"

or:

"The oak tree by the pond down the path beside the cliff…"

These forms specify a referent adroitly, though in real oral communication the information is likely to be less densely packed:

"You know that cliff? Remember that path by the cliff? If you follow it you come to this pond, ….."

Although there is an economy of sorts in dense recursion, it is often at the expense of parseability for the listener, so one has to wonder what the evolution modelers would enter as constraints and whether languages with this feature would be selected for. These seem at most slight economies, and not likely to confer survival advantage on the speakers.

Is there another kind of recursion that is more significant from the point of view of speaker advantage? Bickerton argues that syntax permitted off-line thinking, and retention of representations of events not in the here-and-now, but in the past, or the future, or hypothetical. In this he makes the same point as Newmeyer (this conference) that language has essential cognitive advantages. However, once language is not longer about the immediate situation, certain new problems arise in communication.

It is not too far-fetched to argue that social creatures other than humans monitor the *reliability* of signalers. Vervet monkeys may be encoding referents (eagle, leopard, snake) in their cries, but monkey mothers can ignore the "mistakes" of their babies who make the cry for "eagle" (or whatever it means) at the sight of a falling leaf (18), and others claim that social primates might monitor the excitability and status of their group-mates in deciding whether to react to e,g. alarm calls(19). Animals may make these qualitative judgments, though it is unlikely they refer to notions such as truth or falsity of the call: reliability is all one needs. Young children are also attentive to the reliability of language-users and prefer to learn from the individual who previously showed themselves to be using the code in the same way as the child and her family (11). Nowak et al. take it as a fundamental premise that reliable signalers gain reproductive advantage.

It seems to me that syntactic recursion is most useful to mark that an event was experienced indirectly:

Mary says that a deer is by the tree.

The form makes use of recursion to provide the content of Mary's speech and mark it as her Point of View, allowing the speaker to evade responsibility for a false report. The form signals that the event was only attested through the speech of another. The recursive form above is not the only device, but it is the most common linguistically. Can it be done other ways? Here are several:

Type 1:

"The deer is by the tree."

"You saw the deer?"

"Mary saw the deer."

This relies on questioning about the source of evidence, and would add complexity to conversations if it had to be done continually.

Type 2:

Enact as Mary (e.g. taking on her voice, her stance): *"The deer is by the tree"*

This successfully marks the speech as uttering from another, but relies on mimicry and acting ability. Signed languages such as ASL adopt roles for the speaker, by adopting a different stance and speaking *as* the speaker, e.g. Mary. This is grammaticalized in ASL, but to my knowledge no spoken human language has been shown to rely solely on gesture or paralinguistic features like vocal quality to convey speaker identity.

Type 3:

"The deer is (allegedly) by the tree"

By "allegedly" I am paraphrasing an evidential morpheme, which English does not have. Some languages have evidentials, indicating the source of knowledge, such as hearsay, or directly witnessed, or inferred. Again, it is unclear that any language relies solely on such a morpheme, rather, this is in addition to the usual mechanism of recursive complementation under a verb of communication. One can see the advantage of the latter: if speaker reliability is key, how does it help to indicate simply 'hearsay'?

If my speculation is correct, a good case for the real utility of linguistic recursion may lie in describing *testimony*. If so, then languages with this property would confer an advantage to their speakers over languages that did not. The other advantages of recursion (e.g. economy and efficiency) come for free. Once the machinery is established for distinguishing point-of-view on testimony, it is but a short step to using that same machinery for representing beliefs. Without it, no such representation is possible, however much talk one might hear about minds.

Why invent 'think that"?

If Mary utters the sentence and there is no deer by the tree, what can be concluded? One could just write her off as unreliable, but there are in fact three possibilities that refine that conclusion:

A. If Mary is a child, perhaps one could doubt she can use the words *deer, tree, is* properly, rather like the mother Vervet ignores her young.

B If Mary is an adult, then there is a quandary: why is a speaker mis-using an already-learned code?

1. The first alternative is that Mary had something to gain by saying it. We use language not just to describe, but to persuade, deceive, cajole, seduce. Perhaps Mary has learned that when she says this the group moves away from the fire and vacates a nice warm spot for her. If Mary is using the code to deceive, this represents a violation of the rules of communication to which the group should be highly attuned and resist. (And teach their children about in cultural memes like 'The Boy Who Cried Wolf!')

2. In the second alternative, Mary "made a mistake". This is a sophisticated notion that relies on the notion of intention, but is probably within reach of our closest primate relatives and very young children. We don't, however, always invoke the notion of *false beliefs* to account for mistakes. If I trip over a crack in the sidewalk, it would be odd to say "Jill thought the sidewalk was smooth". We recognize that our perceptual and motor systems have certain error limits, and when we see someone else fall or fail to achieve a goal we recognize the goal and the persistence towards it, and the behavior as an error. But what about when Mary says she saw a deer where there was none? An alternative that we might appreciate *only if we have had like experiences* is that she was
a) not failing to use the code properly
b) not deceiving us but
c) perhaps she saw a shadow *as* a deer.
Seeing "as" is an expression that reflects the fact that there is interpretation in the visual system. This is where Simulation theory might have a point: I catch myself in the act of shouting "deer" when I re-see it as a shadow, and find that experience distinct from failing to know the right referent for 'deer'. Having recognized that fine distinction, I project it onto Mary. But simply saying "Mary made a

mistake" or Mary was wrong" glosses over these fine conceptual distinctions. We need to say, "Mary *thought* she saw a deer". At that point the representation is sufficient to allow computation, prediction and explanation.

It seems quite possible that this last step in the chain requires cultural transmission, i.e. the forms, once invented, were sufficiently useful to be retained across generations. For example, it is possible that not every language has lexicalized the difference between 'say" and "think". But the tale I have spun here does NOT entail cultural transmission of testimony about mind, i.e. of the theory of mind itself. Instead, the theory – that people have different points of view- is "encapsulated" in the linguistic forms themselves, so a child will encounter them in a wide range of environments without requiring special instruction or situational encounters. Any language is the embodiment of inventions of hundreds of generations of ancestors, only some of whom were bright or lucky enough to have the right experiences and persuade others to follow their rhetorical style. Every new child who learns the language benefits anew from that collective wisdom. It is this kind of reasoning that persuaded Ken Hale to write about the death of languages as being as devastating as the death of species: we lose the collective inventiveness of a whole line of human beings when the last speaker dies.

The model I have proposed acknowledges:
a) The genetic basis of language structures.
b) That despite a), there are still language acquisition "problems" to solve.
c) That a mature theory of mind rests on the representational structures that language provides, so language is a fundamental aid to thought.
d) That it is through communication that the value of recursion to *speakers* might have first arisen.

References

Bickerton, D. (1995) Language and Human behavior. Seattle, WA: University of Washington press.
de Villiers, J.G. (in press). Can language acquisition give children a point of view? In J. Astington & J. Baird (Eds.) *Why Language Matters for Theory of Mind*. Oxford Press.
de Villiers, J. G. & de Villiers, P. A. (2000). Linguistic determinism and the understanding of false beliefs. In P. Mitchell & K. Riggs (Eds.), *Children's reasoning and the mind*. Hove, UK: Psychology Press. pp.189-226.
de Villiers,J.G. & de Villiers, P.A. (2003). Language for thought: Coming to understand false beliefs. In D. Gentner & S. Goldin-Meadow, (eds) *Language in Mind: Advances in the Study of Language and Cognition.* MIT Press.
de Villiers, J. G., & Pyers, J. E. (2002). Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cognitive Development, 17,* 1037–1060
de Villiers, P. A. (in press) The role of language in theory of mind development: What deaf children tell us. In J. Astington & J. Baird (Eds.) *Why Language Matters for Theory of Mind*. Oxford Press.
de Villiers, P. A., Burns, F. & Pearson, B. (2003) The role of language in the Theory of Mind development of language-impaired children: Complementing theories. In B. Beachley, A. Brown, & F. Conlin (Eds.), *Proceedings of the 27th Annual Boston University Conference on Language Development*. Pp. 232-242. Cascadilla Press, 2003.
Gold, E.M. (1967) Language identification in the limit. Informat.Control, 10, 447-474.
Hale, C.M. & Tager-Flusberg, H. (2003). The influence of language on theory of mind: A training study. *Developmental Science, 6*, 346-359.
Hauser, M, Chomsky, N. & Fitch, W.T. (2002) The faculty of language: what is it, who has it and how did it evolve? Science, 298, 1569-1579.
Koenig, M.A., Clément, F. & Harris,P. (in press). Trust in Testimony: Children's use of true and false statements. Psychological Science.
Lohmann, H. & Tomasello, M. (2003). The role of language in the development of false belief understanding: A training study. *Child Development*, *74*, 1130-1144.
Nowak, M., Komarova, N & Niyogi, P. (2001) Evolution of universal grammar. Science, 291, 114-118.
Nowak, M., Komarova, N & Niyogi, P. (2002) Computational and evolutionary aspects of language. Nature, 417, 611-617.
Plotkin, J & Novak, M. (2001) Major transitions in language evolution. Entropy, 3, 227-246.
Pyers, J. (2001) Three stages in the understanding of false belief in Nicaraguan signers: The interaction of social experience, language emergence, and conceptual development. R. Senghas (Chair), *The Emergence of Nicaraguan Sign Language: Questions of development, acquisition, & evolution.* Invited symposium conducted at the 31st Annual Meeting of the Piaget Society, Berkeley, CA.
Roeper, T. & Synder, W. (in press) Language Learnability and the Forms of Recursion. In Di Sciullo, A. M. and R. Delmonte (Eds.) UG and External Systems. Amsterdam: John Benjamins.
Seyfarth, R.M. & Cheney, D.L. (1986) Vocal development in vervet monkeys. Animal Behavior, 34:1640-58.
Smith, W, (1977) The Behavior of Communicating. Cambrudge, Ma: Harvard University press.

Tager-Flusberg, H. & Joseph. R. (in press). How language facilitates the acquisition of false belief in children with autism. In J. Astington & J. Baird (Eds.) *Why Language Matters for Theory of Mind*. Oxford

Discussion

benefits and costs
Dan Sperber
May 10, 2004 13:34 UT

When asking what is the adaptive value of recursive syntax, as does Jill de Villiers, it would be interesting to ask not just what is the benefit of recursion but also what is its cost. More precisely, one should ask what are the costs and benefits of having the ability to acquire a language with recursive syntax. Some abilities or organs, such as visual perception and the eye, are obviously costly: both their ontogenetic development and their maintenance use energy, and they may involve extra vulnerabilities (e.g. to specific diseases) for the organisms endowed with them. In general, the benefits of an evolved adaptation must be large enough to offset its costs. Now, the brain is an organ with high consumption of energy, so the question arises: does the acquisition and use of recursive syntax increase or decrease consumption of energy in the brain? In particular do individuals who make a more intensive use of recursion in their use of language make, *ceteris paribus*, a more parsimonious usage of their brain? If so then this might be the function, or at least *a* function of recursion. Jill de Villiers however argues against this. If recursion is costly, how costly is it? The cost could be so low that the trait might persist even without any benefit i.e. be neutral relative to selection and not an adaptation at all. Or the costs might be very low so that the benefits sufficient to cause the selection and persistence of the trait might be too small for us to really notice, in the coarse current state of speculations on the evolution of language. Of course an adaptation may be low-cost and high-benefit. In particular, it might be that recursive syntax is low cost and has the benefit that Jill de Villiers suggests, and more. Still -- and this remark is addressed not just to her but to all the participants to this symposium –- it would useful to systematically reflect on the cost of language-related traits and not just on their benefits, particularly when the hypothesised benefit does not consists in lowering overall cognitive or communicative costs.

> **For listeners too**
> Jill de Villiers
> May 10, 2004 17:41 UT
>
> It seemed to me that economy considerations did not seem a compelling benefit for recursion. I suggested that the cost to parse-ability for the listener might outweigh the advantages of the speaker's saving of words. So in some sense I already tried to weigh costs and benefits across the communicative dyad, but none of these arguments strike me as more than sheer speculation. Perhaps Dan Sperber is right that some future neurophysiologist could provide the answer in terms of energy use, but it would need to be done for the listener too. I sought the advantage of recursion outside economy considerations, because some of the paths that evolution takes seem to defy such calculations. Surely being vulnerable to dying by inhaling food into one's larynx is a fairly high price to pay for discriminable vowels (Lieberman, 1984)? I argue instead that the advantage of recursion may have arisen from the precision with which it could be used to attribute statements to another speaker, and then, fortuitously, to represent the contents of others' beliefs with the same machinery.
>
> Reference Lieberman, P. (1984) The biology and evolution of language. Cambridge, Ma: Harvard press.

Recursion, theory of mind and communication
Anne Reboul
May 17, 2004 8:34 UT

I would like to begin by saying how much I enjoyed de Villiers' paper. There are two main questions that remain in my mind. The first one is the following: given that some autistic people pass standard false belief tests, have those people been tested on linguistic recursion and, if so, have they passed such things with success? The second question, which is almost a follow-up to the first one is: given that one needs syntactic recursion to pass false beliefs tests and given that it is often claimed that autistic people who pass classic false belief test fail tests that involve higher order beliefs, what is the explanation for that fact? After all, recursion being what it is, though it may meet with memory or processing constraints (i.e. embedding cannot be infinite), there is no reason, once it is mastered, for higher order tests being failed.

Computation is costly?

Jill De Villiers

May 17, 2004 15:41 UT

That's a great comment. First, the work of Helen Tager-Flusberg has shown that the same predictors of false belief reasoning work with high functioning children with autism, namely, their success is linked to their mastery of complement structures in syntax. Her data from a longitudinal study (in press) confirm that beautifully and support the direction of effect. The second question is more troubling, and it is not just children with autism who show the problem you mention. Normally developing children also show a delay in handling e.g. second order beliefs. Why, you ask, if recursion is known? Does your next statement contain the clue: "thought it may meet with memory or processing constraints"? It does seem probable that multiple embeddings strain the processing system. I have argued that not only can clauses be considered to have a "point of view" that can allow the truth value of the embedded proposition to change, but so also do the noun phrases embedded in that proposition. This provides a description of the well-known phenomena of referential opacity, wherein the usual free substitution of coreferential nounphrases is not permitted under intentional verbs:

Oedipus married Jocasta/his mother

Oedipus knew he married Jocasta/*his mother

It seems to be the case that children still fail on referential opacity judgments well after they succeed in understanding statements about false beliefs- perhaps even at age six. Understanding such statements may involve a computation across representations, maintaining appropriate indices of point of view without "lapsing" into one's own point of view. To the extent that this is taxing even to a normally developing child, one would imagine it being especially hard for a child with autism to maintain multiple indices representing minds and their contents. But much more work needs to be done on this subject, in particular, is it as hard in production?

Apperly, I. A., & Robinson, E. J. (1998). Children's mental representation of referential relations. Cognition, 63 (3), 287-309.

de Villiers, J.G. (2001) Extension, intension and other minds. In M. Almgren, A. Barrena, M-J. Ezeizabarrena, I. Idiazabal,& B. MacWhinney (eds) Research in Child Language Acquisition: Proceedings of the 8th Conference of the International Association for the study of child language. Somerville, Ma: Cascadilla press.

de Villiers, J.G. (in press) Can language acquisition give children a point of view? In J. Astington & J. Baird (Eds.) Why language matters for theory of mind. Oxford: Oxford University Press.

Kamawar, D., & Olson, D. R. (1999). Children's representational theory of language: The problem of opaque contexts. Cognitive Development, 14, 531-548.

Larson, R.K. & Ludlow, P. (1993) Interpreted logical forms. Synthese, 95, 305-355.

Russell, J. (1987). "Can we say...?" Children's understanding of intensionality. Cognition, 2, 289-308.

Tager-Flusberg, H. (2000b). Language and understanding minds: Connections in autism. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds). Understanding other minds: Perspectives from developmental cognitive neuroscience (2nd. Ed., pp. 124-149). Oxford: Oxford University Press.

Tager-Flusberg, H. and Joseph, R. (in press) How language facilitates the acquisition of false belief in children with autism. In J. Astington & J. Baird (Eds.) Why language matters for theory of mind. Oxford: Oxford University Press.

Think that
Gloria Origgi
May 18, 2004 13:30 UT

Thanks to Jill for her fascinating evolutionary story. I have enjoyed it a lot. I have a doubt about her concluding hypothesis on the emergence of "think that". She supposes that it emerged from a "strategic" analysis of the speaker's competence and benevolence in saying something once we have discovered that what she says is false.

But this implies that people learn to attribute propositional thoughts once they have become sophisticated (rather than gullible) interpreters. Whereas it seems to me that a pragmatic process of attributing an intention to a speaker implies the capacity to recognize that the speaker "thinks that" even if we take for granted that what she says is true (as in the case of a child who interprets what her mother says without questioning its truth or falsity). That is, there are many contexts of interpretation in which we attribute "thoughts that" to the speakers that don't imply the detection of the truth value of a sentence.

the role of falsehood
Jill De Villiers
May 19, 2004 14:16 UT

I think there are several points in these interesting remarks, and I have tried to wrestle with them. First, notice I tried to separate the evolutionary story from the developmental story, The parallel between them is that both developments rely on using recursion first for verbs of communication and then for verbs of belief. The difference between them is that the evolutionary story might have happened over incalculable generations. It supposes a rare genius who makes the fine distinctions –in Gloria's terms, becomes a sophisticated, strategic, interpreter of speech acts. The child may not in fact have to recapitulate that path. All that is necessary for a child is to acquire the appropriate structures that her language uses to represent and attribute false statements, and then the child has available the necessary structures to represent and attribute false beliefs.

But Gloria asks, why false statements, why not true statements? Don't we attribute thoughts when people speak regardless of the truth value? My answer is three-fold. First, when do we attribute thought behind behavior? Usually, when it is not redundant with an account in terms of observable stimuli. It seems odd to say "Jill thinks she is eating an apple" if it is an apple, and I am eating it. In American Sign Language, the verb 'think' can only mean "think falsely" (J. Pyers, personal communication). Of course we use the term sloppily, and attribute thought in cases of disparity even when the internal process is unlikely to resemble propositional reasoning: "the thermostat thinks it's 30 degrees in here!" "That magnolia tree thinks it's Spring". So, indeed "there are many contexts of interpretation in which we attribute "thoughts that" to the speakers that don't imply the detection of the truth value of a sentence". But this is once we can attribute thoughts.

Second, can a child "interpret what her mother says without questioning its truth or falsity"? I agree there is a default assumption that people speak truly, so if I were in an inner room and someone said to me, "Get your umbrella because it's raining", I would probably do so. But anyone older than two would protest upon reaching the door and seeing only sunshine. (Perhaps I have this entire theoretical bias because my own children seemed obsessed with truth. We tried to tempt my son to eat Campbell's "Chicken with stars" soup at age two, and he protested, "Chicken not have stars!").

Third, there are occasions when we use the words "think that' without truth value being a consideration, and perhaps these are the cases that confound the picture. We use "think that" at least in English as a synonym for intention, e.g. 'He thinks that he'll eat an apple" or for advice, "I think you should clean your office". Notice the embedded clauses are future, or irrealis, and so cannot be matched against the world for truth. Children master these well before the kinds with realis clauses, "He thought he was eating an apple" (de Villiers, in press). If I were to hazard another guess, I suspect this is where the evolutionary story and the ontogenetic story might part company, and also where there is less likelihood of linguistic universals. That is, this type of phrasing may be an overlay of propositional attitude talk that has become habitual in certain cultures but not all, perhaps not even all Western subcultures, and plays no particular role in helping children acquire a mature Theory of Mind. I can even suggest a experiment with three year olds who fail false belief tasks, a new modification of the training studies that have been done. Group 1 hears much talk like this:

He thinks he will clean his room. She is thinking she should eat lunch.

Group 2 hears lots of: He said he caught a fish! (it wasn't). He said it is raining! (It isn't).

My prediction is that Group 1 will not pass false belief tasks any earlier than controls, but Group 2 will. Notice Group 1 gets exposure to mental talk and hears the verb 'think", but there is nothing false involved. Group 2 gets the crucial false complement exposure, but with only over speech acts. Any takers with funding?

Recognizing false beliefs
Robyn Carston
May 20, 2004 18:00 UT

I find Jill De Villiers' paper extremely rich and interesting. However, perhaps for lack of appropriate background (or brain-power), I'm not sure I follow all of it. A central claim is that through becoming familiar with linguistic structures of the sort "S said that P" in situations in which it is evident that P is false, a child develops the capacity to represent others as having false beliefs and so to pass the classic false belief task. How does the analogy (or the 'short step') from the one to the other work? False sayings can have three sources, as outlined at the end of the paper. The relevant one for carrying over to false beliefs is the one that involves the speaker in having made a mistake about how things are in the world. Being able to recognise this in others is explained in the paper in terms of recognising one's own capacity for misperception (seeing or hearing something as something it is not). But in the Sally-Ann false belief task Sally's perception was just fine (she correctly saw the chocolate being put in a certain location) so recognising that (at a later time) her belief is false must involve something else. Perhaps it is the realisation that beliefs (generally) persist in the absence of evidence to the contrary, or that a person's not seeing/hearing or being told that Q (generally) implies their not having the belief that Q.

Evidence and belief
Jill De Villiers
May 21, 2004 20:15 UT

That's a very significant comment, bringing us back to deeper questions about how the child understands evidence for belief in the first place. I repeat that the story I proposed for evolution may not necessarily be the same path that each child goes through in developing the concepts of false belief. There are two stories here, one is about evolution, and a claim that an advantage incurred from certain linguistic representations, namely recursion. Within that larger claim, I conjectured about how a distinction between communication and mental states could have evolved. I envisioned a time in early history when a distinction might have been made between saying and thinking, to differentiate the reasons behind a false statement. I suggested that realization might arise in an individual on occasions when her own misperception occurred and was recognized as such. So it is possible that an understanding of mistaken belief arose out of an empathic simulation of mistaken perception, in evolution. But notice that is only by way of explanation after the individual can represent that the individual said p, where p is false.

It doesn't necessarily follow that the child solving all the false belief tasks is doing so via simulation. My theory about ontogeny is that understanding false beliefs depends on having the representational structure that false complements provide. In a case where an individual needs to know what someone believes, i.e. the content of a belief, that seems to have as prerequisite the representational capacity for recursion of propositions. I have argued elsewhere (de Villiers, in press) that the link between say that p and think that p in the child (the 'short step') is taken care of by linguistic analogy: think is in the right subclass of verbs to take false complements, because it models with say.

But are there other prerequisites? Robyn Carston suggests only one variety of false belief springs from mistaken perception, and there are other cases to be considered such as a failure to update old information, a failure of communication, and so on. All of these are fascinating questions to explore, and our present picture from children is still incomplete. That is, even given the representational structure for handling false belief, might there not be differences in the inferences children draw from the evidence conditions? Is this facilitated at all in languages that have obligatory evidential markers? We know that children can draw the inference that someone who didn't see X in a box doesn't know that X is in the box, and that this inference may be available some months before success on the classic false belief task. Oddly enough, Korean children who mark 'witnessed' on their verbs as a matter of necessity do not succeed any earlier than English speakers (Papafragou, Li & Han, 2003). Why not? Could this be another case where language is the repository of essential ideas that are

acquired by children first in language, but are not immediately available to cognition? Perhaps I have raised more questions than answers. I should note that this topic makes everybody's brain hurt.

Theory of mind and language ability. Understanding the bigger picture

Ted Ruffman (University of Sussex)

(Date of publication: 24 May 2004)

Abstract: It is well established that children's 'theory of mind' correlates with their language ability. However, the majority of correlations have been obtained with verbal questions on false belief tasks as the sole or a prominent measure of theory of mind. Given that these theory of mind tasks are highly linguistic by their very nature, the correlations with language are not terribly surprising. Important questions are whether language ability correlates with something more fundamental in children's theory of mind. How could an incremental increase in language ability result in an apparently more fundamental shift in understanding belief? These and other questions are addressed in this paper.

*This paper has been written with **Lance Slade** and **Mele Taumoepeau**.*

There is now abundant evidence that false belief understanding in children is linked to their language ability. Although there is clearly more to a theory of mind than false belief alone, false belief is typically regarded as the aci test. Some have claimed that it is a particular aspect of syntax, sentential complements, that relates to false belief understanding (deVilliers & Pyers, 2002; Hale & Tager-Flusberg, 2003). Yet there are both logical grounds (Astington & Jenkins, 1999; Ruffman, Slade, Rowlandson, Rumsey, & Garnham., 2003; Slade & Ruffman, in press), and empirical grounds (Perner, Sprung, Zauner, & Haider, 2003) for doubting such a claim. Others have claimed that it is general syntax that assists false belief understanding (Astington & Jenkins, 1999). Yet direct tests of semantics (e.g., receptive vocabulary) versus syntax (e.g., word order) have established that general language ability (e.g., syntax + semantics) is important rather than syntax or semantics per se. This has been established both within a single time point (Ruffman et al., 2003), and longitudinally (Slade & Ruffman, in press).

One of the important questions is why language is linked to theory of mind. We have argued that semantics and syntax both relate to theory of mind because, in normal development, syntax is a clue to semantics and semantics is a clue to syntax. These ideas have been referred to, respectively, as syntactic bootstrapping (e.g., Naigles, 1996; 1990; Naigles & Hoffginsberg, 1995; Pinker, 1984), and semantic bootstrapping (e.g., Rondal & Cession, 1990). The idea is that both syntax and semantics develop rapidly in childhood and assist theory of mind by helping the child to reflect on and refine implicit knowledge about mind. That is, we have proposed the following. In normal development children's initial understanding of mind is likely implicit and manifest in their behavior rather than insights that they can verbalize. This understanding could be innate, but learning likely plays a large role given children's interest in the human face (which acts as a window into the mind), or their observations of the social world which allow them to utilize their capacity for statistical learning (e.g., Saffran, Aslin, & Newport, 1996) to pattern match (i.e., connect various social outcomes with certain preconditions). Over time, and as their language develops, children develop a consciously mediated and verbally based theory on the basis of these implicit intuitions.

We have also argued that children with autism understand the world in a different way (Ruffman, Garnham, & Rideout, 2001). They might lack the initial implicit understanding, and as a result, the social understanding that the gifted group of autistic children eventually develop might be based on processes of self-motivated induction or teaching from others rather than reflecting on implicit intuitions to create explicit theories. One consistent finding is that individuals with autism do not look to the eyes in social situations (Klin, Jones, Schultz, & Volkmar, 2003). This might reflect an absence of implicit knowledge or might be the reason such individuals do not develop implicit social knowledge. If this is correct, then both children with autism and normally developing children develop some sort of an explicit, verbally-based understanding of mind, but children with autism cannot base their understanding on a rich set of implicit intuitions. Further, we have argued that language assists in working out explicit theories because it provides the terminology for thinking explicitly about a mental state (Ruffman et al., 2003). Language allows the child to think explicitly about a person pretending "x" versus thinking "x" (for making fine distinctions between different propositional attitudes and contents), and it facilitates understanding of the causal origins and implications (e.g., subsequent actions) of mental states, enabling explicit predictions. There are a number of research strands that come together to support the above ideas. These are discussed below.

Implicit understanding?

We have obtained evidence that children's initial understanding of false belief is implicit (i.e., not conscious). Children were given a standard false belief task in which the character placed an object in a left-hand location and then went away (Ruffman, Garnham, Import, & Connolly. 2001). While absent, the object was moved to the right-hand location. Children were given three measures: (a) they were asked the standard

verbal question regarding where the character would look for the hidden object, (b) their eye gaze was monitored as to where they looked when anticipating the story character's return (as in Clements & Perner, 1994), and (c) they were asked to "bet" 10 plastic counters in any configuration on the location(s) they thought the story character would return. Like Clements and Perner, we found that young children often looked to the correct (left-hand) location when anticipating the character's return, but explicitly claimed he would go to the right-hand location. The betting measure allowed us to determine whether eye gaze indexed low confidence conscious knowledge (e.g., children understood the character might go to the left-hand location but weren't sure about this so said he would go the right-hand location), or genuinely unconscious knowledge. In the youngest children who showed correct eye gaze but incorrect verbal answers, their eye gaze indexed genuinely implicit knowledge. Despite looking to the left-hand location, these children bet all of their counters on the right-hand location (where they said the character would go). They showed no awareness of the knowledge manifest in their eye gaze. Further, a control condition showed that betting was sensitive to small variations in children's certainty. For instance, they bet with much less confidence when guessing whether an object would be red or green when a bag contained 9 red objects and 1 green (9-1), in comparison to when it contained 10 red objects and 0 green (10-0).

Noverbal understanding in autism.

In a different study (Ruffman, Garnham, & Rideout, 2001), we found that children with autism were unimpaired relative to a group of children with MLD (moderate learning difficulty) on verbal questions tapping social understanding, but they were impaired on a measure of eye gaze. Nevertheless, their eye gaze was unimpaired on the 10-0 and 9-1 tasks (tapping non-social knowledge) described above. Further, we found that in comparison to verbal performance, eye gaze was a much better correlate of teachers' ratings of how severely autistic children were. This finding, along with our finding that eye gaze on the social tasks differentiated the autistic and MLD groups whereas verbal performance did not, suggests that the core insight is the *nonverbal* not the verbal insight.

Is the language-theory of mind relation bi- or uni-directional?

There have been three longitudinal studies relevant to this question. Astington and Jenkins (1996) found that language at an early time point predicted later theory of mind at two of three sets of time points, whereas the reverse relation never held. de Villiers & Pyers (2002) found that language predicted later theory of mind on two occasions over their three time points, and that the reverse relation held once. We found that sampling differences (the language items randomly selected from a composite language measure when evaluating language and social understanding) have a large impact on the end result (Slade & Ruffman, in press). In essence, we found that early false belief understanding was as likely to correlate with later language (receptive vocabulary) as the reverse. This finding fits with the idea that children's theory of mind assists them in learning about word meanings (e.g., Baldwin, 1991).

How language relates to verbal and nonverbal social understanding.

We have found that language tends to correlate more highly with verbal measures than it does with nonverbal measures of social understanding. These studies have been conducted both with normally developing children and with atypical children, and have involved a range of ages and nonverbal measures. In one study, 2- to 4-year-old children were given a task that tapped their ability to understand how someone would react in a situation analogous to the visual cliff (Ruffman, 2000; in preparation). We called this measure the Emotion-Behaviour task and tested 39 children in two experimental tasks. In each task there were two rooms (e.g., red and green) and two associated windows, and children were told one room was safe and the other was not. A boy asked his father whether the red room was safe to enter. In one task the father smiled (a happy emotion expression was placed on his face), and in the other he looked fearful. Having seen this expression, the story character set off for one of the rooms (in the general direction of both). The child had been shown that the story character would eventually appear in the window of the room he entered and we were interested in whether children understood how the father's emotional expression would lead the story character to a particular room. To elicit nonverbal understanding, the child then heard a prompt narrated on audio tape which was meant to direct the child's eye gaze towards one of the rooms in expectation of the story character appearing there ("I wonder which window Sam will go to?"), and we videotaped their anticipatory eye movements. The child was then directly asked a verbal question ("Which window will Sam go to?"). Two-, 3-, and 4-year-olds were all above chance on the nonverbal measure, but only the 4-year-olds were above chance on the verbal measure. Over all children, their composite score on a language measure correlated with verbal performance, $r = .54$, $p < .001$, but not with their eye gaze, $r = .10$, *n.s.*

Recently, we have examined emotion recognition in 40 7- to 9-year-olds. Children were given three blocks of trials. In the predictable block they viewed an emotion face on a computer monitor. There were twelve different faces: two each of fear, sadness, anger, disgust, surprise, and happiness. Each face was comprised of an emotion morph (e.g., 60% sadness, 40% fear). There were also two possible emotion labels at the bottom of the screen, one for each of the emotions comprising the emotion morph (e.g., "sad" and

"fear"). These labels corresponded to two keys on the keyboard. After a few seconds the emotion face disappeared and was replaced by a target emotion word. In the predictable block, the target emotion word always matched the dominant emotion of the emotion face (e.g., "sad"). When the target word appeared, children's task was to press the key on the keyboard that matched the target word as quickly as possible. In the predictable block, because the emotion face and target emotion word always matched, the face could act as a prime and reduce reaction times (i.e., understanding of the emotion present in the face would lead to the expectation that the target word would be "sad").

The unpredictable block was identical, and used the identical emotion faces, except that there was no consistent relation between the emotion face and the target word. Half the time they matched, and half the time they mismatched. Reaction times were expected to be longer because the emotion face wouldn't help predict the target word. The predictable and unpredictable blocks were counter-balanced, and were always followed by the verbal trials in which children were sequentially presented with the 12 emotion faces and two labels (e.g., "sad" and "fear"), and were asked to explicitly label each face.

As expected, reaction times were significantly faster in the predictable block than the unpredictable block. Thus, the emotion faces were priming anticipation of the target emotion word, thereby speeding reaction time. This indicates that children possessed knowledge on some level of what emotions were expressed in the faces. More interestingly, reaction times seemed to be tapping knowledge that was mainly implicit in that faster reaction times on the predictable block were in no way related to the number of emotion faces children explicitly identified on the verbal trials when asked directly to label each face. Thus, whether children were correct when verbally labeling 4 to 6 faces, 7 to 9 faces, or 10 to 12 faces, their reaction times were quicker on the predictable block by the same margin relative to the unpredictable block. In other words, it seemed to be implicit knowledge of the emotions that primed children's expectation as to what target word would appear, and sped up reaction times in the predictable block. As above, language (receptive vocabulary) correlated significantly with the percentage of items correct on the verbal measure, $r = .33$, $p < .05$, but not with the percentage of items correct on the nonverbal measure (the percentage of items in which reaction time was quicker on the predictable block), $r = -.15$, *n.s.*

Finally, we examined language-theory of mind relations in autism and MLD in the study described above (Ruffman, Garnham, & Rideout, 2001). In the children with autism, a composite measure of language correlated with verbal performance on the tasks tapping social understanding, $r = .42$, $p < .05$, but not with the measure of eye gaze on the social tasks, $r = .10$, *n.s.* Only in the children with MLD was the pattern different. In this group, language correlated with both verbal performance, $r = .46$, $p < .01$, and with eye gaze, $r = .37$, $p < .05$.

To sum up over all studies, language correlated with verbal performance on 4 of 4 occasions, whereas it correlated with nonverbal performance on only 1 of 4 occasions. The question is why language correlates more consistently with verbal performance. Some of our tasks have more obvious verbal demands than others. The emotion-behavior task described above, and social tasks used in the study of autistic and MLD children, have relatively high verbal demands because children must follow a narrative and then answer a verbal question at the end (e.g., "Which window will Sam go to?"). These tasks are like false belief tasks which also correlate reliably with language ability. This leads to a trivial explanation of the language-theory of mind relation. Perhaps language correlates simply because children can't follow story narratives or can't parse verbal questions. Yet this explanation does not fully account for the pattern of findings because the emotion labeling task described above has relatively low verbal demands. Children must simply label an emotion photo using one of two labels, yet language correlates with verbal labeling (see also Ruffman et al., 2003), and it does so even at 7 and 9 years of age when the emotion terms are within the children's vocabulary. Thus, the fact that language correlates with verbal performance on all tasks suggests that it is not simply the verbal complexity of a task that leads to the correlation.

A second possibility is that the language-theory of mind relation is mediated by children's age. Both things develop over age and it could be age (experience in the world) that is important to theory of mind rather than language per se. Yet we found that the language-theory of mind relation was robust even after accounting for children's age (Ruffman et al., 2003).

A third possibility is that language is a proxy measure of IQ and that is IQ that correlates with theory of mind. Although this explanation is perhaps partially correct, it is not the full story because in contrast to language, spatial IQ does not correlate consistently with social understanding (Jenkins & Astington, 1996; Ruffman, Garnham, & Rideout, 2001; Tager-Flusberg & Sullivan, 1994).

A fourth possibility is that language is related only to theory of mind insights in the initial stages of development (e.g., amongst preschoolers) when language differences between children are more profound. Yet we found language was related to verbal labeling of emotions amongst relatively experienced language users; 7- to 9-year-olds.

Instead, there seems to be something more fundamental about verbal ability itself. As stated above, we think this is that language is needed to formulate, refine and explicitly reflect on ideas about social concepts. In the experiment described above, children's faster reaction times on the predictable block showed that they possessed some knowledge of the emotion expressed in different faces even though they were incorrect when explicitly labeling the faces. Recall that reaction times did not relate significantly to the child's language, whereas verbal performance did. In other words, reaction times were not improving between 7 and 9 years (r = -.11, *n.s.*), even though language was (r = .40, p < .01), and verbal performance was (r = .42, p < .01). These findings are in keeping with studies demonstrating that implicit learning is as good in 4-year-olds as it is in adults (Vinter & Perruchet, 2000), even though language improves dramatically during this period. That is, it is thought that implicit learning is based on processes of pattern induction or statistical learning (Boucher & Dienes, 2003) rather than language. We have posited the same for eye gaze (and reaction times) in tasks tapping social understanding (Ruffman, 2000).

Once again, the idea is that social knowledge typically develops in an implicit form and only later becomes explicit. The reason it is implicit initially is that it develops through relatively slow processes of induction in which many different bits of social information are gradually pieced together through statistical learning processes to arrive at initial implicit insights. That is, it seems less common for social learning to occur through effortful processes of deliberate theorizing or explicit teaching (although the exception as argued above, is in autism). Indeed, we have found that false belief understanding in 3- to 4-year-olds correlates with children's ability to detect statistical patterns in nonsense syllables (Ruffman & Taumoepeau, in preparation). Children were exposed to a string of nonsense syllables based on Saffran, Aslin, and Newport(1996) while they colored (e.g., …bidakutupirogolabutupiro…). Some strings of syllables repeated to form "words" (e.g., "tupiro") such that the probability of "pi" given "tu" was 1.00, whereas the probability of "go" given "tu" was 0. This statistical information was the only clue to what was a "word". Intonation and spacing were held constant between syllables. Children were later asked to identify words such as "tupi" or "tugo" as 'old' or 'new'. Although the false belief task was not an implicit task, performance on the statistical learning task correlated significantly with false belief even after controlling for language ability, pr = .26, p < .05. Again, we hypothesize that this is because explicit false belief is initially based on implicit insights and implicit learning is statistical learning. And to repeat the ideas discussed above, once implicit understanding is in place, the first children to develop explicit understanding are those with better language skills because language provides the terminology to reflect on and refine implicit intuitions.

References

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology*, *35*, 1311-1320.

Baldwin, D. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, *62*, 875-890.

Boucher, L., & Dienes, Z. (2003). Two ways of learning associations. *Cognitive Science*, *27*, 807-842.

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, *9*, 377-395.

de Villiers, J. G., & Pyers, J. E. (2002). Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cognitive Development*, *17*, 1037-1060.

Hale, M., & Tager-Flusberg, H. (2003). The influence of language on theory of mind: A training study. *Developmental Science*, *6*. 346.

Jenkins, J. M., & Astington, J. W. (1996). Cognitive factors and family structure associated with theory of mind development in young children. *Developmental Psychology, 32*, 70-78.

Klin, A., Jones, W., Schultz, R., & Volkmar, F. (2003). The enactive mind, or from actions to cognition: Lessons from autism. *Philosophical Transactions of the Royal Society of London: B*, *358*, 345-360.

Naigles, L. (1990). Children use syntax to learn verb meaning. *Journal of child language*, *17*, 357-374.

Naigles, L. R. (1996). The use of multiple frames in verb learning via syntactic bootstrapping. *Cognition*, *58*, 221-251.

Naigles, L. R. & Hoffginsberg, E. (1995). Input to verb learning: Evidence for the plausibility of syntactic bootstrapping. *Developmental Psychology*, *31*, 827-837.

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 Villier's linguistic determinism on german-speaking children. *Child Development*, *74*, 179-188.

Pinker, S. (1987). The bootstrapping problem in language acquisition. In B. MacWhinney (Ed.), Mechanisms of language acquisition. Hillsale, NJ: LawrenceErlbaum Associates.

Rondal, J. A., & Cession, A. (1990). Input evidence regarding the semantic bootstrapping hypothesis. *Journal of Child Language*, *17*, 711-717.

Ruffman, T. (2000). Nonverbal theory of mind: Is it important, is it implicit, is it simulation, is it relevant to autism? In J. W. Astington (Ed.), *Minds in the making: Essays in honor of David R. Olson* (pp.250-266). Oxford: Blackwell.

Ruffman, T., Garnham, W., Import, A., & Connolly, D. (2001). Does eye direction indicate implicit sensitivity

to false belief?: Charting transitions in knowledge. *Journal of Experimental Child Psychology*. *80*, 201-224.
Ruffman, T., Garnham, W., & Rideout, P. (2001). Social understanding in autism: Eye gaze as a measure of core insights. *Journal of Child Psychology and Psychiatry*, *42*, 1083-1094.
Ruffman, T., Slade, L., Rowlandson, K., Rumsey, C., & Garnham, A. (2003). How language relates to theory of mind. *Cognitive Development, 18*, 139-158.
Ruffman, T., & Taumoepeau, M. (in preparation). False belief understanding is assisted by statistical learning.
Saffran, J. R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science*, *274,* 1926-1928.
Slade, L., & Ruffman, T. (in press). How language does (and does not) relate to theory-of-mind: A longitudinal study of syntax, semantics, working memory and false belief. *British Journal of Developmental Psychology*.
Tager-Flusberg, H., & Sullivan, K. (1994). Predicting and explaining behavior: A comparison of autistic, mentally retarded and normal children. *Journal of Child Psychology and Psychiatry*, *35*, 1059-1075.
Vinter, A., & Perruchet, P. (1999). Implicit learning in children is not related to age: Evidence from drawing behavior. *Child Development*, *71*, 1223-1240.

Discussion

Language and theory of mind: implicitness and explicitness
Anne Reboul
May 27, 2004 9:17 UT

There are plenty of remarks and data which are highly interesting in Ruffman et al.'s paper and I'm sympathetic with the global thesis that though ToM exists in an implicit form (perhaps partly based on implicit learning) before language, language nevertheless plays a role in its complete development by allowing this implicit knowledge to become explicit. However, there still is a question there: granted that this is the role of language, this suggests an explanation of the correlation between linguistic performance and false belief performance, to wit that to succeed in false belief, one needs to be able to explicitly represent other people's (false) belief. Thus, younger children implicitly recognize the right answer (as shown by gaze direction) but are not able to explicitly access it and give the correct verbal answer. However, this does not necessarily explain why autistic children fail the false belief test (in either the standard or higher order versions): if the role of language is to allow the child to explicitly represent a conscious ToM, one would expect autistic children or adults with verbal abilities to get a ToM later on. After all, one can acquire knowledge implicitly, but one can also acquire it explicitly, through language for instance. Given that autistic children generally learn language rather than acquire it, one would expect them to similarly learn ToM rather than acquire it. One final remark: Ruffman et al. note that the basic implicit ToM plays a role in language acquisition and put the difference between autistic and normal children precisely there. Now, it is of course true that autistic children do not acquire language in the standard way and that this could be explained by a deficit in implicit basic ToM. However, it seems that Asperger patients, who also suffer from a deficit in ToM acquire language more or less normally. If this is the case, given the hypothesis that basic ToM is involved in language acquisition and in success in false belief test, why do Asperger patients succeed in language acquisition, while they still fail the false belief test?

> **Response to Anne Reboul**
> Ted Ruffman
> May 28, 2004 15:21 UT
>
> Anne Reboul wonders about the following: If the role of language is to assist explicit understanding of false belief, then shouldn't individuals with autism eventually learn about theory of mind? In fact, there is evidence that they do. For instance, in British special schools teachers have been quick to pay attention to psychological research. Children with autism are now taught extensively about mental states and emotions. It is not surprising, then, that we have recently found (in preparation) that children with autism did significantly better than a control group matched for verbal ability on measures of emotion recognition and false belief. Indeed, other researchers have frequently found children with autism performing at least similarly to control groups (see the introduction to Ruffman, Garnham, & Rideout, 2001). So, clearly, children with autism do learn to pass these tasks. In this situation their understanding was also not completely dependent on their language since language in the two groups

was equivalent but emotion recognition was superior in the children with autism. Thus, teaching has a role to play over and above basic language abilities. In sum, children with autism do learn something about theory of mind although they likely do so more slowly for a number of reasons such as (a) their verbal abilities are generally impaired (though less so in Asperger's Syndrome), (b) they don't pay attention to eyes as much as non-autistic individuals (see Ami Klin's work), (c) they often don't engage much in social interactions, and (d) their brain structures are different (e.g., their amygdala) so that the meaning of social information might not always be grasped.

An important question, though, is how deep their learning goes. We think they sometimes might have just learned to pass particular tasks. For instance, we (in preparation) found that about 2/3 of autistic children were correct on a false belief task but only 1/3 passed a true belief task. True belief should be conceptually simple because the story character witnessed the object's transfer from one location to the other and hence should know it is in the new location. Indeed, the pass rates were reversed in the non-autistic children (1/3 passing false belief and 2/3 passing true belief), a more sensible outcome. It was as if children with autism recognized a paradigm which they mistook for false belief (the paradigm their teaching had utilized), and then ascribed a false belief again. This suggests that a deeper understanding of belief was not achieved despite the teaching. Another finding comes from the Ruffman, Garnham, and Rideout (2001) paper. Eye gaze was better than verbal performance at: (a) differentiating autistic and non-autistic children than verbal performance, and (b) correlating with the severity of autistic symptoms. Again, children with autism can learn to pass theory of mind tasks, but it is an open question how much good this does them.

The second question Anne Reboul raises is this: if ToM helps language acquisition and Asperger's individuals have an impaired theory of mind, why is their language acquisition relatively normal? Baldwin (1991) hypothesizes that understanding intention is integral to language learning (e.g., recognizing what object is intended when a label is uttered). Many have argued that normally developing children use intentions to grasp word meanings by at least 18 months of age (e.g., Tomasello, Strosberg, & Akhtar, 1996). Asperger's individuals have impairments in theory of mind but often pass false belief tasks, a considerably later achievement in normal development. Therefore, the level of theory of mind required for language acquisition is likely fairly basic and well within the grasp of Asperger's individuals. Thus, they acquire language at a relatively normal rate. Further, there are many other routes to language other than theory of mind, for instance, statistical learning, rich and plentiful mother input, etc. These abilities might also be equivalent in Asperger's individuals, and so their language acquisition is unimpaired.

References

Baldwin, D. (1991). Infants' contribution to the achievement of joint reference. Child Development, 62, 875-890. Ruffman, T., Garnham, W., & Rideout, P. (2001). Social understanding in autism: Eye gaze as a measure of core insights. Journal of Child Psychology and Psychiatry, 42, 1083-1094. Tomasello, M., Strosberg, R., & Akhtar, N. (1996). Eighteen-month-old children learn words in non-ostensive contexts. Journal of Child Language, 23, 157-176.

Implicit beliefs?

Jill and Peter de Villiers
May 29, 2004 16:58 UT

There are several questions that are raised by Ted Ruffman's paper. First, there is the question of whether one might expect language in general to effect theory of mind development. There are many possible theories here ranging from the fairly trivial to the deeper: a) the tests require language b) language helps the child contain impulsivity in the tasks c) the child learns the theory from the language about the mind d) the child learns to "talk the talk" of the culture e) the words focus the child's attention on the concepts

It is not clear on any of these accounts why syntax ("word order and the like") would have any particular role to play over and above word meaning, and indeed that is what Ruffman has found, as in our own work. This position must be distinguished from the particular position we have espoused, that sentential complements (tensed, not irrealis) are the crucial bootstrap (de Villiers, in press). Since Ruffman never tested sentential complements, he did not test that hypothesis. He chose not to because he believed sentential complements, even those involving communication verbs, already require false belief reasoning in order to generate the answer. Why then do we find that mastery of the syntax comes first? The only reason left is that perhaps the language tasks are "easier" than the standard false belief tasks.

But consider the now considerable evidence from deaf children, which Ruffman ignores. In the case of deaf children with language delay, one might expect nonverbal false belief tasks would be passed before the language tasks, because language is where the difficulty lies. But the nonverbal tasks pattern with the standard verbal tasks and BOTH are predicted by mastery of sentential complements (de Villiers & de Villiers, 2003; P. de Villiers, in press). This does not bode well for the idea that the complement task requires false belief reasoning, which is accomplished first.

Second, the data are at least somewhat problematic for the idea that implicit understanding of false beliefs is present "in the eyes" for young children before they can develop it explicitly. Deaf children are highly dependent on vision for making sense of the social and physical world, so one might expect this skill of detecting behavioral regularities to be at a premium for them. But they are found to be delayed concomitant with their language skills, in the syntactic domain we have theoretically isolated as necessary. Fortunately for some other theories, vocabulary is also an important independent predictor for the deaf children.

Third, if eye gaze is such an important sign of implicit knowledge, why does it not spill over into any other behavioral sign, such as pointing, answering questions or betting? Povinelli and Vonk (2003) give an important analysis of the work on great apes in which they set rigorous standards for deciding when a creature is responding on the basis of noticing behavioral regularities as opposed to attributing belief to another individual. Interestingly, we have not applied the same rigor in the child development literature (those apes always get the harder tests!). But it is because so many of our "tests" are not really behavioral – we pick up dolls that display no facial expressions, or we read a lean story with crude pictures, or we enact a hide-and-seek game with blank expressions and blindfolded stooges -and demand that the child "project" whatever theory he has developed about behavior onto these largely symbolic manifestations of real behavior. This leads some to suggest the tasks are too artificial, but on the plus side, we are free from the difficulty of trying to separate "behavioral regularity" and "belief" for the child. But Ruffman's tests require that scrutiny.

We agree that there is something right about implicit knowledge. One area in which we have found deaf children succeeding remarkably well is in the domain of the sticker-hiding task (the "penny-in-the-hand" game, Baron-Cohen, 1992), generally assumed to be equivalent to the false belief tasks for typically hearing children. This task stubbornly refuses to conform to the pattern of the other nonverbal false belief tasks we have used, in that deaf children do very well in trying to "trick" the experimenter into choosing the wrong hand, and their cleverness is quite independent of their language skill. The clue comes from a second deceptive pointing task (Moses, 2001) in which they also succeed, suggesting it is deception that is the common factor. Does producing a deceptive response entail figuring out what your partner will do on the basis of a false belief? Or is this the best manifestation of an implicit induction of behavioral regularities, that is a precursor to, but not enough for. false belief reasoning?

Reply to Jill and Peter de Villiers Part 1
Ted Ruffman
Jun 1, 2004 16:28 UT

Jill and Peter de Villiers raise several points in their commentary. First, they point out that we have argued that tensed complements require false belief understanding, and query why it is, then, that an understanding of tensed complements precedes success on false belief tasks (de Villiers & Pyers, 2002). There is a simple explanation of this finding which does not place any special value on acquiring tensed object complements. de Villiers & Pyers' complementation tasks are essentially stripped down versions of false belief tasks without extraneous task demands. Thus, unlike the classic false belief task the child does not have to follow a relatively complex story and narrative. Nor does a complementation task require the child to justify why the story character will hold a false belief (one of the false belief measures in deVilliers & Pyers). Justification is notoriously difficult for children. Further, unlike false belief tasks, the complementation task involves making the mental state of the protagonist salient and it includes an absence of the actual object in question, two factors that Wellman et al. (2001) show have a significant effect on improving false belief performance. Hence, tensed object complement tasks are more sensitive to early false belief understanding, and come in first (Slade & Ruffman, in press). Yet we still argue they are essentially false belief tasks.

Second, de Villiers and de Villiers point out that deaf kids find nonverbal false belief tasks difficult. They argue that this shouldn't be the case on our account because nonverbal tasks are supposed to be easier than verbal tasks. The answer to this question lies in differentiating different types of "nonverbal" tasks. "Nonverbal" does not necessarily entail "implicit", and it is only implicit knowledge that we claim is independent of verbal ability. We have shown that earlier eye gaze in false belief tasks

is typically implicit in that children have no awareness of the knowledge manifest in their eye gaze (Ruffman, Garnham, Import, & Connolly, 2001). Eye gaze is an automatic response to the story narrator's prompt in which the narrator wonders out loud where the story character will return to get his hidden object. The child's gaze is not in response to a direct question about where the story character will look for his hidden object. It indexes an implicit anticipation as to where the story character will return. In this way it is very different from verbal questions (e.g., "Where will Sam look for his cheese?") or betting measures in which the child is explicitly asked to place chips at the location Sam will return to. The same is true of the nonverbal tasks used with deaf kids. Here kids are directly questioned where the story character will look. The key distinction is not whether a task is verbal or non-verbal, but whether a task taps explicit versus implicit knowledge. Direct questions tap explicit knowledge whereas indirect measures have the potential to tap implicit knowledge. We are not surprised, then, that sentential complements relate to performance on certain types of "nonverbal" tasks, as long as the nonverbal tasks pose direct questions to children and therefore tap explicit knowledge.

Reply to Jill and Peter de Villiers Part 2

Ted Ruffman
Jun 1, 2004 16:31 UT

Jill and Peter de Villiers also question whether eye gaze genuinely taps "belief" understanding as opposed to having some simpler basis in behavioural regularities. While not exhausting all possibilities, we have in fact gone some way towards testing such ideas. For instance, we wondered whether eye gaze might tap an understanding of a more basic understanding that one who hasn't seen the object moved would "do the wrong thing" and therefore go to the wrong location (behaviour rather than belief). In the standard two-location task it isn't possible to separate this kind of explanation from a belief explanation because the belief and behaviour hypotheses both lead the character to go to the empty location. Thus, we included a third location (Garnham & Ruffman, 2001). The object starts in location A. The story character then watches it moved to B. When the story character is absent, the child watches the object moved to C. The child then hears the narrator's prompt and we measure anticipatory eye gaze as to where the character will return. Both A and B are consistent with the idea that the story character will do the wrong thing. He hasn't seen the transfer, doesn't know where the object is, so he should be as likely to go to A as B. Yet children look to location B, not location A. Only location B is consistent with the character's "belief", suggesting that children reason on the basis of beliefs rather than a lack of knowledge.

Finally, de Villiers and de Villiers point out that deaf kids do relatively well when deceptively hiding stickers, and their cleverness on this task is independent of their linguistic skill. Similarly, they point out that deceptive pointing is also an early marker of false belief. They ask whether deception requires an understanding of false belief or of behavioural regularities. de Villiers and de Villiers are right to raise this question. In normal development, measures of deception seem to vary in terms of difficulty. Dunn (1989) reports incidents of 2-year-olds acting deceptively (e.g., denying an act they have in fact committed). We and others have argued that many such deceptive acts need not entail an understanding of belief (Perner, 1991; Ruffman, Olson, Ash, & Keenan, 1993). The child might simply know this is a way of avoiding trouble, that is, of changing others' behaviour. Alternatively, deception might involve an understanding of knowledge rather than belief. In hiding a sticker, a child need not think explicitly about what a viewer will think, but only about what he knows. Does he know the sticker is in a particular hand, for instance? Yet another alternative is that deception might tap an understanding of belief but be easier than other belief measures because it is more likely to tap implicit knowledge (again, because the child is not asked directly what someone will think as a result of the deceptive act). In fact, we note that when children are compelled to explicitly consider how deceptive acts affect others' beliefs, they show no advantage on measures of deception relative to standard measures of false belief (Ruffman et al., 1993). As above, this suggests that it is not whether a measure is "verbal" or "nonverbal" that is crucial, but whether the measure requires the child to explicitly reflect on someone's belief.

Recognition of basic emotions vs. surprise

Anne Reboul
Jun 1, 2004 10:53 UT

Thanks to Ted Ruffman's for a very interesting and complete answer to my original comments. There is one last thing I wanted to ask. In his 1995 book, Baron-Cohen reports that though autistic children did recognize "basic" emotions, such as pleasure or distress, which do not prominently involve belief attribution, they did

very poorly at recognition of emotions that do involve belief, such as for instance surprise. You report in your reply to my first comment on the surprisingly good performance of autistic children who have undergone an extensive teaching on emotion recognition. The question was the following: were these children also tested on surprise recognition and, if so, was their performance on that specific emotion comparable or similar to their performance for other emotions recognition?

Reply to: Recognition of basic emotions vs. surprise

Ted Ruffman
Jun 1, 2004 13:33 UT

Anne Reboul asks about children with autism's recognition of surprise in our studies. We tested a group of children with autism and children with moderate learning difficulty (matched for verbal ability) using surprise faces (as well as fear, anger, disgust, happiness, and sadness). In total, there were 12 faces for each emotion. Children with autism were impaired only on disgust. They were superior when recognising facial expressions of all other emotions, including surprise. Why we obtained differences from Simon Baron-Cohen et al.'s (1993) study we are not sure. However, more than 10 years have passed since Baron-Cohen et al.'s study and it is likely that teachers have changed their teaching methods in that time given the findings of psychological studies indicating emotion recognition deficits in children with autism. We cannot be sure of this explanation but think it is at least very plausible.

Word learning without Theory of Mind. Possible, but useless

Gil Diesendruck (Bar-Ilan University, Israel)

(Date of publication: 7 June 2004)

Abstract: A number of psychological mechanisms have been suggested to account for children's word learning. In the present paper, I will argue that while some of these mechanisms might account for word learning in special cases, an understanding of minds is what de facto guides normally developing children's word learning. I will review evidence showing that children rely on cues about speakers' intents, rather than cues pertinent to the other mechanisms, not only for establishing the referents of words, but also for interpreting their meanings. Possible implications for the evolution of these capacities will then be discussed.

In a classic *New Yorker* magazine cartoon, a drowning boy is depicted crying to his dog, "Lassie, get help!", as the dog observes from the edge of the water. In the next panel, we find Lassie lying on a psychiatrist's couch. In the present paper, I argue that word learning without theory of mind is useless in two senses. First, because there seem to be important differences between words acquired via a theory of mind mechanism and words acquired without such a mechanism. In fact, if one of the main functions of words is to express a speaker's disposition towards and about the world and by doing so impact a listener's disposition towards and about the world, then the latter kind of words might be of little use. This is the sense of "useless" illustrated in the cartoon. And second, because typically developing children – who have at their disposal a theory of mind – will favor it when learning words over any other acquisition mechanism. I will first review evidence in support of these two contentions. I will then discuss how theory of mind gets recruited to the job of word learning, arguing that the mapping is not trivial. This discussion will carry implications for the debate on the coevolution of theory of mind and language capacities.

Before I get into the evidence, let me clarify briefly what I mean by "word learning" and by "theory of mind" (ToM). Word learning is the process by which a cognitive system comes to know the linguistic symbols intentionally associated with concepts. As linguistic symbols, words are arbitrary and conventional forms that stand for particular concepts. That is, words are not simple associations between sounds and percepts. Rather, words have the potential to evoke abstract, absent, and generalizable ideas or actions. Last but not least, words are intentional inasmuch as they manifest a particular disposition of the speaker towards the referents of words.

As for ToM, I do not mean the full-blown mindreading, representational, and metacognitive capacities attributable to adults and perhaps even to 5-year-olds. I do not mean false-belief understanding. Rather, I use ToM to refer to all sorts of attributions a cognitive system makes regarding the mental computations performed by a different behaving cognitive system. For infants, this might be limited to a sensitivity to other people's intentions, desires, and emotions.

Word learning without ToM

It is a fact: organisms without ToM can acquire words. Not only Lassie the dog, but also Alex the parrot (Pepperberg & Wilcox, 2000), Kanzi the bonobo (Savage-Rumbaugh, Murphy, Sevcik, Brakke, Williams, & Rumbaugh, 1993), and various other animal species have been shown to understand, and some even produce, word-like symbols. While I agree that these are striking discoveries regarding the cognitive capacities of these animals, it is highly debatable whether these "words" have the paradigmatic characteristics of words described above. One likely obvious shortcoming of this type of words is their capacity to refer to or describe abstract or absent concepts. A further shortcoming more central to the present argument regards the pragmatic characteristics of such words. I will not extend on this matter because others in this forum have already done so in excellent fashion (see Fitch). Reiterating his point, most of the empirical data available suggests that when communicating, animals do not take into account a listener's state of knowledge or disposition towards the world. Their use of word-like symbols serves only to express what they perceive.

Word learning without ToM is not exclusive to nonhuman animals. Certain human children – namely young infants and children with autism – might do so too. Recent work using looking time paradigms shows that under tightly controlled experimental situations, children as young as 14-months of age can learn word-to-object pairings (Schafer & Plunkett, 1998; Werker, Cohen, Lloyd, Casasola, & Stager, 1998). For instance, in Werker et al.'s (1998) study, after habituating to a specific word-object pairing, infants looked longer when either word or object changed. But as Werker et al. carefully conclude, before we can grant that these infants truly learned words, we need first to show that these associative capacities are somehow unique to words as opposed to general to any sound-visual pattern pairing. Second, there would have to be evidence that the

infants generalize the words to other similar objects or events. And third, I would add, there needs to be evidence that these words are used to express various intents.

But perhaps the most striking example of word learning without ToM, and at the same time the clearest case of the consequent pragmatic uselessness of words acquired in such a way, comes from studies on children with autism-spectrum disorders. On the one hand, these children, especially those with Asperger syndrome, can acquire a sizeable vocabulary, yet arguably without help from ToM. For instance, Baron-Cohen, Baldwin, & Crowson (1997) showed that children with autism have difficulty associating a novel label to an object, when the label's referent is not the object they are attending to but rather an object the speaker is attending to. More recently, Preissler (2003) reported that the word learning difficulties exhibited by children with autism in a task similar to Baron-Cohen et al.'s (1997), were related to their difficulties in a non-verbal intentional understanding task. These latter difficulties notwithstanding, Preissler found that these children did avoid assigning two labels to an object, presumably contradicting the claim that this response results from inferences about speakers' referential intents (Clark, 1990; Diesendruck & Markson, 2001). On the other hand, a central characteristic of these syndromes is the impairment in communication: language is commonly not used to express the range of possible dispositions towards the world, and to influence others' dispositions towards the world.

Taken together, the main point to be taken from this brief review is that when words are acquired without ToM, they are relatively useless for the sake of effective communication. A secondary implication of this review, however, is that it leaves open the question of what mechanisms animals, infants, and children with autism *do* use to acquire words. The word learning literature offers two candidate mechanisms: associative learning and lexical constraints.

Associative learning seems to be a plausible candidate account for the word learning found in all these different populations. The small size of vocabulary acquired (except for children with Asperger), the intensive amount of training required (in all cases), the specificity of word-object links (in young infants), and the limited reliance on inferential processes (in all cases), are all consistent with an associationist mechanism.

As for lexical constraints, studies on parrots (Pepperberg & Wilcox, 2000), bonobo monkeys (Savage-Rumbaugh et al., 1993), and on children with autism (Preissler, 2003), imply that a form of mutual exclusivity might be present in these different populations. Questionable in this regard is to what extent these reported mutual-exclusivity-like phenomena – namely, the avoidance of two names for a single referent – result from the intensive training in labeling that participants in these studies underwent. For instance, in Preissler's (2003) studies on children with autism, participants performed 16 baseline/training trials before the actual testing trials in which they eventually avoided two names for an object. For the sake of comparison, people with Williams Syndrome show a similar response pattern after only two familiarization trials (Stevens & Karmiloff-Smith, 1997). These differences are consistent with the possibility that while the mutual-exclusivity-like phenomena found in parrots, bonobos, and children with autism might result from associative learning, those found in other populations might have a different source – arguably ToM.

Word learning when ToM is available

A current dominant view in the field is that children rely on multiple cues when acquiring words.For instance, it has been argued that children rely on lexical-specific constraints (Markman, 1989), syntactic information (Hall & Graham, 1999), and mechanisms of attention and memory (Smith, Jones, & Landau, 1996). At the same time, a number of researchers emphasize how cues to speakers' communicative intents influence this acquisition process (Akhtar & Tomasello, 2000; Baldwin & Moses, 2001; L. Bloom, 1998; P. Bloom, 2000). An important question that arises from these theoretical perspectives regards the relative power of these different cues. A direct way to address this question is to evaluate how children respond when faced with conflicting cues.

ToM vs. constraints. Mutual exclusivity is one of the lexical constraints presumed to guide children's word learning (Markman, 1989). Mutual exclusivity stipulates that children avoid assigning two names for an object. In a series of studies, Diesendruck & Markson (2001) investigated an alternative explanation for this type of response. Following Clark's (1990) Principle of Contrast, we hypothesized that this response might result from children reasoning that when speakers use two different names – or any other referential symbols – it is because they likely have different referential intents in mind, and consequently the names likely refer to different objects. In support of this hypothesis, we found that this response was not unique to the case of names, but also appeared when children were asked to assign facts to objects. That is, the response was not lexical specific but instead more general to referential contexts. Furthermore, we found that children's avoidance of assigning two facts to an object disappeared when one of the speakers asking children for an object associated with a fact, did not know that the other object presented to children already had a fact

associated with it. That is, children took into consideration a speaker's state of knowledge in order to infer his referential intent.

A recent study extended this conclusion also to the case of object names (Diesendruck, 2004). In this study, the state of knowledge of a speaker was manipulated by varying whether the speaker was bilingual or monolingual. Specifically, a bilingual speaker taught bilingual children a novel *English* name for one of two novel objects. In two critical conditions, a different speaker, who was present during the introduction of the English name, then asked children for the referent of a novel *Hebrew* name. The only difference between the conditions was whether this second speaker was himself bilingual or monolingual. We found that when the second speaker was bilingual, children tended to select the as-yet unnamed object as the referent of the Hebrew name. When he was monolingual, however, children selected randomly between the objects. Children seem to have reasoned that given the fact that the second bilingual speaker understood the English name provided by the first speaker to one of the objects, then his use of a different name – even if in a different language – probably reflected a different referential intent. When the second speaker was monolingual, however, children assumed he did not understand the English name used by the first speaker, and thus could not draw a clear inference about his referential intent.

These studies indicate that at least in these limited contexts, children's responses seem to be guided by inferences regarding speakers' intents rather than by a priori lexical specific constraints.

ToM vs. syntactic cues. An important source of information for children about the meanings of new words is the syntactic or morphological frame in which words are embedded (Brown, 1957). For instance, young children interpret novel words differently if the words are introduced as a count noun (e.g., "This one is a zav"), a proper noun (e.g., "This one is named zavy"), an adjective (e.g., "This one is very zavy"), or a verb (e.g., "This one is zaving"). In fact, Hall & Graham (1999) found that children rely on such syntactic cues when inferring the relation between labels. Specifically, they found that under certain circumstances, children accept a proper name and an adjective for the same entity, implying that perhaps children allow two words for the same object as long as they are from different form classes.

A second study reported in Diesendruck (2004), investigated the effect of speakers' knowledge state on children's inferences of the meanings of words from different form classes. In that study, an experimenter taught children either a novel proper noun or a novel count noun for one of two unfamiliar animate creatures. A second speaker, who was either present or absent during the experimenter's naming, then asked children for the referent of a different name. We found that when the second speaker was present, then it did not matter the lexical form of the noun taught by the experimenter: in both cases children selected the as-yet unnamed creature in response to the second speaker's request. The crucial finding came when the second speaker was absent when the experimenter named one of the creatures. In these circumstances, the lexical form of the name used by the experimenter mattered. Specifically, when the name was a count noun, children again selected the as-yet unnamed creature in response to the second speaker's request. However, when the name was a proper noun, children selected randomly. Children seem to have reasoned that count nouns are conventional linguistic forms known by all speakers of a language, and thus even an absent speaker is presumed to know them (Clark, 1990; Diesendruck & Markson, 2001). In contrast, proper nouns are only known by individuals familiar with the referents, and thus an absent speaker new to the situation might not know the nouns (Birch & Bloom, 2002). This is an initial piece of evidence that inferences about speakers' state of knowledge and consequent referential intents, modulate children's reliance on syntactic cues to the meanings of words.

ToM vs. attentional mechanisms. Given the presumed centrality of associationist and attentional mechanisms for word learning in ToM-less organisms noted above, this conflict is of particular theoretical relevance. Indeed, this conflict has received considerable empirical attention, most of which supporting the claim that children privilege cues to speakers' intents over automatically attention-grabbing associations.

Baldwin's (1991, 1993) studies with 18-month-olds showed that temporal contiguity was not necessary for a name-object link to be formed, and that a link was only formed when an interacting person presented the name. These studies further revealed that children associate a novel name not with the object they are attentive to but rather with the object the speaker is attentive to. In fact, 2-year-olds rely on social cues (e.g., eye gaze) for linking names to objects, even when these cues direct children away from salient attention-grabbing aspects of the situation (Hollich, Hirsh-Pasek, & Golinkoff, 2000; Moore, Angelopoulos, & Bennett, 1999).

In contraposition to these studies, Samuelson & Smith (1998) offered an attentional learning account of phenomena previously proclaimed as resulting from intentional inferences. Specifically, Akhtar, Carpenter, & Tomasello (1996) reported that 2-year-olds relied on a speaker's state of knowledge to infer that a novel name referred to the one object, out of four, that the speaker had not seen. Samuelson & Smith argued that this response was due to the novelty of the context in which the target object was presented, which made it memorable and attention-grabbing. In a recent study, Diesendruck, Markson, Akhtar, & Reudor (2004) demonstrated that the response was not simply due to a change in context. Specifically, 2-year-olds

associated a novel name with the target object only when the change in its presentation context resulted from an intentional rather than an accidental action. Moreover, even when the change was intentional, children made the association only when the speaker doing the naming was the one who produced the change. The association was not formed when a different speaker provided the name.

Summarizing this section, in most of the studies in which ToM was directly pitted against other cues, ToM seemed to come out as the dominant mechanism used by children to learn words.

How is the mapping of ToM to word learning done?

Having argued for the centrality of ToM in word learning, what is left to discuss is how children recruit ToM for this task. It seems to me that this mapping involves at least three steps. First, children need to have a sufficiently developed ToM *prior* to substantial vocabulary acquisition. Second, children need to adequately employ their ToM for the sake of understanding *human* actions. And third, children need to realize that words are a kind of intentional human "product", which they should consequently deal with through their ToM capacities.

Do young children have the ToM it takes? The short answer is, "yes." There is evidence that 9-month-olds encode the goals of agents (Woodward, 1998), that 10-month-olds parse dynamic actions in terms of intentional structure (Baldwin, Baird, Saylor, & Clark, 2001), that 14-month-olds distinguish between accidental and intentional actions (Carpenter, Akhtar, & Tomasello, 1998), and that 18-month-olds attribute different intentions, emotions, and desires to different people (Meltzoff, 1995; Moses, Baldwin, Rosicky, & Tidball, 2001;Repacholi & Gopnik, 1997).

Do infants restrict ToM to humans? The answer here is, "may be not from the start!" On the one hand, a few of the studies cited earlier indicate that intentional attribution seems to be restricted to humans (Meltzoff, 1995; Woodward, 1998). On the other hand, a number of studies show that young infants attribute goal-directed behavior not only to humans but also to plain circles moving on a computer screen (Csibra, Gergely, Biro, Koos, & Brockbank, 1999), and to other non-human objects (Johnson, Booth, & O'Hearn, 2001). In fact, 1-year-oldseven rely on previously witnessed behavior of non-human agents to interpret these agents' subsequent actions in a different context (Kuhlmeier, Wynn, & Bloom, 2003). Whatever the reasons for the differences in findings or for infants' "overextension" of ToM to non-human agents, the studies intimate that during the first 12 to 18 months of their lives, infants' ToM capacities undergo fine tuning, so as to eventually become consistently and systematically recruited for the processing of human actions.

When do infants treat words as intentional expressions? Recent studies indicate that early in the word learning process, infants treat a variety of symbols as equally suitable for expressing referential intents. In particular, it has been found 12- to 18-month-olds treat emotional expressions, sounds, and gestures as referential (Campbell & Namy, 2003; Moses et al., 2001; Namy & Waxman, 1998; Woodward & Hoyne, 1999). Importantly, between 20- to 26-months of age, children start narrowing this range of acceptable referential symbols to words (Namy & Waxman, 1998; Woodward & Hoyne, 1999).

Evidently, then, all three steps involved in the mapping of ToM to word learning are not trivial and take time. Curiously, the three processes converge to reach the minimal level of required sophistication at around the 18 to 24 month age period, perhaps not coincidentally a period in which substantial vocabulary expansion occurs.

Implications for the coevolution of language and ToM

The argument presented so far has straightforward implications for the coevolution issue. Yes, "words" could be acquired without ToM. But these words probably would have very little similarity to what we standardly consider as words. For this latter kind of words to be acquired, a basic mindreading capacity had to be in place. Moreover, once such a capacity was in place, humans would typically and consistently recruit this capacity for learning words. This conclusion notwithstanding, as the last section of my paper points out, such recruitment might not be done automatically. Rather, the recruitment seems to involve a series of specific mappings. The fact that all the processes involved in the mappings develop independently with unique functional properties, supports the hypothesis that there need not be a special/modular mechanism devoted only for the intentional processing of words (Bloom, 2000; cf. Sperber & Wilson, 2002).

References

Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development, 67*, 635-645.

Akhtar, N., & Tomasello, M. (2000). The social nature of words and word learning. In R. M. Golinkoff & K. Hirsh-Pasek (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 115-135). Oxford: OxfordUniversityPress.

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development, 63*, 875-890.

Baldwin, D. A. (1993). Early referential understanding: Young children's ability to recognize referential acts for what they are. *Developmental Psychology, 29*, 832-843.

Baldwin, D. A., & Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development, 72*, 708-717.

Baldwin, D. A., & Moses, L. J. (2001). Links between social understanding and early word learning: Challenges to current accounts. *Social Development, 10*, 309-329.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press.

Baron-Cohen, S., Baldwin, D. A., & Crowson, M. (1997). Do children with autism use the direction of gaze strategy to crack the code of language? *Child Development, 68*, 48-57.

Birch, S. A. J., & Bloom, P. (2002). Preschoolers are sensitive to speakers' knowledge when learning proper names. *Child Development, 73*, 434-444.

Bloom, L. (1998). Language acquisition in its developmental context. In W. Damon (Series Ed.), D. Kuhn & R. S. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2, Cognition, perception, and language* (5th edition, pp. 309-370). New York: Wiley.

Bloom, P. (2000). *How children learn the meanings of words.* Cambridge, MA: MIT Press.

Brown, R. (1957). Linguistic determinism and the part of speech. *Journal of Abnormal and Social Psychology, 55*, 1-5.

Campbell, A. L., & Namy, L. L. (2003) The role of social-referential context in verbal and nonverbal symbol learning. *Child Development, 74*, 549-563.

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen through eighteen month infants differentially imitate intentional and accidental actions. *Infant Behavior and Development, 21*, 315-330.

Clark, E. V. (1990). On the pragmatics of contrast. *Journal of Child Language, 17*, 417-431.

Csibra, G., Gergely, G., Biro, S., Koos, O., & Brockbank, M. (1999). Goal attribution without agency cues: the perception of 'pure reason' in infancy. *Cognition, 72*, 237-267.

Diesendruck, G. (2004). *The principle of conventionality in word learning: Two empirical examinations*. Manuscript under review.

Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental Psychology, 37*, 630-641.

Diesendruck, G., Markson, L., Akhtar, N., & Reudor, A. (2004). Two-year-olds' sensitivity to speakers' intent: An alternative account of Samuelson and Smith. *Developmental Science, 7*, 33-41.

Hall, D. G., & Graham, S. A. (1999). Lexical form class information guides word-to-object mappings in preschoolers. *Child Development, 70*, 78-91.

Hollich, G. J., Hirsh-Pasek, K., & Golinkoff, R. M. (2000). Breaking the language barrier: An Emergentist Coalition Model for the origins of word learning. *Monographs of the Society for Research in Child Development, 65 (3)*, Serial no. 262.

Johnson, S. C., Booth, A., & O'Hearn, K. (2001). Inferring the unseen goals of a non-human agent. *Cognitive Development, 16*, 637-656.

Kuhlmeier, V., Wynn, K., & Bloom, P. (2003). Attribution of dispositional states by 12-month-olds. *Psychological Science, 14*, 402-408.

Markman, E. M. (1989). *Categorization and naming in children*. Cambridge, MA: MIT Press.

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-olds children. *Developmental Psychology, 31*, 838-850.

Moore, C., Angelopoulos, M., & Bennett, P. (1999). Word learning in the context of referential and salience cues. *Developmental Psychology, 35*, 60-68.

Moses, L. J., Baldwin, D. A., Rosicky, J. G., & Tidball, G. (2001). Evidence for referential understanding in the emotions domain at 12 and 18 months. *Child Development, 72*, 718-735.

Namy, L. L., & Waxman, S. R. (1998). Words and gestures: Infants' interpretation of different forms of symbolic reference. *Child Development, 69*, 295-308.

Pepperberg, I.M., & Wilcox, S. E. (2000). Evidence for a form of mutual exclusivity during label acquisition by grey parrots. *Journal of Comparative Psychology, 114*, 219-231.

Preissler, M. A. (2003). *Mutual exclusivity as a word learning constraint in children with autism*. Poster presented at the Biennial Meeting of the Society for Research in Child Development, Tampa, FL.

Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology, 33*, 12-21.

Samuelson, L. K., & Smith, L. B. (1998). Memory and attention make smart word learning: An alternative account of Akhtar, Carpenter, and Tomasello. *Child Development, 69*, 94-104.

Savage-Rumbaugh, E., Murphy, J., Sevcik, R. A., Brakke, K. E., Williams, S. L., & Rumbaugh, D. M. (1993). Language comprehension in ape and child. *Monographs of the Society for Research in Child Development, 58*(3-4, Serial No. 233).

Schafer, G., & Plunkett, K. (1998). Rapid word learning by fifteen-month-olds under tightly controlled conditions. *Child Development, 69*, 309-320.

Smith, L. B., Jones, S. S., & Landau, B. (1996). Naming in young children: A dumb attentional mechanism? *Cognition, 60*, 143-171.

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

Stevens, T., & Karmiloff-Smith, A. (1997). Word learning in a special population: Do individuals with Williams syndrome obey lexical constraints? *Journal of Child Language, 24*, 737-765.

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word-object associations by 14-month-olds infants. *Developmental Psychology, 34*, 1289-1309.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*, 1-34.
Woodward, A. L., & Hoyne, K. L. (1999). Infants' learning about words and sounds in relation to objects. *Child Development, 70*, 65-77.

Discussion

Language learning vs. language acquisition: what theory of mind is necessary?
Anne Reboul
Jun 8, 2004 14:32 UT

I've found Gil Diesendruck's paper extremely interesting and I want both to identify points that I take to be of general agreement not only to her and myself but more generally to all participants in the conference until now, and to ask a general question. The point of general agreement (I take it) is that there is a major distinction between words which enter a human or non-human animal knowledge through some associative process and words which enter a human and non-human animal knowledge through some sort of theory of mind. Let's call the first type of words and the first process respectively learned words and lexicon learning and the second type of words and the second process respectively acquired words and lexicon acquisition. Again, I suppose there is a general agreement to the effect that the process in non-human animals and in specific human cases (autistic children or very young children) is a process of lexicon learning (rather than lexicon acquisition) and that the words thus known are learned words, while the process in most human cases is a process of lexicon acquisition and the words are acquired rather than learned. The main point, aside from terminological verbiage is that there is a difference between the use one may make of words learned relative to the use one may make of words acquired, hence the specificity of, e.g., autistic people's discourse relative to normal people's. Note that this doesn't contradict Jill de Villiers' point about recursion: the kind of theory of mind necessary is relatively elementary and certainly does not entail passing the false belief test. Just a word in passing: the distinction made above could be seen as a support for a gricean account of meaning with learned words as more or less intermediate between natural and non-natural (acquired words) meaning. Now for the question and I must own that it is unfair to ask it of Gil when it could obviously have been adressed to any of the preceeding papers. Obviously this question adresses the kind of theory of mind that is necessary for lexicon acquisition. Granted, it does not include passing the false belief test, and, if one agrees with Jill de Villiers does not presuppose linguistic embedding or recursion (which does not exclude non linguistic types of recursion). But the question is: how is that capacity to be described and or defined? According to Bloom, first words are produced at around nine months old and according to Baron-Cohen, at that early age, only the intentionality detector and the gaze direction detector are available (shared attention is not entirely acquired before 18 months). Let us suppose that the first words are learned rather than acquired and that they somehow take on the same format as acquired words later on. The interesting point is the exact characterization of both the type of theory of mind available at, let's say, 18 months of age: is it merely shared attention mechanism or does it go further? What type of other mechanims develop during the following months and years, at what age, how do they interact with previous mechanisms

and what role do they ultimately play in "full" theory of mind, when children pass the false belief test and later on when they pass the opaque context tests?

I should presumably point out that this question is not only adressed to Jill but to all contributors up to now, myself included… So maybe everyone should feel free to answer it!

Three fronts in which "knowledge" develops
Gil Diesendruck
Jun 9, 2004 15:35 UT

Anne Reboul asks two very important questions, which I only touched upon in my paper: what kind of theory of mind is in place at around 18-months, and what develops from that point on? As I note in my paper, it seems to me that the basic competence that needs to be in place is an understanding of human actions as intentional. But in fact, the understanding needs to be more specific: the child has to be able to distinguish between different KINDS of intentions. Specifically, the difference between communicative intents and other active intents, and also among communicative intents themselves (e.g., referring vs. requesting). As far as I can tell, the literature on this topic hasn't yet examined this issue systematically, concentrating instead on when children develop an understanding of intentions generally defined.

As for the second question, of what develops further, I see at least 3 crucial fronts which contribute substantially to word learning as children grow. First, one crucial aspect of theory of mind itself that develops is the child's capacity to assess other people's states of knowledge. Most of the evidence we have thus far shows that by 24-months, there is some sensitivity to what others know, a sensitivity that influences how children interpret people's referential intents. This capacity is crucial for the understanding that words are conventional; i.e., that words are known by all members of a linguistic community. A second system of knowledge that develops tremendously after 18-months is syntactic knowledge. As the child becomes able to discriminate among the syntactic frames in which words are embedded, he/she can use this knowledge to figure out the meanings of words. Last but not least, as the child's conceptual understanding expands, so do the possible word meanings he/she entertains. Evidently, the more information the child has to work with, the more complex the interactions among these different sources.

Just ToM, or ToM with a pragmatic sub-module?
Dan Sperber
Jun 8, 2004 18:36 UT

Gil concludes his outstanding contribution with the following sentence: "*The fact that all the processes involved in the mappings develop independently with unique functional properties, supports the hypothesis that there need not be a special/modular mechanism devoted only for the intentional processing of words.*" This is compatible with the job being done by general ToM, or by general ToM together with a mechanism specialised for inferential comprehension, whether verbal or non-verbal. Assume that, as Deirdre Wilson and I have argued, there is such an pragmatic mechanism. In a sense, it is a sub-mechanism or sub-module of ToM, since its job is to infer a special type of mental state, viz. speaker's meanings. On the other hand, it uses specific inferential procedures (this is what makes it a distinct module). In particular -- or so we claim -- it allows the addresse of an act of communication to home in on the interpretation intended by the communicator on the assumption that this interpretation should meet the addressee's expectations of relevance, without, in the general case, using assumptions about the mental states of the communicator as premises in this inference (unlike Gricean-type interpretation). Two questions then:

- The fact that people with autism find it hard or impossible to use assumptions about the mental states of a speaker, or just direction of gaze, *as premises* in comprehension does not suffice to show that their comprehension does not result in the attribution of a mental state (a speaker's meaning) *as a conclusion*. Couldn't their non-standard word learning still be a pragmatic process carried out by a pragmatic module impoverished in its inputs, rather than something similar to non-human animal word-learning?

- Couldn't "mutual exclusivity" be a by-product of relevance-guided interpretation rather than a *bona fide* constraint, and couldn't, at least in some cases, the participants in Gil's experiments arrive at their responses on a relevance basis, without engaging in Gricean reasoning about what the speaker might have meant?

What accounts for developmental change?
Gil Diesendruck
Jun 9, 2004 16:11 UT

In his commentary, Dan Sperber writes that a pragmatic sub-module "allows the addresse of an act of communication to home in on the interpretation intended by the communicator on the assumption that this interpretation should meet the addressee's expectations of relevance, without, in the general case, using assumptions about the mental states of the communicator as premises in this inference".

My questions about this account have to do primarily with the last part of this claim.

Moore et al. report a study in which children are shown two objects: an active/attractive toy and a plain object. In the crucial condition, the experimenter utters a novel name while looking at the plain object. Moore et al. found that while 24-month-olds associate the name with the plain-but-attended-to object, 18-month-olds associate the name with the attractive-but-not-attended-to toy. On Sperber's account, how can we explain this developmental change? Is it the case that the criterion for cognitive effect changed? Or do developments in executive control make the inhibitory effort more manageable? Or what changes is children's understanding of the communication game -- i.e., you should pay attention to what people are looking at no matter what? Evidently I don't have an answer to these possibilities, but my worry with putting the weight of the interpretation on the mind of an 18-month-old addressee is that it seems it would lead to a substantial number of misinterpretations (a la Moore et al) which would then need to be corrected somehow.

My second concern has to do with the qualification that the relevance-based inference works "in the general case". How do children know what IS the general case? How do they get to it? How do children know when the general case is violated? What do they DO when it is violated? Do they switch to the general ToM module? One finding that seems to me to suggest that the "general" case and the "exceptional" case might be treated through the same infential process comes from our studies on the mutual exclusivity phenomenon. Both in Diesendruck & Markson (2001) and Diesendruck (2004) studies, we conducted separate analyses on children's pattern of responses on the very first trial -- i.e., before they got experience in our "communicative contexts". In all studies, we found that children's pattern of responses in the first trial was the same as their average pattern across trials.

I'll conclude by agreeing with Dan that the interpretation of communication as an intentional act is by no means trivial -- a point I make in my paper and in my response to Reboul's commentary. Nonetheless, I'm unconvinced that this fact alone should lead us to postulate inferential capacities a priori set to process only this kind of input.

Reply to Gil
Dan Sperber
Jun 15, 2004 22:13 UT

Here is a possible relevance-theoretic account of the shift from 18 months where the child takes the novel word to refer to the attractive-but-not-attended-to toy to the 24 months where the child takes the novel word to refer to the plain-but-attended-to object: what develops is the sophistication of the expectations of relevance. The 18-month-old expects utterances to be relevant to her own train of thought and perceptions (and competent caretakers manage to communicate successfully with them by adapting to this expectation, and to more sophisticated expectations at later ages). The 24-month-old expects the utterance to be relevant to her train of thought and perceptions as modified by the interaction with the speaker. At a later age (let me guess, around four) the child expects that the utterance to have seemed to the speaker to be relevant to her, expect by default the speaker to have succeeded in being relevant, but is capable of interpreting the utterance even if, in fact, the speaker was mistaken about what would be relevant to her. At a still later stage, the hearer expects the utterance to have seemed to the speaker to be such that it would seem relevant to the hearer (taking into account the possibility of deception). So what gets more and more sophisticated is the exact content of expectation of relevance, but the default remains actual relevance to the hearer, and the relevance-guided process of inferential comprehension (follow a path of least effort in constructing an interpretation of the utterance and stop when your expectations of relevance have been met) remains the same.

How does the older child (or the adult) know that the default expectation of relevance is not the appropriate one, asks Gil? There are two ways: either the hearer knows that the speaker has a

mistaken idea of what is relevant to her (for instance, the hearer knows that the speaker believes he has some relevant novel information that she in fact already possesses); or the interpretation based on the default assumption contradicts what the hearer knows otherwise of the speaker (for instance the speaker says “Jack is coming”, and I first assume he means Jack Smith, which is more relevant to me, but this interpretation is incompatible with my knowledge of the fact that the speaker does not know Jack Smith). General ToM considerations may play a role in my rejecting an interpretation, but they don’t in my constructing either a default one or an alternative one. ToM-based interpretation is found only in devious cases like Grice’s example of “*a testimonial about a pupil who is candidate for philosophy job” [reading]: “Dear Sir, Mr. X’s command of English is excellent and his attendance at tutorials has been regular, yours, etc””*, where the interpreter has to think about the strategy of the communicator. Failing to understand the implicature of such a message is not a failure of verbal comprehension but of strategic thinking.

On expectations of relevance increasing in sophistication in development, see Sperber (1994) Understanding verbal understanding. In Jean Khalfa (ed.) *What is Intelligence?* Cambridge University Press, 179-198. On reasons to assume that comprehension involves a sub-module rather than general ToM, see Sperber & Wilson (2002) Pragmatics, Modularity and Mind-reading. In *Mind and Language.* 17. 3-23. (Both texts are available at www.dan.sperber.com).

Reply to Dan's reply
Gil Diesendruck
Jun 17, 2004 9:36 UT

Dan says that "what gets more and more sophisticated is the exact content of expectation of relevance." Specifically, in his examples of how this expectation might change, he talks about changes in the listener's assumptions about what the speaker knows, believes, intends, and the sheer capacity to take into consideration not only his (the listener's) own thoughts, but also the speaker's.

Evidently, these all seem like ToM related developments. And if so, then would you agree, Dan, with the following formulation of the interaction between Relevance and ToM? Relevance defines the parameters (e.g., effort, effect) by which the inferential solution to the implicature(in this case word meaning) is made, but ToM determines the developmental and perhaps context specific nature/value of these parameters.

We still disagree
Dan Sperber
Jun 18, 2004 22:55 UT

Gil writes :” *would you agree, Dan, with the following formulation of the interaction between Relevance and ToM? Relevance defines the parameters (e.g., effort, effect) by which the inferential solution to the implicature (in this case word meaning) is made, but ToM determines the developmental and perhaps context specific nature/value of these parameters.*” I am afraid not quite. My claim is that inferential comprehension (which, by the way, is not just about implicatures but also about the explicit context which is always underdetermined by linguistic meaning) is a distinct sub-module of ToM. So yes, there are interactions between ToM and its comprehension sub-module, but no, the nature and value of the effort and effect parameters (which have to do with the role of relevance in cognition in general, and not just in communication) are not at all determined by ToM. On the other, yes, ToM plays a role in addressee’s expectations of relevance becoming more and more sophisticated.

Two kinds of
Andrei Popescu-Belis
Jun 12, 2004 11:03 UT

Thank you, Gil, for your stimulating paper on the relations between word learning and the development of a theory of mind (ToM). I would like to outline a paradox that may arise from a too literal reading of your paper, especially from someone not fully familiar with the field (e.g., myself).

While you acknowledge that some kind of words can be learned without a theory of mind, you dismiss them are mere associations that are useless for full-fledged communication. At the end of your paper, you suggest that "these words ... would have very little similarity to what we standardly consider as words". One could believe in this case that you would accept as real words only those that are learned thanks to a ToM, or

those that are used in relation to a ToM. So the statement that 'real word learning requires a ToM' would become in this reading a mere tautology.

The distinction of two types of words, or more accurately of two types of learning, is nicely framed by Anne Reboul in her comment (I am not sure though whether the terms 'learning' vs. 'acquisition' are the most adequate choice here). My feeling is that the contrasting of, say, A-words learned without a ToM, vs. B-words learned with a ToM, could prove quite unproductive for the understanding of both language and ToM evolution. On the contrary, I would be interested in commonalities between the two types of words, in the features of words and word learning that generalize the best, and that could provide the missing steps on the path of language evolution.

It seems to me that the gap between A-words and B-words may be narrower that it appears from your paper. For a start, I would avoid focusing on the distinct uses of A-words vs. B-words as an argument for their difference. Obviously, an individual without a ToM (e.g. autistic people) cannot be supposed to use words for inferential communication, but this does not rule out the possibility that the individual makes use some kind of communicative labeling device that shares a significant amount of common properties with... words. Would one deny that apes have hands, on the ground that they never use them for writing?

The gap between A-learning and B-learning may have become even narrower since yesterday. In the latest issue of Science, Kaminski et al. report the case of Rico, a dog that acquired about 200 words, and bring interesting evidence for exclusion learning and "fast mapping" capacities. For instance, Rico was able to infer (?!) the referent of a new word among a group of eight objects, seven of which already had names that he knew. Of course, as Paul Bloom notes in his comments (same issue), the use of these names by Rico is quite limited. One can definitely not expect it to use them in a creative way, but in the experiments, such words are generally interpreted exclusively as commands to fetch the respective objects. It seems these dog-words behave more like common names, rather than proper names, though they lack the conceptual autonomy (or referentiality) of human words. But there are striking similarities between Rico's learning and human learning of words.

Furthermore, how would the distinction between A-learning and B-learning operate in adult word learning? Would second language learning by adults be entirely based on B-learning, on the grounds that a full ToM is available? I can definitely think of cases when complex inferences allowed me to guess the meaning of a foreign word, but dictionary look-up and syntactic cues remain useful too.

To summarize: couldn't we think of a variable set of strategies for word learning rather than oppose only two different modes? Isn't A-learning an important factor in bootstrapping the ToM itself, a strategy that could remain available even to adults, but could be in some cases superseded by B-learning? This could be seen as a reductionist proposal, which attempts to identify building blocks in the evolutionary emergence of language, rather than look for complex pre-requisites to any kind of word learning. One could even think of an evolutionary game, in which successful (i.e. rewarded) word learning would prompt infants to develop their attentional skills - a view which nearly reverses the priorities of learning words and developing a ToM.

References

J. Kaminski, J. Call, J. Fischer (2004) - Word Learning in a Domestic Dog: Evidence for "Fast Mapping". Science, 304, 11 June 2004, p.1682-3.
P. Bloom (2004) - Can a Dog Learn a Word? Ibid., p.1605-6.

Completion

Andrei Popescu-Belis
Jun 12, 2004 11:06 UT

The full title is:
Two kinds of "words"? For some reason everything after the first quote was skipped. Sorry for this incident. APB.

Two kinds of processes

Gil Diesendruck
Jun 14, 2004 11:51 UT

Andrei raised a number of very interesting questions regarding my paper. I'll respond here to some of them. Starting from his final question, as to whether children employ a variety of strategies to learn

words, let me point out that this position is indeed the dominant one in the field nowadays. Note, however, that this position doesn't solve the problem of how children then coordinate among these multiple strategies. The position I was defending is that once children understand what communication is all about (i.e., intents), they can then use this understanding to define the relative importance of the various strategies.

Andrei's first point is also important: if we define "real" words in terms of intents, then isn't it tautological to say that children acquire words via ToM? Being aware of this "catch", I defined word learning in my paper not in terms of intents but rather in terms of dispositions. What I mean is that, to understand a word, a listener must figure out what is the particular relation between the speaker and the referent of the word. THe listener must realize what the speaker wants me to DO with the referent. Does the speaker want me to look at it? to bring it? to catch it? This is the sense of disposition I had in mind. Clearly, in our Western adult lexicon, "intents" capture this meaning quite straightforwardly, but it's nonetheless an interpretation of what listeners should do in word learning. What they ACTUALLY do, is respond/behave appropriately. ToM is one possible mechanism that can explain why listeners behave appropriately. If other mechanisms can account as effectively for these word learning behaviors, then ToM might not be necessary.

Which brings me to Rico -- the formidable 9-year-old German border collie capable of understanding 200 words and fast mapping by exclusion. As Andrei pointed out, in his commentary to the Science paper, Paul Bloom gave a number of reasons why to be cautious in interpreting Rico's capacities. I want to add two rather specific cautions.

First, a purely methodological matter: if I understood the procedure correctly, in the fast mapping trials Rico was presented 7 familiar objects and 1 completely novel object. On the first trial, Rico's trainer asked him to fetch a familiar object. Then either on the 2nd or 3d trial, the trainer asked him to fetch a [novel word]. On 70% of these novel word trials, Rico indeed brought back the 1 novel object. My question is simply this: could this response be due to a novelty preference? That is, is it possible that Rico simply "preferred" the novel object -- because it was new! Here's a simple procedure to test this possibility: familiarize Rico with the novel objects also (without labeling them); then give him the fast mapping trials. If I'm correct, the preference for the novel object would by then be extinguished, and Rico might not bring any object in response to a novel word.

I have to say that watching the video (at Science's website) of Rico performing the task reinforced my suspicion. What you see in the video is that from the very first moment Rico enters the room, the novel object (a pink rabbit) attracts his attention, even though he was asked to fetch a familiar object.

A second issue that is also somewhat apparent in the video is how Rico picks out the novel object: he goes straight to it. When children are asked to retrieve the referent of a novel word, what they often do is search around the room! It's as if they realize that there must be something there that responds to the novel name the speaker has used, and perhaps if they see the thing, they will recognize it. It's only after they have searched -- and not found -- that they decide that perhaps the best candidate for that novel word is the novel object laying right in front of them.

The bottom line is a general one: many processes can lead to similar behaviors; and in order to decide between processes, one needs to analyze the behaviors into its components and implications, breaking down their apparent similarity.

Reply to Gil: Starting simple
Andrei Popescu-Belis
Jun 16, 2004 22:31 UT

I was very interested in the possible explanation provided by Gil Diesendruck for Rico's behavior, namely an interest in novelty – probably a quite universal behavioral feature – and I think this is a serious objection to the thesis contained in the title of the Science paper (Word Learning in a Domestic Dog: Evidence for "Fast Mapping"). It would probably take more than two pages to provide all the necessary experimental evidence that “fast mapping” was responsible for Rico’s behavior. Still, in Gil’s explanation, the uttered novel word must still be recognized as novel by Rico in order to trigger the “fetch new object” behavior – and this word varied in the trials.

With respect to the first part of the answer, I wonder whether a modular view of ToM would not render more acceptable some of the consequences of a ToM-based lexical learning. The view that “The listener must realize what the speaker wants me to DO with the referent. Does the speaker want me to look at it? to bring it? to catch it?”, could be limited to the simplest version of DOing that was mentioned, namely looking, which requires only a simple component of a ToM, such as shared attention (see also my reply to Anne Reboul above). The coordination between various ToM-based strategies for lexical learning could then be just a matter of availability of the respective modules, depending on the infant’s age. Simpler strategies are the first ones available, and they are complemented later by more complex ones, especially in situations where these are required for disambiguation.

Reply to Andrei
Gil Diesendruck
Jun 17, 2004 11:06 UT

First, a couple of points about Rico. Indeed, Josep Call just informed me that Rico was never familiarized with the novel objects, so in a way my explanation stands: Rico's tendency to pick the novel object in response to a novel word could have resulted from his preference for the unfamiliar object. The second point is that all Rico had to recognize from that request is that the trainer was not using a familiar word. That is, Rico didn't have to recognize the novel word as a NOVEL WORD, but simply as something different from the familiar names of the objects. That is, if Rico could understand the phrase, "Get me SOMETHING from the other room", my prediction is that he would respond in the same way as in the novel word condition.

Andrei suggests that perhaps a modular view of ToM (e.g., Baron-Cohen's account) might be best suited to explain the developmental findings regarding word learning. That might be true. I, however, am not yet willing to commit to a strong version of an evolutionary based and neuroanatomically instantiated ToM modules.

Declarative pointing and ToM
Ingar Brinck
Jun 15, 2004 10:45 UT

Gil Diesenbruck uses a wide concept of ToM. But it seems that some kind of rather low-level ToM is necessary to understand attention as directed, and thus a non-symbolic referential intention, while a more advanced ToM is required to understand words that relate to e.g. folk psychology. Unless we distinguish between kinds of ToM, we will not be able to undersdtand at what stage a full-blown ToM becomes active in language acquisition. It will as well be difficult to account for variations in language acquisition, due to e.g. age, autism and other kinds of illnesses that affect language acquisition. Moreover, to account for the difference between humans and other primates as regards communication skills and symbol acquisition, we also need a more refined view. For instance, to say that apes use signals, while humans use symbols, will not further our understanding much, especially not if our aim is to understand the evolution of language. Declarative pointing (DP) is a prime example of the child’s use of a referential intention to direct the attention of the other to a shared object. But DP does not require much of a ToM, not intentions about intentions, but only the capacity of having attentional states about intentional ones. DP is intimately related to word learning, and in the normal case occurs by 12 months. It is typical for the human species. It is referential and has an basic indicating function that can be put to various uses in different contexts. A child who masters DP has grasped of the distinction between a general communicative intention and the intention that the act of pointing is used with in particular contexts. Thus declarative pointing is in several ways comparable with indexical words. Yet it is acquired at an age when the child has not yet developed much of a ToM. Being able to use DP without being dependent on support from the environment relies on having a certain kind of joint attention. This will occur if speaker and hearer: (i) attend to each other’s states of attention as states that are directed at objects in the environment and by being so directed are about the objects, and, moreover, attend to these states also when they are directed at non-salient entities. (ii) make attention contact, and (iii) alternate gaze between each other and the object. Elsewhere I have argued that such joint attention is intersubjective in two senses: it comprises interattentionality and interaffectivity. Thus it takes into account those states that display the intentional mental states of the other, but not those states themselves. This indirect understanding or sensitivity to mental states can in quite a few cases take on a similar role as does full-blown ToM. To conclude: Word learning without ToM is crucial to get language acquisition going (disregarding the case of autism), while a ToM is required to become a fully competent speaker.

Brinck, I. 2004, “The pragmatics of imperative and declarative pointing”, Cognitive Science Quarterly, 3(4), 1-18. Brinck, I. (to appear), “Joint attention, triangulation and radical interpretation: A problem and its solution”, Dialectica, 1-20.

DP isn't necessary nor sufficient for word learning
Gil Diesendruck
Jun 17, 2004 9:50 UT

I totally agree with Brinck's argument that a basic ToM-based capacity might be enough to get word learning off the ground. And I'm sorry if that point wasn't clear in my original paper or other replies. I also fully agree that we don't yet know how to best characterize this "basic" ToM capacity. I disagree with Brinck, however, on how important DP is for word learning.

First, a principled disagreement: while DP seems like an indication that the child distinguishes between general communicative intention and other intentions -- as noted by Brinck -- in order to learn words children need to make a finer distinction -- as I pointed out in one of my earlier replies. For one, the child needs to distinguish between referential intents and other sorts of communicative intents. In itself, DP doesn't indicate that a child can make these distinctions. In other words, DP is probably not sufficient for word learning.

Second, empirical evidence is beginning to accumulate that DP or joint attention more broadly might not be necessary for word learning. Specifically, Nameera Akhtar and colleagues have shown that young word learners (2-year-olds) can learn words simply by "overhearing" others label objects. This type of finding helps us understand also why children raised in cultures where "motherese" is not so common, nonetheless acquire words quite effectively.

Dogs, words and ToM
Anne Reboul
Jun 16, 2004 7:13 UT

I've followed with interest the exchange between Andrei Popescu-Belis and Gil Diesendruck. I would merely like to point out that in a previous paper in Science (ref. below), Hare and colleagues tested a (seemignly) low-level communicative ability, the ability to follow and understand a pointing gesture. This ability is known to be absent in non-human primates, but they wanted to test it on the paramount domestic animal, i.e. the dog and, what is more to examine whether it might be at least partly innate. The experiments compared the ability of woolves — raised in a human environment, unless my memory fails me —, domestic dogs and dog puppies on their ability to use a pointing gesture by a human experimenter as a cue for food retrieving (they'd been very careful to mask all smells). The results were perfectly clear: woolves performed at chance, while both dogs and puppies performed well above chance levels. The conclusion seem to be (unsurprisingly) that domestic dogs, presumably through the human-guided process of evolution which has lasted since dogs were first domesticated (seval thousands years) have been selected for their social abilities, including their communicative abilities. Thus, though I agree with Gil and Bloom that the results of the Rico experiment should be taken cautiously, they may nevertheless, in the light of Hare et al.'s experiment, not support a ToM-less view of words acquisition. Just a word of caution: I certainly do not mean that dogs have a full ToM or would be able to pass the false belief test or are able of deception or anything like that: I'm merely pointing out that dogs are likely animals to have some communicative abilities, however primitive. This is hardly surprising: they were before domestication social animals and whatever communicative skills they may have evolved as non-domesticated animals have probably been augmented during the long domestication process.

Dogs, words, and (some) components of ToM
Andrei Popescu-Belis
Jun 16, 2004 22:27 UT

The additional experiments with dogs described by Anne Reboul clearly point at the need to distinguish various degrees or components of the ToM ability, some of which could be present in dogs or apes. Referring to Simon Baron-Cohen’s model (1995), the "mindreading system" could be at least composed of an intentionality detector (ID), an eye-direction detector (EED), a shared-attention mechanism (SAM), and a theory-of-mind mechanism (ToMM). Note that with this terminology, ToMM per se is only one of the four mechanisms, albeit the highest-level one.

So, to explain why dogs "understand pointing gestures", one could reasonably argue that an EED mechanism emerged or was enhanced by human-guided selection of man's best friend. It is also clear that EED is of great help in human lexical acquisition, though following Baron-Cohen's modular view, EED-based learning is certainly not yet ToMM-based learning. One could argue that some lexical acquisition can occur even without EED, based on word-to-referent pairing, provided referents have enough perceptual salience to the infant (this could occur for the names of the parents for instance).

Computational evolutionary games illustrate nicely the various possibilities of lexical "learning". Lexical diffusion through word-to-object pairing is a straightforward phenomenon, quite easy to model computationally when new objects are presented one by one to agents, together with their "name" (quotes should probably appear everywhere in this paragraph). However, lexical learning still occurs when several new objects are presented at a time, provided the number of new objects is small with respect to the number of known objects and of presentations. These simulations seem to show that associative learning of "words" could work even when based on perceptual salience only, but it is greatly enhanced (especially when the referent/background distinction is hard) by the availability of a pointing mechanism (Popescu-Belis 2002).

Which is of course nicely illustrated by the literal interpretation of the well-known Chinese saying: "When the wise man points at the moon, the fool looks at the finger."

Baron-Cohen S. (1995) – Mindblindness. MIT Press, Cambridge, MA.

Popescu-Belis A. (2002) - Modèles dynamiques de l'émergence de conventions linguistiques. In Guillot A. & Daucé E., eds. (2002), Approche dynamique de la cognition artificielle, coll. "Traité des sciences cognitives", Hermès, Paris, p.181-199.

On behaviors and mental states

Gil Diesendruck
Jun 17, 2004 11:27 UT

Anne's and Andrei's comments on this issue highlight the tricky business that is psychology. As behaviorists like to remind us cognitivists, in the end of the day we are all behaviorists in the sense that the only data we have IS behavior. So when does acting based on pointing involve intentional understanding and when doesn't it?

This is tricky, but there might be nonetheless several ways to make this distinction, and I want to point one: context specificity. Both studies on dogs required the dogs to perform a similar action: fetching. In fact, in Hare et al's studies it wasn't simple fetching, but the evolutionary crucial task of fetching food. Do dogs follow pointing when no food (or fetching) is involved? Do dogs even "entertain" the possibility that pointing could have an implication different from fetching or finding food (e.g., of referring)? I dont' know the answer to these questions, but recall that young HUMAN infants -- with a nonetheless less matured brain than puppies -- do not follow pointing at all. Once they start doing it, however, they do so across various contexts and understand a variety of communicative intents implied.

Furhter support for the context specificity of "mindreading-like" behavior comes from recent studies on primates. For instance, on the one hand, Povinelli has systematically studied captive primates and concluded that, without extensive training, chimpanzees fail at a number of basic tasks having to do with seeing-knowing relations -- tasks that 2-year-old human children succeed. On the other hand, recent studies coming from Tomasello and Call's lab, and Hauser and Santos's labs, show that in more naturalistic settings, and when performing an evolutionary important task (e.g., retrieving food), chimpanzees and new-world monkeys SUCCEED in tasks very similar to those used by Povinelli.

So, do these animals have mindreading capacities? Or is it possible that they are reacting to specific behaviors or visual configurations of crucial stimuli (e.g., is my competitor looking my way or not)? It seems to me that even the experts don't yet agree.

On concepts and language

Véronique Boulenger (Institut des Sciences Cognitives, Lyon) and Tatjana Nazir (Institut des Sciences Cognitives, Lyon)

(Date of publication: 21 June 2004)

Abstract: One of the central issues in ToM concerns the role of concepts such as belief and knowledge. The goal of the paper is to take a closer look the relation between concepts and language. Today, the claim that evolutionary pressures have resulted in specialized neural circuits dedicated to processing different categories such as 'animals', 'fruit/vegetables', and 'tools' goes almost with no further notice. As even "old world monkeys" such as Baboons can use analogical reasoning to match symbol arrays, the speculated neural circuits in humans may then have evolved prior to language. Yet, in human infants there are clear links between word learning and conceptual organization. We will review category specific semantic deficits as observed in brain damaged patients together with findings in functional neuroimaging to assess whether these results can provide hints as to the evolution of human language.

Categorization: a fundamental principle of cognitive economy.

Many animals can detect basic characteristics of objects or events and store them as generalized classes beyond specific details of sensory inputs. Hence, crickets sharply divide a range of sound frequency into either "mate" (<16 kilohertz) or "bat" (>16 kilohertz) - a predator - even though the tones vary along a continuum. While such categorical perception is known primarily from human speech perception the cricket-example suggests that forming perceptual categories is a rather widespread feature of sensory systems (Wyttenbach et al., 1996). The ability to classify similar but not identical things as equivalent (by reacting to them in the same manner) is, in fact, an optimal solution to minimize processing requirements, which is advantageous especially for small information-processing systems (Cook, Wright & Kendrick, 1990; Delius, Siemann & Jitsumori; 2000). The inclination to compress the amount of information to be retained is observed in many animals although the strategy used to attain this "cognitive economy" (Huber, 2004) varies between (and within) species: Categorization can be done at different levels of abstraction, requiring different degrees of instruction.

The seminal work by Herrnstein & Loveland (1964), for instance, showed that pigeon could be trained to classify photographs on the base of whether or not they contained a human being. This classification skill does not simply apply to the trained pictures but generalized to novel pictures not used in the training phase. By relying on low-level visual features pigeons can even learn to discriminate color slides of paintings by Picasso from those of Monet and generalize this ability to novel paintings by other cubists (e.g., Braque) and impressionists (e.g. Cezanne) (Watanabe et al. 1995). More developed animals such as monkeys, by contrast, can learn to take multiple stimulus dimensions into account in order to make sharp distinctions between categories such as "animals" or "trees". (see Miller et al., 2003). In contrast to pigeons this ability is not simply based on grouping of the stimuli according to physical similarity because monkeys can be trained to categorize computer generated visual stimuli along a morphing continuum of various blends of two different objects (Freedman et al. 2001). Hence, by merging different amounts of a "cat" and a "dog" image, monkeys can be trained to develop discrete category boundary between the two classes of animals even as the similarity between two stimuli of one category (e.g. two different types of cats) is lower than the similarity between stimuli belonging to the other category (e.g., a cat and a cat-like dog). The sharp boundaries between categories is even evident at the neural level as single neurons in the (inferior temporal cortex and the prefrontal) cortex of the trained animals react selectively to a member of a given category with little differentiation between various examples of the category but sharp transition in neural activity between categories (Freedman et al., 2001; Vogels, 1999).

In humans (and possibly in other animals) the evolutionary pressure to minimize processing requirement seem to have lead to the development of specialized and functionally dissociable neural circuits for the processing of objects for which rapid identification could have had survival advantage (Caramazza & Shelton, 1998). Hence, brain damage in humans can lead to impairments at recognizing or identifying one category of objects more than another. For instance, while some patients have difficulties in correctly distinguishing cucumbers from green peppers without having problems in identifying animals, other patients do fine with vegetables and fruits but identify a giraffe as kangaroo or claim that the natural color of an elephant is orange. Post traumatic impairments at recognizing or identifying one category of objects more than another have been reported for semantic categories as broad as living vs. non-living things to more narrow categories such as fruits/vegetables, animals, and tools. Functional brain imaging studies with healthy subjects confirm these observations by revealing distinct cortical networks activated by living and

non-living things. Non-living things activating regions in the parietal and frontal lobes of the language dominant left hemisphere, while living things activate a network of cortical regions mainly in the occipital lobes and along the temporo-parietal junctions (see Figure 1).

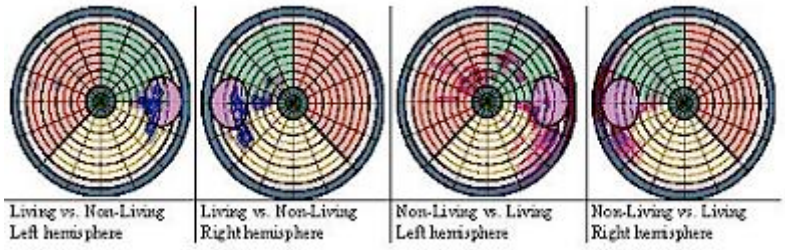

Figure 1: A 2D map of the two hemispheres (Geometric atlas; Toro & Burnod, 03) with the main lobes indicated by different colors (pale red= frontal lobe; pale green= parietal lobe; pale yellow = temporal lobe; grey = occipital lobe. The central circle represents the insula). The blue and violet shaded zones in the circles specify cortical activity observed across a series of experiments with healthy subjects while they were observing pictures of living things or non-living things, respectively.

A priori, the development of segregated neural circuits for specific categories could emerge entirely as the result of experience and learning (see the monkey-example above). Accordingly, artificial neural networks can develop separate neural representations for arbitrary categories provided that they are adequately trained (Polk & Farah, 1995). In humans, however, the organization of conceptual knowledge about living and non-living things seems to be specified in the genome, as lifelong selective deficits for living things (with normal knowledge about non-living things) can be observed after brain injuries that occurred few days post natal. That is, prior to any experience the cortical substrate for acquiring semantic memory for nonliving-things can develop in the absence of the development of function of semantic memory for living-things (Farah & Rabinowitz, 03).

In conclusion, a strong common bias to reduce processing loads (through the search of commonalities between objects) may have led to the emergence of categorization skills in different species. The specific solution to achieve this goal, however, varies between kinds. It is yet unknown when categorization appeared in the phylogenetic scale and whether it appeared only once. However, as such, categorization does not require language skills. Does language rely on categorization (Herrnstein, 1984)?

Categorization in preverbal infants

Studies with preverbal children have shown that as soon as infants are able to parse words from the speech stream, words have a powerful influence on conceptual organization (Waxman, 2004). Using a novelty-preference task Waxman and colleagues have shown that children as young as 9 months old form category-based commonalities simply by establishing links between novel words and objects. In the novelty-preference task children are familiarized to a series of objects from a given category (e.g. animals), and are then tested with two new objects of which one is a member of the familiarized category (e.g., another animal) and the other a member of a new category (e.g. a fruit). If infants notice the category-based commonalities among the familiarization objects, in the test phase they should show a preference for the fruit. During the familiarization phase the presentation of the member of the given category is either accompanied by a neutral sentence such as: “look at this”, or a sentence that introduced a novel word: “this is a *blicket*” (Waxman and Markow, 1995; 1998). While children in the neutral condition show no preference for one or the other object in a subsequent test phase, infants in the novel-word condition show a clear preference for the object that belongs to the new category (i.e. the fruit). The presence of words seems even to allow infants to generalize beyond shared shape similarity to nonvisible properties of objects (Graham et al., 2004). According to Waxman and colleagues, these experiments show that infants begin the task of word learning equipped with an (universally) available expectation linking novel words to a range of commonalities between objects. In the authors’ terms, “words serve as invitation to form categories” (Waxman and Markow, 1995). Slightly turning the argument we could therefore say that as infants are strongly predisposed to seek for optimal solutions to minimize processing requirements, words provide a perfect clue allowing them to *rapidly* form categories.

Although nothing in the here described issue allows speculations about the emergence of language, an innate drive to minimize information processing could at least foster its evolution.

Why words? A series of speculative hypotheses

Interestingly, the successful categorization of objects by young human infants' as described above hinges on the use of *verbal* stimuli and does not occur for equally attention-engaging non-verbal auditory stimuli (Balaban & Waxman, 1997; 2002). That is, although both verbal and non-verbal auditory stimuli capture infants' attention, only words support the establishment of categories. And this, even before children understand their meaning. What is so special about words that they allow the quick establishment of categories?

Though highly speculative, a possible reason for the advantage of verbal over non-verbal sounds in the formation of categories could lie in the role of (vocal) imitation/simulation. Human infants are predisposed to produce verbal sounds and by this the status of a spoken word might be different from the status of a non-verbal (non-reproducible) sound. In fact, studies with human adults have demonstrated that observing and executing a motor action activate partly overlapping cortical regions in several areas of the brain (Iacoboni et al., 1999). This observation has been taken to assume that observing actions (that can be executed by the observer) involves a "resonance" mechanism that allows to directly mapping a perceived action onto an internal motor representation of the same action. This internal simulation, in turn, is the base of imitation. Coherent with this assumption, distinct patterns of neural activity are observed when adult humans are asked to report whether or not people (imitable) or dogs (not imitable) can perform a common set of behaviors. While people and dogs are capable of performing many of the same actions (e.g. run, sit, bite), the representation of this knowledge seem to be associated with distinct patterns of neural activity (Mason et al., 2004).

The "resonance" or "mirror neuron" system has initially been observed through analysis of the activity of single neurons in rhesus monkeys (Rizzolatti et al., 1999). Yet, while many species can copy the models' choice of object (goal of the action), imitation of the perceived motor action of a model is rare in non-human animals (Premack, 2004). In human infants, by contrast, automatic imitation of elementary actions such as facial and manual gestures is already present at only few days of age (e.g., Meltzoff & Moore, 1977). Hence, if this "resonance" mechanism truly exists, it has gained sophistication in humans compared to monkeys and might provide an important step towards the understanding of human evolution in general (Ramachandran, 01).

By way of such a system verbal stimuli could thus gain their special status (independently of the fact that it is language!) over the equally attention-engaging but non-"resonating", non-verbal auditory stimuli in infants' categorization. Moreover, at the same time as linking words to commonalities between objects satisfy the hypothesized bias to reduce processing loads (by forming categories), it also provides infants with a mean to establish a rudimentary lexicon that can be fine-tuned subsequently. Finally, although categorization by help of words may initially rely on physical aspect of the stimuli (e.g. visual similarity between objects), once the infant understood the principle, words could help to form abstract categories beyond simple grouping-by-similarity and may eventually lead to the development of concepts.

Lastly, the rapid formation of categories by help of verbal stimuli in young infants requires that someone tell the infant: "this is an X". In other words, there is a need for a "teacher". Like imitation, teaching seems to be strictly human and does not occur in other animals, not even in chimpanzees (Inoue-Nakamura et al., 1997). Teaching reverses the flow of information found in imitation by providing feedback to the imitator (Premack, 04) and therefore requires some degree of intentional communication. Teaching entails the ability to represent other minds. In the frame of the "resonance" mechanism hypothesis, miming and teaching rely on much the same principle. Anytime you watch someone else doing something (or even starting to do something as the system is also activated by still pictures that only imply motion, Nishitani and Hari, 2002), the resonance mechanism allows the "reading" and understanding of other's intentions, and thus to develop a sophisticated "theory of minds."

References

Balaban MT, Waxman SR. Do words facilitate object categorization in 9-month-old infants? J Exp Child Psychol 1997; 64: 3-26.
Caramazza A, Shelton JR. Domain-specific knowledge systems in the brain the animate-inanimate distinction. J Cogn Neurosci 1998; 10: 1-34.
Cook, R. G., Wright, A. A. & Kendrick, D. F. (1990). Visual categorization by pigeons. In M. L. Commons, R. J. Herrnstein, S. Kosslyn & D. Mumford (Eds.), Quantitative analysis of behavio: Vol. 8. Behavioral approaches to pattern recognition and concept formation (pp. 187-214). Hillsdale, NJ: Lawrence Erlbaum.
Delius, J. D., Jitsumori, M., & Siemann, M. (2000). Stimulus equivalencies through discrimination reversals.

In C. Heyes & L. Huber (Eds.), The evolution of cognition (pp. 103-122). Cambridge, MA: MIT Press.
Farah, M Rabinowitz, (2003). Genetic and environmental influence on the organization of semantic memory in the brain: Is "living things" an innate category? Cognitive neuropsychology, 20, 401-408
Freedman DJ, Riesenhuber M, Poggio T, Miller EK. Categorical representation of visual stimuli in the primate prefrontal cortex. Science. 2001 Jan 12;291(5502):312-6.

Graham SA, Kilbreath CS, Welder AN. Thirteen-month-olds rely on shared labels and shape similarity for inductive inferences. Child Dev. 2004 Mar-Apr;75(2):409-27.
Herrnstein, R. J. (1984). Objects, categories, and discriminative stimuli. In H. L. Roitblat, T. G. Bever, & H. S. Terrace (Eds.), Animal cognition (pp. 233-261). Hillsdale, NJ: Lawrence Erlbaum Associates.
Herrnstein RJ, Loveland DH. Complex Visual Concept in the Pigeon. Science 1964; 146: 549-51.
Huber, L. (2004) Visual Categorization in Pigeons. In Robert G. Cook (Ed). Avian visual cognition. Comparative Cognition Press.
Iacoboni M, Woods RP, Brass M, Bekkering H, Mazziotta JC, Rizzolatti G. Cortical mechanisms of human imitation. Science. 1999 Dec 24;286(5449):2526-8.
Inoue-Nakamura N, Matsuzawa T. Development of stone tool use by wild chimpanzees (Pan troglodytes). J Comp Psychol 1997; 111: 159-73.
Mason MF, Banfield JF, Macrae CN. Thinking about actions: the neural substrates of person knowledge. Cereb Cortex. 2004 Feb;14(2):209-14.
Meltzoff AN, Moore MK. Imitation of facial and manual gestures by human neonates. Science 1977; 198: 74-8.
Miller EK, Nieder A, Freedman DJ, Wallis JD. Neural correlates of categories and concepts. Curr Opin Neurobiol. 2003 Apr;13(2):198-203.
Nishitani N, Hari R. Viewing lip forms: cortical dynamics. Neuron. 2002 Dec 19;36(6):1211-20.
Polk TA, Farah MJ. Brain localization for arbitrary stimulus categories: a simple account based on Hebbian learning. Proc Natl Acad Sci U S A 1995; 92: 12370-3.
Premack D. Psychology. Is language the key to human intelligence? Science 2004; 303: 318-20.
Rizzolatti G, Fadiga L, Fogassi L, Gallese V. Resonance behaviors and mirror neurons.
Arch Ital Biol. 1999 May;137(2-3):85-100.
Toro R, Burnod Y. Geometric atlas: modeling the cortex as an organized surface. Neuroimage 2003; 20: 1468-84.
Vogels R. Categorization of complex visual images by rhesus monkeys. Part 2: single-cell study. Eur J Neurosci. 1999 Apr;11(4):1239-55
Watanabe, S., Sakamoto, J., & Wakita, M. (1995). Pigeons' discrimination of paintings by Monet and Picasso. Journal of the Experimental Analysis of Behavior, 63, 165-174.
Waxman, Sandra R. (in press). Everything had a name, and each name gave birth to a new thought: Links between early word-learning and conceptual organization. In D. G. Hall & S. R. Waxman (Eds.), From many strands: Weaving a lexicon. Cambridge: MIT Press.
Waxman SR, Markow DB. Words as invitations to form categories: evidence from 12- to 13-month-old infants. Cognit Psychol 1995; 29: 257-302.
Waxman SR, Markow DB. Object properties and object kind: twenty-one-month-old infants' extension of novel adjectives. Child Dev 1998; 69: 1313-29.
Wyttenbach RA, May ML, Hoy RR. Categorical perception of sound frequency by crickets. Science 1996; 273: 1542-4.

Discussion

On resonance, words and categories

Gloria Origgi
Jun 21, 2004 18:09 UT

It is unclear to me how children's mirror system could help to understand the role of words in category formation. Nazir and Boulenger say: *"By way of such a system verbal stimuli could thus gain their special status (independently of the fact that it is language!) over the equally attention-engaging but non-"resonating", non-verbal auditory stimuli in infants' categorization"*. What if the stimulus is non-verbal but "resonating", as for example a gesture of pointing towards a new item that has been introduced by the experimenter?

More generally, what is the relation, according to the authors, between (1) the referential capacity of recognizing a piece of verbal input as a “name” for an object , (2) the capacity to resonate with a perceived action and (3) the predisposition to reproduce language?

Resonating stimuli: words and gand gestures: a reply to Origgi & Reboul

Tatjana Nazir
Jun 25, 2004 10:39 UT

A non-verbal but “resonating” stimulus has a similar impact on infants’ categorization as novel words do. Hence, non-iconic hand gestures (e.g. repeated simultaneous extension of index and middle finger from a closed fist, etc.) for instance, helps infants’ object categorization as readily as words (Namy & Waxman, 98). Also, young infants use symbolic gestures and words in the same manner (i.e., to request or label) and show little overlap between the two. That is, if a gesture had been acquired for a particular referent they don’t have a word for that referent. Interestingly, however, this apparent equivalence between words and hand gestures hold to a certain age only and disappears in older infants. Hence, while 18-month-olds spontaneously accept both novel words and arbitrary hand gestures as names for object categories, 26-month-olds continue to do so for words but not for hand gestures (except when given additional practice). This developmental change is indicative of an advantage to use words as prime symbolic form (either because of the existence of a specialized cognitive module for language, or, alternatively, because words are simply the better stimulus).

In the same line, hearing babies who receive no systematic exposure to spoken language in early life but are exclusively exposed to sign language, “babble silently on their hands” (Petitto et al., 01; 04). That is, while they show clear differences in the rhythmic frequency of their hand activity when compared to speech-exposed hearing babies, they do not vocally babble like speech-exposed babies do. Babbling is a key mechanism that allows babies to discover the patterned structure of natural language (e.g., de Boysson-Bardies, 1993). But babies babble according to the input they receive; they are not born tuned to verbal sounds. If the input is spoken language, they babble vocally. If the input is sign language, they babble on their hands. In other terms: they engage in (still poor) imitation/simulation. (How would babies babble if the patterned structure of natural language were exposed to them via abstract symbols on a computer screen?)

In short, “resonating” stimuli, whether verbal sounds or hand gestures, are simply more salient than non-resonating stimuli (because they provide information beyond pure perception). Assuming that we are predisposed to seek for optimal solutions to minimize processing requirements, we used this argument to speculate why words (as opposed to non-reproducible sounds) provide a strong(er) clue allowing infants to rapidly form categories. Other stimuli may also do the job, but “resonating” stimuli do it better.

de Boysson-Bardies, B. Ontogeny of language-specific syllabic production. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P MacNeilage, & J Morton (Eds). Developmental neurocognition: specch and face processing in first year of life (pp. 353-363). Dordrecht: Kluwer Academic.

Namy, LL & Waxman, S. Words and gestures: infants' interpretations of different forms of symbolic reference. Child Dev. 1998 Apr;69(2):295-308.

Petitto LA, Holowka S, Sergio LE, Ostry D. Language rhythms in baby hand movements. Nature. 2001 Sep 6;413(6851):35-6.

Petitto LA, Holowka S, Sergio LE, Levy B, Ostry DJ. Baby hands that move to the rhythm of language: hearing babies acquiring sign languages babble silently on the hands. Cognition. 2004 Aug;93(1):43-73.

Words, pointing and concepts

Anne Reboul
Jun 24, 2004 9:46 UT

I'm sympathetic with the ideas in Veronique and Tatjana's paper and am particularly interested by the last section and by their analysis of Waxman's experiments. My query bears on exactly the same sentence as Gloria's. I just want to point out that the fact that words belong to language may be more crucial than the authors allow because of all the well-known fact that children, from a very early age, are more interested in

hearing verbal as opposed to non-verbal sounds. A possible explanation for that, in keeping with the authors' thesis, would be that verbal sounds do indeed resonate while non-verbal sounds do not. This would make language special only in as much as it is imitable and not as engaging, for instance, a specfic cognitive module. My worry is over whether this would not be extending the very notion of 'resonance' rather farther than it should be. In the experiments which have established the specific interest of infants for verbal sounds, the infants were just hearing the sounds, not seeing people producing them. It is not clear to me whether and how plain hearing of verbal sounds would engage "resonance", especially given that very young infants are actually quite limited in the sounds that they themselves can produce. Indeed, when babling begins at around six months of age (quite late relative to the specific interest of infants for verbal sounds, which has been evidenced very early after birth), infants are still pretty limited in the range of syllables they can produce (Mac Neilage & Davis 2002). Thus, though the idea of resonnance is interesting (the authors make it clear for instance that it is involved in discriminating between conspecifics and non-conspecifics, an ability which arguably might be a condition for or a component of ToM), its relations with imitation may be in need of specification. I'd be interested in having the authors' opinion about this.

REFERENCES MacNeilage, P.F. & Davis, B.L. (2002) "On the origins of intersyllabic complexity", in Givon, T & Malle, B.F. (eds): The evolution of language out of pre-language, Amsterdam, Benjamins, 155-170.

The Mirror System Hypothesis. Linking Language to Theory of Mind

Michael Arbib (University of Southern California)

(Date of publication: 14 September 2004)

Abstract: In a recent BBS article, Arbib (2005) offers a lengthy exposition of the Mirror System Hypothesis (MSH) which links the brain mechanisms for language to the mirror system for grasping. This discussion article takes advantage of Commentaries on that exposition to ground a discussion of how this hypothesis might be extended to pay more attention to Theory of Mind (ToM).

Introduction

The original Mirror System Hypothesis (Arbib & Rizzolatti, 1997; Rizzolatti & Arbib, 1998) states that:

H1. The parity requirement for language in humans is met because Broca's area evolved atop the mirror system for grasping with its capacity to generate and recognize a set of actions.

Arbib (2005) follows Arbib (2002) in distinguishing a *language-ready brain* (equipping the child to learn a language) from a *brain that "has" language* (in the sense of, e.g., an innate "principles and parameters" Universal Grammar). It also stresses the role of "complex imitation" in the evolution of the language ready brain, and offers an account based on the following stages. (When I speak of a "stage" in phylogeny, I do not have in mind a discontinuous change in the phenotype, but rather the coalescence of multiple changes that can be characterized as forming a global pattern that may emerge over the course of tens of millennia.)

The first three stages are pre-hominid:

S1: Grasping.

S2: A mirror system for grasping shared with the common ancestor of human and monkey.

S3: A simple imitation system for object-directed grasping through much repeated exposure. This is shared with common ancestor of human and chimpanzee.

The next three stages then distinguish the hominid line from that of the great apes:

S4: A complex imitation system for grasping.

S5: *Protosign*, a manual-based communication system, breaking through the fixed repertoire of primate vocalizations to yield an open repertoire.

S6: *Proto-speech*, resulting from the ability of control mechanisms evolved for protosign coming to control the vocal apparatus with increasing flexibility.

The final stage is claimed to involve little if any biological evolution, but instead to result from cultural evolution (historical change) in *Homo sapiens*:

S7: *Language*: the change from action-object frames to verb-argument structures to syntax and semantics; the co-evolution of cognitive and linguistic complexity.

The Mirror System Hypothesis is simply the assertion that the mechanisms which get us to the role of Broca's area in language depend in a crucial way on the mechanisms established in Stage **S2**. The above seven stages provide just one set of hypotheses on how this dependence may have arisen.

Looking in more detail at Arbib (2005), we find that in addition to analyzing the above stages, it replaces the original MSH by four hypotheses:

H2. Language readiness evolved as a multi-modal manual/facial/vocal system with protosign providing the scaffolding for protospeech – these then co-evolved in an *expanding spiral* to provide "neural critical mass" for protolanguage.

H3. Protolanguage was holophrastic – "protowords" were semantically more akin to phrases or sentences of modern language than words "as we know them".

H4. Biological evolution gave humans a language-ready brain, but the emergence of human languages from protolanguage was a matter of history, not biology.

H5. While the original MSH focused on macaque F5 and Broca's area, F5 is part of a larger F5-PF-STS system in the macaque and this "lifts" to a larger frontal-parietal-temporal language-ready system in the human brain.

Between them, H2-H5 constitute an extended MSH. What needs stressing is that *these four hypotheses are almost independent* – and thus each must stand on its own. My Response to the commentaries in BBS

addresses arguments pro and con each of these four hypotheses. In the present discussion article, I want to focus on those commentaries which, while generally endorsing the role of the mirror system in evolution of the language-ready brain, stress the need to further consider mechanisms supporting conversation, motivation, prosody and theory of mind.

Complex Imitation

I want to stress (as Arbib 2002 did not sufficiently do) that the ability for complex imitation has a perceptual side whose importance cannot be overemphasized. This is the ability for *complex action analysis*: recognizing another's performance as combining actions which can be approximated by variants of actions already in the repertoire a set of familiar actions and then repeat them. Complex imitation then rests on the ability to exploit that analysis to ground imitation of the observed action (which can be further tuned by experience).

When the child is acquiring language, the whole process of complex imitation comes into play as the child acquires phonology and lexicon, and learns which "sentential actions" may be deployed to achieve its goals. However, when adults talk to each other, it is only the *complex action analysis* (recognizing what the other said) that comes into explicit play. Thus Bickerton (2005) is quite right to observe that when someone addresses you, you do not just imitate what they said. The human mirror system creates a representation that can be used for feedback control, imitation (which monkeys do not exhibit) or generating some appropriate response while inhibiting mimicking. Only in pathology does this inhibition fail, yielding compulsive imitation (echopraxia; Podell et al., 2001).

Conversation

The developing child must learn both affordances (opportunities for action as presented in the sensory stream) and *effectivities* (what the body can do; Shaw & Turvey, 1981) as two sides of the mirror system. By directing the child's attention to its own effectivities in relation to affordances, the caregiver narrows the search space for learning, and thus enhances that learning (Zukow-Goldring, 1996). These practices may pave the way to early word learning (Zukow-Goldring, Rader, & Cain, 2001). The prolonged period of infant dependency in humans combines with caregiving to provide conditions for complex social learning.

But in some sense the success of the caregiver depends on both the attention of the caregiver to the child and of the child to the caregiver. Nagy and Molnar (2004) demonstrate that newborn infants communicate by using "imitation" right after birth. The cycle of turn taking in "imitating" a small repertoire of "almost innate" gestures is crucial in establishing the social pattern of turn taking (cf. "motherese"; Falk 2004). I agree, but suggest that this neonatal imitation is based on moving single effectors and thus differs from goal-directed imitation. (Studdert-Kennedy, 2002 discusses data consistent with the view that the infant at first imitates sounds by moving one articulator and only later coordinates articulators.) Social reciprocity in neonatal imitation may be a necessary precursor for complex imitation, establishing that "I am like the other" (Zukow-Goldring, 2005). This suggests an innate basis for "conversation" that precedes its pragmatic function; Nagy (2004) suggests that language develops from these early intersubjective "conversations" (Trevarthen, 2001). Biological evolution may have selected for neonatal imitation as a basis for complex imitation.

Motivation

Prudkov (2005) downplays the key role I hypothesize for complex imitation in the evolution of the language-ready brain. Instead, he argues that the complexity of languages builds on the ability of the human brain to construct diverse goals. He suggests that animals can only form learned motivations when basic drives are activated. However, animals can acquire secondary reinforcers, etc. Chimpanzees have the ability to develop non-innate subgoals (e.g., cracking nuts). The mirror system is well-linked to the motivational system in the macaque. Arbib (2005) shows that the F5 mirror system for grasping is best understood within the larger F5-PF-STSa mirror system for manual and oro-facial actions. Rizzolatti et al. (2001) observe that STSa is also part of a circuit that includes the amygdala and the orbitofrontal cortex and so may be involved in the elaboration of affective aspects of social behavior. Thus Prudkov's transition to "non-innate motivation" may be less proximate for the evolution of the language-ready brain *per se* than complex imitation, which made possible the rapid acquisition of new skills.

Theory of Mind

Let me first state my prejudice that "Theory of Mind" (ToM) is a slippery concept, and that I think it is better looked at as an assemblage of diverse skills for relating one's own action to those of others. Thus different creatures may have different potentials for ToM, and the particular bunch of ToM skills exhibited by a group of humans may reflect their history perhaps even more than their biology.

Arbib (2002) argues that pantomime plays a key role in the transition from complex imitation of praxis (using the hands for manipulation) to complex imitation for communication (protosign as scaffolding for protospeech – see Fogassi & Ferrari (2004), Arbib (2004) and MacNeilage & Davis (2004) for further debate on this notion). However, Fabrega (2005) stresses that successful pantomime presupposes social cognition,

awareness of self, and goal setting. In other words, if we take mirror neurons "seriously" as a tool for recognizing the actions of others (I claim that was an exaptation of their earlier role in providing visual feedback for dexterous hand movements) – whether for imitation or communication then the social role ("I am like the other") must be an explicit part of our account. In this respect, it is worth noting, with Arbib & Mundhenk (2004) that mirror neurons per se only recognize an action – it is the task of other parts of the brain to recognize the agent of the action and to bind the neural representations of agent and action appropriately.

Thinking this through returns us to the topic of conversation: utterances are usually part of an ongoing interaction between 2 or more speakers or signers. In line with this, Kotchoubey (2005) emphasizes pragmatics, e.g., what we say depends on the mental state of our "hearer": However, his claim "We do not use language to transmit information, but to persuade and motivate." seems a false dichotomy. "Look at this beautiful flower" combines information "This flower is beautiful" and persuasion "Look at this flower". Kotchoubey (p.c.) stresses that his starting point is cooperation between two or more humans, reinforcing the claims of MSH for relating praxic and communicative actions.

Williams (2005) sees the greatest evolutionary advantage conferred by spoken language as its ability to communicate mentalistic concepts (Theory of Mind, ToM). He stresses selection pressure for social maneuvering where I have emphasized physical activities. However, I think he would agree that this is not a simple case of either/or. Surely the two "domains of discourse" complement each other. Williams (2005) notes the possible role of the mirror neuron system in mentalizing and this is indeed a theme that has been developed by Gallese (e.g., 2003) and others (e.g., Meltzoff & Decety, 2003). We need to investigate whether an account can be given of a shared evolution of "mirror systems" suiting both ToM and complex imitation. I hypothesize that the ancestral mirror system for manual praxis was distinct from the putative mirror system for facial expression of emotion. The former would support pantomime and thence on to multi-modal symbols; and then the availability of symbols could enrich the latter to yield rudiments of ToM.

Williams (2005) and Théoret & Fecteau (2005) see autism as providing a window on the role of the mirror system in ToM and language. (Théoret & Fecteau add analysis of blindness). Deficits in autism are prominent in speech associated with social communication but praxic aspects of language are fairly well preserved. Perhaps what is affected is not so much language per se as the integration of this with affect and ToM. Interestingly, autistics may exhibit stereotypic mimicking (which monkey's *don't* have). Thus it must be reiterated that a fully functional human mirror system can inhibit mere repetition (echopraxia and echolalia) and instead relate the perception of perceived actions to the planning of an appropriate course of action.

Prosody

Kotchoubey (2005) and Fitch (2005) note that my emphasis on cognitive-symbolic aspects of language ignores prosody. Kotchoubey notes that prosody subserves both affective prosody (emotional expression) and linguistic prosody (as in distinguishing between an assertion and a question) and that both forms of prosodic information are processed mainly in the right temporal lobe. In similar vein, Gilissen (2005) notes that human vocal behavior resembles monkey calls in the emotional intonations superimposed on the verbal component though I would stress [sic] that such calls lack the semantic openness that is the hallmark of protolanguage and, a fortiori, language. Kotchoubey (p.c.) observes that in many languages,intonation is the only distinction between question and declaration. He thus suggests that linguistic prosody is a part of the right hemisphere so closely controlled by the left that they cannot work without each other. This is reminiscent of the coupling of gesticulations to the syntax and semantics of a specific language (Kita & Özyürek, 2003).

Gilissen (2005) cites Falk's (2004) evolutionary perspective on the hypothesis that, as human infants develop, a special form of infant-directed speech (motherese) provides a scaffold for their eventual acquisition of language. This enriches our discussion of the role of the caregiver in neonatal "conversation". Gilissen says that the special vocalizations of human motherese are in marked contrast to the relatively silent mother/infant interactions that characterize chimpanzees, yet suggests a possible link between monkey calls and motherese. This apparent contradiction suggests that the *affective* content of motherese (and protolanguage) builds upon the monkey vocalization system, but the *information* content of motherese (and protolanguage) has a complementary evolutionary history. Kotchoubey (2005) suggests that the left hemispheric subsystem develops as described by MSH to subserve the cognitive-symbolic function, whereas the right hemispheric subsystem is a direct successor of monkey vocalization mechanisms and gives language its intonational color. It is a long standing observation (Hughlings Jackson 1879-80) that imprecations survive damage to the human brain that blocks normal speech. Arbib (2002) thus suggested that the language-ready brain integrates action-oriented and affect-oriented systems in a pattern of cooperative computation.

Fitch (2005) adopts Darwin's hypothesis that our prelinguistic ancestors possessed an intermediate "protolanguage" that was musical and that music scaffolds the early structural and imitative aspects of language (prosody). He sees the semantic stage as coming later. However, even if we accept the

importance of “musicality”, it does not follow that the co-evolution of vocal and manual gesture is tied more closely to music than pantomime and linguistic communication – but it does encourage us to investigate how dance and music might enrich MSH.

References

Arbib, M.A. (2002) The mirror system, imitation, and the evolution of language. In Nehaniv, C. & Dautenhahn, K. (Eds.). *Imitation in Animals and Artefacts.* MIT Press, Cambridge, MA:229-280.
Arbib, M.A. (2004) Interweaving Protosign and Protospeech: Further Developments Beyond the Mirror, *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems* (to appear).
Arbib, M.A., 2005, From Monkey-like Action Recognition to Human Language: An Evolutionary Framework for Neurolinguistics, *Behavioral and Brain Sciences*. (In press.)
Arbib, M.A., and Mundhenk, T.N., 2004, Schizophrenia and the Mirror System: An Essay, *Neuropsychologia* (in press).
Arbib, M.A. & Rizzolatti, G. (1997) Neural expectations: a possible evolutionary path from manual skills to language. *Communication and Cognition* 29: 393-423.

Bickerton, D., (2005): Beyond the Mirror Neuron – The Smoke Neuron?*, Behavioral and Brain Sciences,* (in press).
Fabrega, Jr, H., (2005): Biological evolution of cognition and culture: off Arbib’s mirror neuron system stage?*, Behavioral and Brain Sciences,* (in press).

Falk, D. 2004. Prelinguistic evolution in early hominins: Whence motherese. *Behavioral and Brain Sciences*
Gilissen, E., (2005): Imitation systems, monkey vocalization, and the human language*, Behavioral and Brain Sciences,* (in press).
Williams, J., (2005): Language is fundamentally a social affair. *, Behavioral and Brain Sciences,* (in press).
Jackson, J.H., 1878-79, On affections of speech from disease of the brain. Brain 1: 304-330; 2: 203-222, 323-356.
Kita, S., & Özyürek, A. 2003. What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking, Journal of Memory and Language 48(1):16-32.
Kotchoubey, B., (2005): Pragmatics, prosody, and evolution: Language is more than a symbolic system*, Behavioral and Brain Sciences,* (in press).
MacNeilage, P.F. & Davis, B.L. (2004) The frame/content theory of evolution of speech: Comparison with a gestural origins theory*. Interaction Studies: Social Behavior and Communication in Biological and Artificial Systems.* (to appear)
Nagy, E., (2004): The first dialogue: ‘conversation’ through imitation with newborn infants (personal communication to M.A. Arbib).
Nagy, E., & Molnar, P. (2004) Homo imitans or homo provocans? The phenomenon of neonatal initiation. *Infant Behavior and Development, 27:*57-63.
Podell, K, Wisniewski, K, & Lovell MR (2001) The assessment of echopraxia as a component of executive control deficit in traumatic brain injury, *Brain and Cognition* 47(1-2):349-353.
Prudkov, P.N., (2005): Motivation rather than imitation determined the appearance of language*, Behavioral and Brain Sciences,* (in press).
Rizzolatti, R., Fogassi, L. & Gallese, V. (2001) Neurophysiological Mechanisms Underlying the Understanding and Imitation of Action*, Nature Reviews Neuroscience*2:661 -670.
Shaw, R., & Turvey, M. (1981) Coalitions as models of ecosystems: A realist perspective on perceptual organization. In M. Kubovy & J. R. Pomerantz, Eds.), *Perceptual organization* (pp. 343-415). Hillsdale, NJ: Erlbaum.
Théoret, H. and Fecteau, S., (2005): Making a case for mirror neuron system involvement in language development: what about autism and blindness ?*, Behavioral and Brain Sciences,* (in press).

Trevarthen, C. (2001) The neurobiology of early communication: Intersubjective regulations in human brain development. In. Kalverboer, A. F. & Gramsbergen, A. (Eds). *Handbook on Brain and Behavior in Human Development.* pp. 841-882. Dodrecht, The Netherlands: Kluwer.
Fitch, W.T. (2005) Protomusic and Protolanguage as Alternatives to Protosign*, Behavioral and Brain Sciences,* (in press).
Zukow-Goldring, P. (1996). Sensitive caregivers foster the comprehension of speech: When gestures speak louder than words. *Early Development and Parenting*, 5 (4), 195-211.
Zukow-Goldring, P. 2005. Assisted Imitation: Affordances, Effectivities, and the Mirror System in Early Language Development, in *Action To Language via the Mirror Neuron System* **(**Michael A. Arbib, Editor), Cambridge University Press (in press).
Zukow-Goldring, P., (2004): Assisted imitation: Linking action sequences, communication, and the emergence of language *(personal communication to M.A. Arbib)*.

Zukow-Goldring, P., Rader, N., & Cain, T. (under review). Dynamic gestures and early word learning: A test of a perceptually based theory.

Discussion

Protosign and cultural evolution: how plausible are they?

Anne Reboul
Sep 17, 2004 11:12 UT

Michael Arbib's contribution is extremely interesting and has the (slightly frustrating) advantage that it gives a precis of his future publication and peer reviews in BBS. I have two queries related to S5 and S7 and H2. My first question, concerning S5 and H2, is on why one should suppose a protosign (i.e. manual communication) to be necessary. Granted apes and monkeys communicative repertoires seem both limited and fixed, but why should gestures be necessary to break through these limits? Why not go directly to protospeech? Granted, the common ancestor to humans and apes must have been as limited as e.g., chimpanzees in its abilities to produce sounds, but why couldn't the vocal tract of hominids have co-evolved with protospeech? Why should the hypothesis of a gestural communication system stage be more plausible that the co-evolution of the vocal tract and protospeech? My other query is about S7: how plausible is cultural evolution for language? For instance, Goldin-Meadow has shown in numerous papers that deaf children raised by hearing parents and not exposed to sign language communicate with their family through signs which manifest non-random linguistic properties. This seems, to say the least, slightly weird if language is culturally evolved (and presumably culturally acquired): these children are, if any thing, accultured from a linguistic point of view. Another question is the emergence of creoles from pidgins. Thus, though I'd be happy to buy Arbib's theory, I think that these questions cannot be ignored and I'd like to know how he proposes to treat them.

The Path to Protospeech May Be Indirect

Michael Arbib
Sep 22, 2004 22:25 UT

Anne Reboul raises three questions. Here is my initial response to the first, “Why should the hypothesis of a gestural communication system stage be more plausible that the co-evolution of the vocal tract and protospeech?”

Many people would argue that protospeech evolved directly from something akin to the vocalizations we see in monkeys today (e.g., MacNeilage & Davis (2004) offer an explicit challenge to Arbib (2004)). The main arguments against this are:

Human language use is multimodal – speech is normally accompanied by manual and facial gestures (McNeill, 1992). These “cospeech gestures” or “gesticulations are language-specific, closely coupled to the syntactic structure of speech (Kita & Ozyurek, 2003).

All primates have a rich repertoire of arm, hand and face movements, and skills in active goal-seeking using manipulation. They also have a great deal of “oral dexterity”, but (a) manual control is under visual feedback, providing a wide range of hand movements available for the inspection of others even in the absence of intended communication, whereas (b) the ingestive use of “oral articulators” has few auditory correlates, and so their use for vocalization requires a dramatically different mode of control Thus, speech employs a system of articulators for which there is no rich behavioral repertoire of sound producing movements to build upon.

Recent neurophysiology helps us expand our view of the MSH to build on the orofacial as well as the manual mirror system of macaques: see Kohler et al. (2002) for manual mirror neurons responsive to sounds, Ferrari et al. (2003) for orofacial motor neurons responsive to observation of orofacial communicative gestures, and Fogassi & Ferrari (2004). However, there is no evidence that this offers control of an articulatory system with the combinatorial richness offered by manipulation.

I thus see great power in the notion that the use of the manual system for pantomime and then conventional gesture (see my reply to Origgi, “Intentional Agents and the Mirror System”) provided the scaffolding for the brain’s exploitation of vocalization to represent an open-ended semantics by combining otherwise meaningless gestures.

Arbib, M.A. (2004) Interweaving Protosign and Protospeech: Further Developments Beyond the Mirror, Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems (to appear).

Ferrari PF, Gallese V, Rizzolatti G, Fogassi L. (2003) Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex.. Eur J Neurosci. 17(8):1703-14.

Fogassi, L., and Ferrari, P.F., 2004, Mirror neurons, gestures and language evolution, Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems (to appear)

Kita, S., & Ozyurek, A. 2003. What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking, Journal of Memory and Language 48(1):16-32.

Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. Science, 297(5582), 846-848.

MacNeilage, P.F. & Davis, B.L. (2004) The frame/content theory of evolution of speech: Comparison with a gestural origins theory. Interaction Studies: Social Behavior and Communication in Biological and Artificial Systems. (to appear)

McNeill, D. 1992. Hand and Mind: What gestures reveal about thought. Chicago: University of Chicago Press.

Is Home Sign a Language?

Michael Arbib
Sep 22, 2004 22:43 UT

Anne Reboul raises three questions. Here is my initial response to the second, “Deaf children raised by hearing parents and not exposed to sign language communicate with their family through signs which manifest non-random linguistic properties. Doesn’t this contradict the view that language is culturally evolved?”

I am no expert on the work of Goldin-Meadow so offer some amateurish thoughts for others more expert to build on. I would simply observe the following: (a) the children do have language-ready brains, though deprived of auditory input; (b) they are raised by human caretakers who use pointing and action and gesture and language in ways that are different from those the child comes up with, but which provide a “language-enriched environment” very different from that which I imagine early hominids would have had and which educate the child in “what can be symbolized.” (c) But even with these 2 advantages, I believe that the children come up with a “protolanguage” rather than a full language. Do they use relative clauses, subjunctives, past pluperfects?

In short, I think Goldin-Meadow’s observations do “push the envelope” as to what “ingredients” the language-ready brain may provide, but I don’t think they give any evidence for, e.g., an innate “principles and parameters” Universal Grammar “wired into the brain” (see my response to Reinhart-Waller, “A Lad with a LED or a LAD?”).

Finally, let me note that apes raised by humans master a form of protolanguage completely different from the forms of communication seen in the wild – again emphasizing the role of culture in shaping the development of the infant brain.

Creoles and Pidgins

Michael Arbib
Sep 22, 2004 22:58 UT

Anne Reboul raises three questions. Here is my initial response to the third, “What about the emergence of creoles from pidgins?”

I assume that this question is inspired by Dennis Bickerton’s claim that Hawaiian Creole emerged “all at once” with a grammatical structure that was not apparent in the pidgin that preceded it, and that this supported his Bioprogram Hypothesis which is akin to the assumption of an innate Universal Grammar. I.e., children who filtered pidgin through their innate Universal Grammar based Language Acquisition Device imported the grammar not from their culture but from their genomic Universal Grammar. However, my understanding is that while Bickerton is widely respected by people who study pidgins and creoles, those “Universalists” who agree with his Bioprogram Hypothesis are definitely in the minority. There seems to be good evidence the grammatical structure of a Creole will bear a strong resemblance (albeit simplified) to one or more of the languages from which it comes. Moreover, novel features of the creole grammar can often be traced to a process of grammaticalization acting on a phrase of the pidgin. I don’t have the reference at hand to check the details, but Toksin Pisin (New

Guinea) had in its pidgin days the word bymbi (from the English bye and bye) added to a sentence to mark it as referring to the future; as Toksin Pisin became a creole, this was contracted to "by" (approximately!) and added to the verb.

Holm, J. (2000) An Introduction to Pidgins and Creoles, Cambridge: Cambridge University Press.

Conversation, social imitation and intentionality

Gloria Origgi
Sep 17, 2004 11:31 UT

Thanks to Michael Arbib for this very clear and concise paper. The paper adresses some of the commentaries to his BBS longer article: "From Monkey-like Action Recognition to Human Language: An Evolutionary Framework for Neurolinguistics" that will be published in 2005, in particular those who stress the need for a more fine grained analysis of mechanisms supporting conversation, motivation, prosody and theory of mind.

I have a question about his claim on the role of social imitation in conversation, Arbib agrees that social reciprocity in neonatal imitation may be a precursor of the ability of complex imitation needed to learn a language, thus conversation could have an innate basis that precedes its pragmatic function. But these pre-linguistic conversations with babies acquire a "pragmatic" dimension quite early: children are surprised for a "violation" of an expected conversational move. They are sensitive to the intentional dimension of the prelinguisitc conversation, that is, to the fact that they are in front of intentional agents.

I do not see how you can go from a readiness to imitate a repertoire of gestures to the sensitivity to the intentional aspects of conversational interaction. Could you be more explicit of this point?

Intentional Agents and the Mirror System

Michael Arbib
Sep 22, 2004 11:15 UT

Gloria Origgi notes that children are sensitive to the fact that they are in front of intentional agents, and then asks how I can go from a readiness to imitate a repertoire of gestures to the sensitivity to the intentional aspects of conversational interaction.

Perhaps the key point is that the Mirror System Hypothesis is not based on the recognition of gestures! It is based on the recognition of actions, where action = movement + goal (Arbib & Rizzolatti, 1997). Thus in the model of Oztop & Arbib (2002), the monkey mirror system processes visual input concerning the trajectory of the hand relative to an affordance of (opportunity for grasping) an object. More generally, then, I see a mirror system as stage (ii), or beyond, in the following evolutionary progression:

(i) a generalized feedback system whereby the organism evaluates its progress towards some goal and can correct its movement accordingly;

(ii) a mirror system which rests on the organism's ability to map its body on (or from) that of another, and recognize which action (i.e., both choice of goal and movement to achieve it) another organism is conducting; and

(iii) a capacity for imitation, i.e., to go from recognizing a novel action by another to adding the action to its repertoire, i.e., becoming able to achieve the observed goal by some "coordinated control program" [Arbib, 1981] abstracted from the observed pattern of movement.

Note that a mirror system may be quite specific, so that different mirror systems might mediate manipulatory actions, facial expression of emotion (see "Emotion is Not a Language", my response to Chris Lofting), and vocalizations (see "The Path to Protospeech May Be Indirect", one of my responses to Anne Reboul).

Returning to Origgi's comment that children are sensitive to the fact that they are in front of intentional agents, I I would say more generally that "any creature with a mirror system is able to recognize the intentions of conspecifics to the extent that those intentions fall within the scope of that mirror system" – my point being that the notion of "intentional agent" is a rather sophisticated abstraction from a multiplicity of such recognitions.

How, then, can I trace an evolutionary path from ancestors with basic mirror systems to human children who are both "intention-aware" and "language ready"? Elsewhere (e.g., Arbib, 2005) I suggest how hominids might have evolved from recognizing goals and evaluating how well a movement reaches it (iv) to inferring the goal from a pattern of movement and (v) come to pantomime movements

(a big shift) to communicate the idea of the goal to the observer – this then led to two "smaller" shifts, the first from pantomime of hand movements to pantomime of other actions or objects (e.g., using the hands to imitate flapping wings to signify "bird flying" or "bird" or "flying") and then the development by a community of conventional gestures to disambiguate the pantomimes. Once this system of conventional gesture is in place, the path is open to protosign (and protospeech – again, see "The Path to Protospeech May Be Indirect") within a general context of communicating about goals and thus about intentions.

Arbib, M. A., 1981, Perceptual structures and distributed motor control, in Handbook of Physiology – The Nervous System II. Motor Control (V. B. Brooks, Ed.), Bethesda, MD: American Physiological Society, pp.1449-1480.

Arbib, M.A., 2005, From Monkey-like Action Recognition to Human Language: An Evolutionary Framework for Neurolinguistics, Behavioral and Brain Sciences. (In press.)

Arbib, M., and Rizzolatti, G., 1997, Neural expectations: a possible evolutionary path from manual skills to language, Communication and Cognition, 29, 393-424.

Oztop, E., and Arbib, M.A. (2002) Schema Design and Implementation of the Grasp-Related Mirror Neuron System, Biological Cybernetics, 87:116–140.

Reading actions and reading speech acts

Gloria Origgi
Sep 23, 2004 12:20 UT

Arbib writes : *"How, then, can I trace an evolutionary path from ancestors with basic mirror systems to human children who are both "intention-aware" and "language ready"? Elsewhere (e.g., Arbib, 2005) I suggest how hominids might have evolved from recognizing goals and evaluating how well a movement reaches it (iv) to inferring the goal from a pattern of movement and (v) come to pantomime movements (a big shift) to communicate the idea of the goal to the observer – this then led to two "smaller" shifts, the first from pantomime of hand movements to pantomime of other actions or objects (e.g., using the hands to imitate flapping wings to signify "bird flying" or "bird" or "flying") and then the development by a community of conventional gestures to disambiguate the pantomimes."*

If this is so, then it seems that "intention-aware" and "language-ready" children have a readiness for an intentional reading of those special interactions that have to do with speech and conversation.

My point is that children have a sensitivity to "language"-based intentional interactions that seems to be more fine-grained than a simple ability to imitate or recognize intentional actions. Speech interactions, even in the case of the pre-linguistic child, are treated in a different way than any other interaction. This suggests (as Origgi-Sperber 2000 have argued) that pragmatic competences can follow inferential routines that are quite distinct from other theory of mind-based competences. That's why I was puzzled by the idea of a development of pragmatic competences from complex imitation and social reciprocity.

Origgi, G; Sperber, D. (2000) "Evolution, Communication and the Proper Function of Language" in P. Carruthers, A. Chamberlain *Evolution and the Human Mind* Cambridge UP.

Inferencing and the Mirror System are Partners in Understanding

Michael Arbib
Sep 24, 2004 22:49 UT

Origgi observes that "Speech interactions, even in the case of the pre-linguistic child, are treated in a different way than any other interaction."

What is the evidence for this? Trivially, speech is different from non-vocal communication and from certain forms of vocal communication (crying, laughing, screaming, etc.). But I would have thought that in a deep sense speech is part of a complex facial-vocal-gestural communication system which preserves the properties I think Origgi has in mind even when the vocal component is missing.

Origgi then "suggests (as Origgi-Sperber 2000 have argued) that pragmatic competences can follow inferential routines that are quite distinct from other theory of mind-based competences. That's why I was puzzled by the idea of a development of pragmatic competences from complex imitation and social reciprocity."

There are two issues here:

a) I agree completely with Origgi and Sperber on the importance of inferential processes but see these as integral to the full utility of a mirror system – and vice versa – rather than being in competition. It is no good to recognize an action unless one can infer the likely reasons for another's actions and the likely consequences for others and oneself; but equally the inference needs accurate observations to get started, and the mirror system provides a representation of an action that links it into one's experience with both observation and execution of the action, thus giving access to a wider knowledge base for inference than would be available for either alone.

b) But, having said that, I am not clear what evidence there is that speech (or language use) "can follow inferential routines that are quite distinct from other theory of mind-based competences." Wouldn't playing poker be a "theory of mind-based competence" that exercises similar inferential capabilities without necessary employing language even implicitly in the process?

What happened to Emotion as Language?

Chris Lofting
Sep 17, 2004 12:30 UT

Of note in Arbib's prose is the lack of reference to emotion as being a language, the language of the non-verbal; a language where the qualities to be used to derive meaning through speech have their roots.

Through analysis of the properties of the fight/flight dichotomy, a dichotomy 'ruling' the amygdala of our brains and strongly associated with the derivation of, and transmission of, meaning, we can derive a set of generic emotions possible for communication in non-speaking life forms. Applying the above dichotomy recursively (and so using self-referencing) we can come up with (after Plutchick):

anger sexual love acceptance surprise anticipation rejection grief/sadness fear

These qualities can be formed into compounds and as such give us a language through the use of emotions. (for refs etc see http://www.austarmetro.com.au/~lofting/myweb/emote.html)

Since the fight/flight dichotomy is sourced in other neuron-dependent, non-speaking, life forms so the roots of language are 'deeper' than focused on some gesture etc. to a degree where the foundations are sourced in the dynamics of neurons. As such, the tiny zebra fish has a 'brain' that makes the same associations as ours do in the categorising of the KNOWN from the UNKNOWN. IOW there is a 'meaning' system developed at the level of the neuron that needs complex development to be exploited enough for the level of communications that we, as a fully conscious species, use; BUT it can still serve 'lower' life forms in their communications with each other through emotion.

In the context of Arbib's focus on mirroring etc., if we review carefully the above qualities we find that anger and sex share the same source, a more generic focus on context REPLACEMENT. Here a life form can take over a context, assert its 'will' on that context through replacement. This replacement is done through either (a) erradication (through anger) or (b) replication (through sex - we flood the context with images of our selves).

The complement for replacement is reflected in the qualities of sadness/fear that seem to reflect a focus not on replacement but on coexistence; fear elicits a 'need' to dissapear and that 'into' the context. As such, integrating with the context aids in PROTECTION and we coexist with our 'enemy'. Sadness/grief covers the loss of a 'loved one' or a loved 'ideal' etc etc where that love one continues to exist 'in here' in the form of memories.

What is noticable in this in the context of Arbib's concepts is that the quality associated with sexual love is a quality encompassing mirroring, copying, reflecting, replicating etc. IOW the notion of 'mirror' and 'copying' is already encoded in the realm of emotions, a realm covering many other life forms besides us.

The differences between us and lower life forms is more of DEGREE where our high neural complexity allows us to be VERY precise in expression through the development of discrete units of communications (sound etc) that take the immediate expression of emotion and stretch it out, refine it, and re-transmit it. This reflects communications through spectrum exchange; we break the full communication into parts and then re-transmit. This process is applicable to ANY neuron-dependent life form but strongly so in us due to our developed 24/7 consciousness (lower life forms seem to have 'snippets' of awareness spanning hours etc)

This issue of PRECISION is reflected (!) in our brains with, in most, the left and front areas of our brains being more 'discrete' in operation, more particular, more 'pointed' in precision, when compared to the right and back areas of our brains that are more 'continuum' oriented, more general, more 'approximate' in precision.

The development of the 'left' element, the more precise element, occurs with the development of the spoken word - infants are born with a well developed right side that 'gives way' to the still developing left as it refines its focus on the discrete, and so on differentiating as compared to integrating.

It is this development of precision in one side (the left in most) that elicits in us a marked bias in 'sidedness' and so the RIGHT side dominates overall (being controlled by the left hemisphere). In 'lower' life forms this degree of precision and so bias is not as strong, handedness being more context-sensitive that the almost universal bias we have developed to right-handedness, footedness, etc etc.

These issues of emotion as language and the development of precision are fundamentals that cannot be left out of reflections on language development; Arbib has left them out which is, IMHO, unfortunate.

Chris Lofting.

main site: http://www.austarmetro.com.au/~lofting/myweb email: chrislofting@ozemail.com.au lofting@austarmetro.com.au

Emotion is Not a Language

Michael Arbib
Sep 22, 2004 11:40 UT

Chris Lofting offers a set of generic emotions possible for communication in non-speaking life forms:

anger sexual love acceptance surprise anticipation rejection grief/sadness fear

and asserts that “These qualities can be formed into compounds and as such give us a language through the use of emotions.” He makes various interesting observations such as “the quality associated with sexual love is a quality encompassing mirroring, copying, reflecting, replicating etc. In other words, the notion of 'mirror' and 'copying' is already encoded in the realm of emotions, a realm covering many other life forms besides us,” and he asserts that “The differences between us and lower life forms is more of DEGREE where our high neural complexity allows us to be VERY precise in expression through the development of discrete units of communications (sound etc) that take the immediate expression of emotion and stretch it out, refine it, and re-transmit it.”

I agree that there probably is a mirror system for the facial expression of emotion, and that a full development of the Mirror System Hypothesis of language evolution should take this into account. In other respects, I think his essay is rather unhelpful. The word “language” can and has been used metaphorically for any system which supports a set of messages - whether “the language of the bees”, “the language of flowers” or (Lofting’s choice) the “language of emotions”. However, I want to reserve the term language for a system of communication that is “something like” human language, in its use of words to refer to objects and actions and so much more, and its use of some form of syntax to support the creation of new messages which others can understand thanks to a compositional semantics.

It is hardly a satisfactory evolutionary theory to simply say “emotions are expressive so it's just a matter of degree to get to language”. One wants to offer plausible evolutionary intermediates that drive that change in degree, and I personally want to understand the neural underpinnings. If we look at the neurology, we find that different lesions can yield a person who retains the symbolic use of language but is devoid of emotional expression or a person who has emotional imprecations and emotion but has lost compositionality.

I think the interesting challenge concerning emotion is not the one Lofting offers, but rather to understand how the human brain evolved to be both “language ready” and to have a set of emotions immensely enriched by the interaction of their neural substrate with cortical mechanisms supporting cognition. The reader may find some pointers to the relevant literature in Arbib & Fellous (2004) which also contains speculations on the implications of the neurobiology of emotions for the present and future design of “emotional robots”. What may be of particular relevance is the discussion of the two-fold role of emotion:

1. Emotional expression for communication and social coordination.

2. Emotion for organization of behavior (action selection, attention, social coordination and learning).

Arbib, M.A., and Fellous, J.-M., 2004, Emotions: From Brain to Robot, Trends in Cognitive Science (December 2004, in press).

Reply to Arbib
Chris Lofting
Sep 27, 2004 1:30 UT

My reply was too long so I have put it on my website - URL is:

http://www.austarmetro.com.au/~lofting/arbib.htm

Chris.

What's the difference between language-ready brain and brain having LED
Gerry Reinhart-Waller
Sep 22, 2004 1:48 UT

>>Arib (2005) follows Arbib (2002) in distinguishing a language-ready brain (equipping the child to learn a language) from a brain that “has” language (in the sense of, e.g., an innate “principles and parameters” Universal Grammar). It also stresses the role of “complex imitation” in the evolution of the language ready brain, and offers an account based on the following stages. (When I speak of a “stage” in phylogeny, I do not have in mind a discontinuous change in the phenotype, but rather the coalescence of multiple changes that can be characterized as forming a global pattern that may emerge over the course of tens of millennia.) >>

Am having much difficulty is distinguishing between a "language-ready brain" vs. what Chomsky claims to be a brain with LED. Aren't both the same thing (except for the semantics?).

If the son of a violinist decides to toy with the instrument, is he toying with his "language ready brain" or with something deeper as Chomsky implies. Doesn't much matter, does it? Aren't both shades of each other? They certainly do differ from the son of someone who doesn't play the instrument.

Arib’s Reply: A Lad with a LED or a LAD?
Michael Arbib
Sep 22, 2004 12:01 UT

Reinhart-Waller has difficulty is distinguishing between a "language-ready brain" vs. what Chomsky claims to be a brain with LED. Presumably, he means a LAD, Language Acquisition Device, and he talks of a lad learning to play the violin. This is a good example: I doubt that anyone would claim that the human brain is genetically programmed to play the violin, but would instead say that humans discovered a technology for making instruments that could be manipulated to make pleasing sounds, and that over hundreds of years improved the technology, and developed a repertoire of increasing subtlety which could be passed on to succeeding generations. The evolution of the human brain did not respond to adaptive pressures for better violin playing – but it did provide the ability to tune motor skills on the basis of auditory feedback.

I offer the same distinction with respect to language. If by LAD one simply means “a brain that can learn language”, then having a LAD is just the same as having a “language-ready brain” (just as a human child has, as described above, a violin ready brain [and body]). But, as I stated explicitly, what I was rejecting was the hypothesis that the LAD incorporates an innate “principles and parameters” Universal Grammar – i.e., I deny that the genome defines neural circuitry that evolved specifically to support the basic categories and syntactic rules of all possible human languages so that a child acquires the syntax of its native language by having a brain “set the switches” to the appropriate values, with those switches and all possible settings of those switches explicitly laid out in the genome.

A Lad with a LED
Gerry Reinhart-Waller
Sep 23, 2004 16:24 UT

Chomsky most definitely envisions a brain with a LAD (Language Acquisition Device) yet through a mistype my lad, the violin player, was given a LED as in "Light Emiting Diode". Not such an absurd thought though....a brain that is digitized has the ability to program itself in a myriad of directions not unlike the numerals on a digital alarm clock. What is needed is a map of what the presumed pathway should be. Whether or not this chart is universal, as Chomsky might say, depends on whether numerals 0-9 are recognized in all languages. If they are, then we could be talking about "Universal Grammar" whereby the grid has already been established and the pathway is of individual choice.

From parity and complex imitation to pantomime

Ingar Brinck

Sep 22, 2004 23:09 UT

The MSH relates the human capacity for language to that for intentional action, emphasizing the affinities between different kinds of social cognition such as cooperation, intentional communication, and tool use. Language use and complex social action are learnt by imitation and require the capacities to distinguish means from ends and parts from wholes, recombine parts and transform wholes in ways that fit the circumstances, transfer actions from one context to another, and to cooperate.

Social cognition is tuned to the typical traits of the species. Tool use in nonhuman primates and humans are radically different. Vygotsky remarked that by using external representations, humans can break lose from the given structure of the attention field, reconstruct their perception and master their attention. Symbolic activity has an organizing function that produces new forms of behavior. Tools and symbols function as mediators, but while tools are externally oriented, language is a means to control the behaviour of both oneself and others (Mead). Language is a second-order tool.

These observations point to a weak link in Arbib's account: the step from S4, a complex imitation system for manual action, to S5, protosign and pantomime. Arbib remarks that the goal of complex imitation is instrumental, while that of pantomime is communicative. Note that praxic action aims at changing a state of the world directly, while communicative action aims to do so indirectly.

Taking the step from S4 to S5 requires more than the ability to infer the object or goal of the action. Monkeys have been shown to have this ability (Gallese, personal comm.). In contrast to complex imitation, pantomime requires grasping that one object (action) can stand for or signify another object (action). It is a symbolic activity. I cannot see how the MSH in its present form can account for it. The mirror property of parity only accounts for how known actions can be recognised. Complex imitation explains how an action formed from a recombination of known parts can be recognized when perceived in relation to some observed or inferred goal. But it does not explain how we can understand an action like moving one's arms up and down as symbolizing a (flying?) bird. Sign use needs interpretation: seeing that x counts as y.

Two facts about gestures complicate the picture. First, not all gestures are symbols. For instance, only declarative pointing requires symbolisation. Imperative pointing is instrumental. Second, gestures are triadic or dyadic. Triadic gestures draw the attention of others to something in the environment. Dyadic gestures constitute a way of requesting behaviour of others towards oneself. These are not reciprocal: the capacity to produce them does not imply the capacity to understand others' use of them and vice versa. Parity and complex imitation account for how dyadic gestures that are part of a ritualised behaviour can initiate interactive sequences. But dyadic gestures are not similar to pantomime.

Grasping the function that an unconventionalised pantomime is meant to perform demands perceiving or 'reading' the intention behind the gesture, and perhaps even understanding how the intentions of the agent is related to one's own states and actions. This brings us to the topic of ToM. As Arbib asserts, a genuine ToM consists, at least partly, in a set of culture-dependent skills. I moreover believe that intention reading can be accomplished by attention reading. Nevertheless, simpler forms of ToM such as joint attention is necessary for grasping the notion of reference and to interpret pantomime. Attention is a control mechanism that can be used voluntarily and be directed at the attention and, indirectly, mental states of others.

Brinck, I. 1999. Attention and tool-use in the evolution of language. http://www.lucs.lu.se/spinning/index.html

Brinck, I. 2004. Joint attention, triangulation and radical interpretation. Dialectica, 58(2).

Brinck, I. 2004. The pragmatics of imperative and declarative pointing. Cognitive Science Quarterly, 3(4).

Brinck, I. & Gärdenfors, P. 2003. Co-operation and communication in apes and humans. Mind & Language, 18(5).

Mead, G. H. 1934. Mind, Self, and Society. University of Chicago Press.

Tomasello, M. 1996. The cultural roots of language. In Velichovsky & Rumbaugh, Communicating meaning. Lawrence Erlbaum.

Vygotsky, L. S. 1978. Mind in Society. Harvard University Press.

From complex imitation to pantomime: Possibly necessary but not fully explained

Michael Arbib

Sep 23, 2004 6:45 UT

Ingar Brinck starts with an invigorating reminder of the insights of Vygotsky and Mead. He then

focuses on stages S4, a complex imitation system for manual action, and S5, protosign and pantomime, in my account of the Mirror System Hypothesis (MSH). The goal of complex imitation is instrumental, while that of pantomime and protosign is communicative. Since these are very different, Brinck confesses that he cannot see how MSH in its present form can account for the transition from S4 to S5.

I don't think it does account for it. My point is that the original MSH (Arbib & Rizzolatti, 1997; Rizzolatti & Arbib, 1998) starts from the observation that human Broca's area contains a mirror system for grasping homologous to the mirror system for grasping in F5 of macaque, and then suggests that the homologous brain region of the common ancestor of macaque and human had an F5-like mirror system for grasping, and then suggests that evolution of the brain along the hominid line may have underwritten the evolution of a system for recognition of manual communicative gestures, and then on to sign language and language more generally.

The point of Arbib (2002, 2005) was that MSH needed to be refined into a series of more precise hypotheses about the transitions that occurred. I hypothesize that the transition from S4 to S5 was one of the crucial transitions, noting that there can be good selection pressure to yield creatures with complex imitation skills (stage S4) without positing an "omega of language" as the driving force. I then suggest that with S4 in place, the ability to recognize the movement component of an action to infer the goal could be enough of a biological change to then secure stage S5 on the basis of various learning abilities. In other words, I agree with Brinck that this is a significant change. I hypothesize that it did occur but have few ideas on what might have underwritten such a stage transition. I welcome suggestions from other discussants.

Brinck has useful things to say about varieties of grasping, and about attention and Theory of Mind. Another aspect would be the integration of the mirror system for recognizing facial expressions of emotion with the more instrumental observation of what another is doing as a basis for inferring his/her intentions.

What all this says is that I do not believe there is a single evolutionary step that yielded humans with language-ready brains. Although the hominid evolutionary record is woefully incomplete, we know that there were many species of hominids in the course of the last 5 million years, each representing evolutionary experiments – and we have yet to learn of all those which yielded features that survived into the human genome. But the point is that many changes occurred. Again, if we accept the current estimate of 30,000 genes in the human genome, with a 1% difference from the chimpanzee genome, that suggests that about 150 mutations separate us from the chimpanzee-human common ancestor, with another 150 separating the chimpanzee from that ancestor. Many of the "human 150" have nothing to do with brain evolution, but if we guess that 50 are relevant we still have a lot to understand beyond the changes I posit from Stage S2 to Stage S6, and the supplementary changes discussed by Brinck.

[To change the subject: I spent 2 hours this afternoon (September 22) in a public debate with a Creationist. A daunting experience – about 70 of the 80 people in the audience were Creationists – but an interesting one. It seems hard for many non-scientists to realize that the fact that those of us seeking evolutionary insights into complex systems do not yet have all the answers does not imply that the theory of evolution (in general form) is wrong.]

Arbib, M.A. (2002) The Mirror System, Imitation, and the Evolution of Language, in Imitation in Animals and Artifacts, (Chrystopher Nehaniv and Kerstin Dautenhahn, Editors), The MIT Press, pp. 229 - 280.

Arbib, M.A. (2005) From Monkey-like Action Recognition to Human Language: An Evolutionary Framework for Neurolinguistics, Behavioral and Brain Sciences. (In press.)

Arbib, M., and Rizzolatti, G., 1997, Neural expectations: a possible evolutionary path from manual skills to language, Communication and Cognition, 29, 393-424.

Rizzolatti, G., and Arbib, M.A. (1998) Language Within Our Grasp, Trends in Neurosciences, 21(5):188-194.

Mirror neurons are not deus ex machina: I. The high stakes

Maxim Stamenov

Sep 24, 2004 13:20 UT

The discovery of Mirror Neurons (MNs) in monkeys unlike any other in the neurosciences stirred a whole flood of hypotheses and speculations in cognitive and social sciences. Attempts were made and scenarios were constructed that related, one or other way, MNs to all major specifically human capacities (for a

sample, cf. e.g., Stamenov & Gallese, 2002; Gallese, Keysers & Rizzolatti, 2004; more will undoubtedly find place in the comments to Arbib, 2005 in BBS). That a class of neurons in a certain model-dependent brain network, called sometimes Mirror Neurons System (MNS), is ascribed such a significance must already put the critical reader on a guard what is at stake. To the best of my knowledge, what is currently known about the way of performance of MNs does not support unequivocally any of the evolutionary scenarios offered for humans either for the origin of instrumental and/or social intelligence, or of imitation, or of language, or of empathy, etc. I think that it would be not a too gross exaggeration to claim that if there is a common denominator to all of the functions ascribed to MNs in all these contexts by all interpreters, it would boil down to that of serving as a deus ex machina in the explanation of 'specifically human capacity X', either in phylo- or/and in onto-genesis.

For example, one of the hypotheses taken sometimes bona fide as proven onboard is that MNS supports what is taken to be 'neonatal imitation' in humans. But is the neonatal 'imitation' a token of imitation? On critical examination it may turn out that it is as much imitation, as mimicry is role playing. What we currently have is a certain apparent analogy, but how far it goes and what it means remains to be put under closer scrutiny. On the other hand, I am not aware of experimental studies that have shown the possible involvement of MNs in the Broca's area of infants' brains during the cases of 'neonatal imitation' (please note that an activation in Broca's area would be per se not enough; it must be due to MNs in Broca). Both on conceptual and empirical side the case remains open. And here we are talking about a case where experimental verification is not out of question (and what to talk about language origin and evolution and the phylogenesis of the theory of mind mechanims).

But this is not the end of the story about MNs and their possible uses (due to the 700-words limitation). For continuation, please check the second discussion note on "The Grass Roots".

References

Gallese, V., K. Keysers & G. Rizzolatti. 2004. A unifying view about the basis of social cognition. Trends in Cognitive Sciences, 8(9), 396-403.

Stamenov, M. & V. Gallese (Eds.). 2002. Mirror Neurons and the Evolution of Brain and Language. Amsterdam & Philadelphia: John Benjamins. http://www.benjamins.com/cgi-bin/t_bookview.cgi?bookid=AiCR%2042

Neonatal “imitation” is not “real” imitation

Michael Arbib
Sep 25, 2004 3:16 UT

I agree with Stamenov’s caution that the literature on mirror neurons has extrapolated the concept far beyond the available data. While there has been much work on relevant brain imaging, the surprising fact is that neurophysiology of the macaque mirror system is pretty much confined to the Parma group.

Let me comment on one specific statement: “[Some people assert that] that MNS supports what is taken to be 'neonatal imitation' in humans. But is the neonatal 'imitation' a token of imitation?” Indeed, I would like to be careful about distinguishing the processes that are available to the child at different ages. I think that neonatal imitation is very different from the “complex imitation” that I consider in my article - where one recognizes that a novel action achieves a goal (and various subgoals) and comes to add that action to one’s own repertoire to achieve that goal oneself. Neonatal imitation is based on moving single effectors and thus differs from goal-directed imitation. (Studdert-Kennedy, 2002 discusses data consistent with the view that the infant at first imitates sounds by moving one articulator and only later coordinates articulators.)

Arbib (2005) touches all too briefly on a developmental sequence of models of the grasping system: ILGM (Oztop, Bradley & Arbib, 2004) -> MNS (Oztop & Arbib, 2002) -> “real” imitation.

However, it could be argued that the reciprocity seen in neonatal imitation is a necessary precursor for complex imitation, establishing that “I am like the other” (not, I think, “the other is like me” since that would unduly restrict the other’s capabilities) – and so can hope to do what they can do. Thus even while I would distinguish the two forms of imitation, I might agree that the biological evolution that selects for developmental pathways could well select for neonatal imitation as a basis for yielding a creature that has complex imitation available as a widely used tool for cultural transmission.

Arbib, M.A., 2005, From Monkey-like Action Recognition to Human Language: An Evolutionary Framework for Neurolinguistics, Behavioral and Brain Sciences, in press.

Oztop, E., and Arbib, M.A. (2002) Schema Design and Implementation of the Grasp-Related Mirror Neuron System, Biological Cybernetics, 87:116–140.

Oztop, E., Bradley, N.S., and Arbib, M.A., 2004, Infant grasp learning: A computational model Exp. Brain Res. 158: 480–503.

Studdert-Kennedy, M. 2002. Mirror neurons, vocal imitation and the evolution of particulate speech, in Mirror Neurons and the Evolution of Brain and Language, (M. Stamenov and V. Gallese, Eds.), Amsterdam/Philadelphia: John Benjamins, pp.207-227.

Neonatal "imitation" is not "real" imitation

Michael Arbib
Sep 25, 2004 3:16 UT

I agree with Stamenov's caution that the literature on mirror neurons has extrapolated the concept far beyond the available data. While there has been much work on relevant brain imaging, the surprising fact is that neurophysiology of the macaque mirror system is pretty much confined to the Parma group.

Let me comment on one specific statement: "[Some people assert that] that MNS supports what is taken to be 'neonatal imitation' in humans. But is the neonatal 'imitation' a token of imitation?" Indeed, I would like to be careful about distinguishing the processes that are available to the child at different ages. I think that neonatal imitation is very different from the "complex imitation" that I consider in my article - where one recognizes that a novel action achieves a goal (and various subgoals) and comes to add that action to one's own repertoire to achieve that goal oneself. Neonatal imitation is based on moving single effectors and thus differs from goal-directed imitation. (Studdert-Kennedy, 2002 discusses data consistent with the view that the infant at first imitates sounds by moving one articulator and only later coordinates articulators.)

Arbib (2005) touches all too briefly on a developmental sequence of models of the grasping system: ILGM (Oztop, Bradley & Arbib, 2004) -> MNS (Oztop & Arbib, 2002) -> "real" imitation.

However, it could be argued that the reciprocity seen in neonatal imitation is a necessary precursor for complex imitation, establishing that "I am like the other" (not, I think, "the other is like me" since that would unduly restrict the other's capabilities) – and so can hope to do what they can do. Thus even while I would distinguish the two forms of imitation, I might agree that the biological evolution that selects for developmental pathways could well select for neonatal imitation as a basis for yielding a creature that has complex imitation available as a widely used tool for cultural transmission.

Arbib, M.A., 2005, From Monkey-like Action Recognition to Human Language: An Evolutionary Framework for Neurolinguistics, Behavioral and Brain Sciences, in press.

Oztop, E., and Arbib, M.A. (2002) Schema Design and Implementation of the Grasp-Related Mirror Neuron System, Biological Cybernetics, 87:116–140.

Oztop, E., Bradley, N.S., and Arbib, M.A., 2004, Infant grasp learning: A computational model Exp. Brain Res. 158: 480–503.

Studdert-Kennedy, M. 2002. Mirror neurons, vocal imitation and the evolution of particulate speech, in Mirror Neurons and the Evolution of Brain and Language, (M. Stamenov and V. Gallese, Eds.), Amsterdam/Philadelphia: John Benjamins, pp.207-227.

Mirror neurons are not deus ex machina: II. The grass roots

Maxim Stamenov
Sep 24, 2004 14:18 UT

The most curious in the story about MNs, I think, is not the breadth of interpretations with high stakes it triggered per se but that on the basis of a discovery that is very predominantly documented how it is supposed to work on non-human primates, the cognitive and social scientists rushed to build scenarios how MNs made us different from them! At the same time, it remains debatable what is actually the function that MNS serves in macaques (the main group of experimental subjects) themselves: To attune to or to learn some rather specific class of object manipulation? To attune to or to learn conspecifics' way of behavior (and experience)? To trigger the process of social interaction through activation of a joint attention mechanism? Or something else, maybe? Nobody came to the idea to ask what are the evolutionary precursors of MNS in monkeys themselves (the only exception I am aware of is Hurford, 2004). Nobody also took care to ask her/himself is there a function MNs implement that we may potentially share with our ancestors up to the present day. In order to develop a balanced account (if we share a certain class of neurons or certain brain network in homologous brain areas) we must try for different, as well as same function(s). My own hypothesis (cf. Stamenov, in preparation) is that the primary function of MNS we share with our ancestors is that of extraction of fragments of implicit body-image from an inborn neural body-schema-toolkit. I mention

this alternative only in order to point out that one can attribute a function to MNS that is not related to any of the functions ascribed to it up to the present day on such a broad scale. And the real challenge is how to prove who has better chances to be on the right track. I don't think we will manage to convince each other about the function(s) of MNs in building alternative large-scale unified hypothetical scenarios.

One of the ways to start to build step-by-step a bridge between the 'grass roots' and the 'high stakes' of this topic, I think, is contained in two other ideas of Prof. Arbib – to investigate the possible relation of MNs in humans to schizophrenia and to autism (cf. the reference in his paper and Arbib & Yahya 2002 in his Internet web site) – which I find quite promising and appropriate. I would study here certain phenomena of body image fragmentation, voice hallucinations, echolalia and echopraxia that may clear the ground for better understanding of the potential involvement of MNS-like mechanisms in the acquisition and performance in norm and pathology of speech, actions and action sequences. I also think that the complementary exploration of the idea that we may still share certain function of MNS with our ancestors also should not be discarded without an attempt to verify it experimentally (and this would be a proof of function that we are talking about identical class of MNs in homological brain areas). Very promising are the current studies of Rizzolatti and colleagues (cf. Gallese, Keysers & Rizzolatti, 2004) that investigate the problem are there other MNS-like mechanisms for other purposes, e.g., for emotional attunement (that also becomes problematic in schizophrenia and autism), we share with our ancestors. In this context, constructing large-scale scenarios relating theory of mind, language evolution and neurophysiology (MNS) look to me premature – too many things must be taken for granted and/or managed via a whole cascade of hypotheses each of which remains to be proven on independent grounds.

For "Conclusion", please check the next, and last, Posting III.

References

Arbib, M. & J.S.M. Yahya. 2002. The "beyond-the-mirror" system and autism. http://www-hbp.usc.edu/people/arbib-papers.htm.

Gallese, V., K. Keysers & G. Rizzolatti. 2004. A unifying view about the basis of social cognition. Trends in Cognitive Sciences, 8(9), 396-403.

Hurford, J. 2004. Language beyond our grasp: What mirror neurons can, and cannot do, for language evolution. http://www.ling.ed.ac.uk/~jim/mirrormit.html.

Stamenov, M. (in preparation). Body schema, body image and mirror neurons. In: H. De Preester & V. Knockaert (Eds.), Body Image and Body Schema. Amsterdam & Philadelphia: John Benjamins.

From Feedback to Mental Disorders

Michael Arbib

Sep 25, 2004 3:36 UT

Stamenov says: "What is actually the function that MNS serves in macaques themselves? My own hypothesis (cf. Stamenov, in preparation) is that the primary function of MNS we share with our ancestors is that of extraction of fragments of implicit body-image from an inborn neural body-schema-toolkit." My own hypothesis is that the MNS for grasping evolved first to provide the subtle visual feedback needed for the control of dexterous hand movements. As such, it provided data on the relation of the moving hand to the affordances of the object to be grasped or manipulated rather than the retinal view of hand or goal object per se. I would then suggest that the ability to use this same machinery for recognizing hand-object relations when others were grasping ("I am like the other") was the basis for exaptation to the ability to recognize the actions of others. I think "attention to the hand" may be part of the "inborn neural body-schema-toolkit" but that the details of the grasp repertoire were not inborn but is the fruit of experience (cf. the models of development cited in my reply to Stamenov I).

Turning to the relation of mental disorders to the mirror system. Unfortunately, the study of autism with Salvador Marmol Yahya was never completed, but readers will find interest in the Commentaries by Williams and by Théoret & Fecteau on my BBS article. To these, I responded as follows: "Williams and Théoret & Fecteau see autism as providing a window on the role of the mirror system in ToM and language. (Théoret & Fecteau add analysis of blindness.) Deficits in autism are prominent in speech associated with social communication but praxic aspects of language are fairly well preserved. Perhaps what is affected is not so much language per se as the integration of this with affect and ToM. Interestingly, autistics may exhibit stereotypic mimicking (which monkey's don't have). Thus it must be [stressed] that a fully functional human mirror system inhibits mere repetition (echopraxia and echolalia) and instead relates the perception of perceived actions to the planning of an appropriate course of action."

I have also given some thought to the relation of the mirror system to schizophrenia Arbib & Mundhenk, 2004). The essay is built around the observation that the mirror system per se is posited to recognize an action, not the agent of an action – and so for action recognition to work properly, the neural representation of an action must be bound to the neural representation of the correct action.

Arbib, M.A., and Mundhenk, T.N. (2004) Schizophrenia and the Mirror System: An Essay, Neuropsychologia, in press.

Mirror neurons are not deus ex machina: III. The point
Maxim Stamenov
Sep 24, 2004 14:27 UT

In the previous two postings, I juxtaposed two approaches to MNs' study – top-down (= "high stakes") or bottom-up (= "grass roots"). In theory, everybody would agree that the best solution is to try to do it both ways at the same time and keep talking to each other. The problem with the "high stakes" in this case, however, is that they are so high (What makes humans unique as a biological species without making from them extraterrestrials?) that the use of such a relatively low brain 'variable' like MNs to justify human exclusiveness makes from them functional correlates of deus ex machina in the models and explanations. Those involved in experimental studies of MNS who work from "grass roots", on the other side, cannot use such highly general models in order to formulate experimentally verifiable hypotheses. Thus the possible joint venture of professionals using different methodologies (e.g., experimental vs. clinical single-case vs. descriptive vs. argumentative-interpretive) becomes quite implausible.

The fascination with MNs seems to me highly justified and their study has the potential of an integrative interdisciplinary program. The best service we (i.e., those that are not involved in experimental studies; I am a linguist) can offer to those that do experiments on the subject is to criticize their interpretations of the data and offer alternative ones that can lead to formulation of further or alternative verifiable/falsifiable hypotheses (on certain set of topics, as listed provisionally in the Posting II). Such an interdisciplinary multi-team collaboration looks to me currently much more plausible and effective on the "grass roots" basis. I put 'multi-team' because a personal or single-team research program capable of dealing top-down and bottom-up with MNs (as sketched) looks to me impossible to fund and manage.

Summary of the point of Postings I-III: The fascination with MNs seems to me highly justified and their study, under certain conditions, has the potential of an integrative interdisciplinary multi-team program. In such a program (if it comes to be realized), the interface between language and theory of mind in phylo- and ontogenesis and its neurophysiological basis will come as one of the last topics to be potentially considered due to its very high complexity and current lack of knowledge on many aspects of its implementation and way of functioning in brain/mind.

Bridging the Levels
Michael Arbib
Sep 25, 2004 3:50 UT

A relevant methodological advance is a technique for using computational models derived from primate neurophysiological data to predict and analyze the results of global data on brain and behavior (e.g. from PET and fMRI). Models tied to human brain imaging data often focus on a few "boxes" based on brain regions associated with exceptionally high blood flow, rather than analyzing the cooperative computation of multiple brain regions. For analysis directly at the level of such data, a schema-based model may be most appropriate. To further address neurophysiological data, the Synthetic PET imaging method (Arbib et al., 1994) uses neural models that are based on primate neurophysiology to predict and analyze results from PET human brain imaging taken during performance of a variety of behaviors. The problem is to find an integrated measure of activity in each simulated neural group which provides a predictor for the 3D volume of the image to which the neurons in this group correspond. The key hypothesis is that PET is correlated with the integrated synaptic activity in a region (Brownell et al., 1982), and thus reflects in part neural activity in regions afferent to the region studied, not just intrinsic neural activity in the region itself. Arbib, Fagg and Grafton (2003) show how the synthetic PET imaging approach was applied to a model of primate grasp control. The synthetic PET measures were computed for a simulated conditional/non-conditional grasping experiment using the FARS model (see Section 4.2), and then compared to the results of a similar human PET study (Grafton, Fagg and Arbib,1998). It further showed how the human PET results may be used to further constrain the computational model. Arbib et al. (2000) describe the possible extension of the Synthetic PET method to fMRI and provides an analysis of the imitation of motor skills, paying particular attention to data on the mirror system in monkey.

I thus propose an “evolutionary modeling strategy”: Develop models of monkey circuitry rooted in detailed neurophysiology and neuroanatomy; determine which regions of the human brain are homologous to which regions of the monkey brain (e.g., Arbib & Bota, 2003, 2004; Deacon, 2004); extend these homologies to generate models of human circuitry which will either strongly resemble the monkey model (high degree of homology) or be an informed variant of such a model (low degree of homology). In each case, the resultant models of interacting brain regions in the human are to be tested by Synthetic PET and Synthetic fMRI. The starting hypothesis is that the processing of a range of basic perceptual and motor strategies will be similar in monkeys and humans; that processes supporting imitation will be almost absent in monkey, rudimentary in chimp and well-developed in human; and that only a part of the mechanisms that support reflection and language in the human will be present in chimpanzees. A continuing challenge will be to chart the overlap and distinction between the relevant neural representations, to see how processes that operate on one representation “ground” the processes that act on “higher” representations, reflecting perceptual and motor schemas to linguistic schemas.

Arbib, M.A., Billard, A., Iacoboni, M., and Oztop, E. (2000) Synthetic Brain Imaging: Grasping, Mirror Neurons and Imitation, Neural Networks (13: 975-997.

Arbib, M.A., and Bota, M. (2003) Language Evolution: Neural Homologies and Neuroinformatics, Neural Networks 16:1237–1260.

Arbib, M.A., and Bota, M. (2004) Response to Deacon: Evolving mirror systems: homologies and the nature of neuroinformatics, Trends in Cognitive Sciences, 8:290-291.

Arbib, M.A., Bischoff, A., Fagg, A. H., and Grafton, S. T. (1994) Synthetic PET: Analyzing Large-Scale Properties of Neural Networks, Human Brain Mapping, 2:225-233.

Arbib, M.A., Fagg, A.H., and Grafton, S.T. (2003) Synthetic PET Imaging for Grasping: From Primate Neurophysiology to Human Behavior, in Exploratory Analysis and Data Modeling in Functional Neuroimaging, (F.T. Sommer and A. Wichert, Eds.), Cambridge MA: The MIT Press, pp.232-250.

Deacon, T. (2004) Monkey homologues of language areas: computing the ambiguities. Trends in Cognitive Sciences 8:288-290.

Grafton, S. T., Fagg, A. H., & Arbib, M. A. (1998) Dorsal Premotor Cortex and Conditional Movement Selection: A PET Functional Mapping Study. Journal of Neurophysiology 79:1092-1097

Bridging the Levels

Michael Arbib

Sep 25, 2004 3:50 UT

A relevant methodological advance is a technique for using computational models derived from primate neurophysiological data to predict and analyze the results of global data on brain and behavior (e.g. from PET and fMRI). Models tied to human brain imaging data often focus on a few “boxes” based on brain regions associated with exceptionally high blood flow, rather than analyzing the cooperative computation of multiple brain regions. For analysis directly at the level of such data, a schema-based model may be most appropriate. To further address neurophysiological data, the Synthetic PET imaging method (Arbib et al., 1994) uses neural models that are based on primate neurophysiology to predict and analyze results from PET human brain imaging taken during performance of a variety of behaviors. The problem is to find an integrated measure of activity in each simulated neural group which provides a predictor for the 3D volume of the image to which the neurons in this group correspond. The key hypothesis is that PET is correlated with the integrated synaptic activity in a region (Brownell et al., 1982), and thus reflects in part neural activity in regions afferent to the region studied, not just intrinsic neural activity in the region itself. Arbib, Fagg and Grafton (2003) show how the synthetic PET imaging approach was applied to a model of primate grasp control. The synthetic PET measures were computed for a simulated conditional/non-conditional grasping experiment using the FARS model (see Section 4.2), and then compared to the results of a similar human PET study (Grafton, Fagg and Arbib,1998). It further showed how the human PET results may be used to further constrain the computational model. Arbib et al. (2000) describe the possible extension of the Synthetic PET method to fMRI and provides an analysis of the imitation of motor skills, paying particular attention to data on the mirror system in monkey.

I thus propose an “evolutionary modeling strategy”: Develop models of monkey circuitry rooted in detailed neurophysiology and neuroanatomy; determine which regions of the human brain are homologous to which regions of the monkey brain (e.g., Arbib & Bota, 2003, 2004; Deacon, 2004); extend these homologies to generate models of human circuitry which will either strongly resemble the monkey model (high degree of homology) or be an informed variant of such a model (low degree of

homology). In each case, the resultant models of interacting brain regions in the human are to be tested by Synthetic PET and Synthetic fMRI. The starting hypothesis is that the processing of a range of basic perceptual and motor strategies will be similar in monkeys and humans; that processes supporting imitation will be almost absent in monkey, rudimentary in chimp and well-developed in human; and that only a part of the mechanisms that support reflection and language in the human will be present in chimpanzees. A continuing challenge will be to chart the overlap and distinction between the relevant neural representations, to see how processes that operate on one representation "ground" the processes that act on "higher" representations, reflecting perceptual and motor schemas to linguistic schemas.

Arbib, M.A., Billard, A., Iacoboni, M., and Oztop, E. (2000) Synthetic Brain Imaging: Grasping, Mirror Neurons and Imitation, Neural Networks (13: 975-997.

Arbib, M.A., and Bota, M. (2003) Language Evolution: Neural Homologies and Neuroinformatics, Neural Networks 16:1237–1260.

Arbib, M.A., and Bota, M. (2004) Response to Deacon: Evolving mirror systems: homologies and the nature of neuroinformatics, Trends in Cognitive Sciences, 8:290-291.

Arbib, M.A., Bischoff, A., Fagg, A. H., and Grafton, S. T. (1994) Synthetic PET: Analyzing Large-Scale Properties of Neural Networks, Human Brain Mapping, 2:225-233.

Arbib, M.A., Fagg, A.H., and Grafton, S.T. (2003) Synthetic PET Imaging for Grasping: From Primate Neurophysiology to Human Behavior, in Exploratory Analysis and Data Modeling in Functional Neuroimaging, (F.T. Sommer and A. Wichert, Eds.), Cambridge MA: The MIT Press, pp.232-250.

Deacon, T. (2004) Monkey homologues of language areas: computing the ambiguities. Trends in Cognitive Sciences 8:288-290.

Grafton, S. T., Fagg, A. H., & Arbib, M. A. (1998) Dorsal Premotor Cortex and Conditional Movement Selection: A PET Functional Mapping Study. Journal of Neurophysiology 79:1092-1097

The role of unconventional communication

Cristiano Castelfranchi
Sep 27, 2004 13:51 UT

An important notion seems missing not only in Arbib's quite strong and very challenging picture of the origin of language, but in general in our reasoning about stages and conditions involved in this process. Moreover, I think that Arbib uses a too vague notion of 'protosign' and manual-based 'protolanguage', maintaining at the same time a too limited notion of communication.

In my view, we should reason about the origin of language and the origin of intentional communication keeping in mind a clear notion of a very basic form of communication: *when mere practical behavior, simple effective actions are used for communicating, without and before specialized signs or 'gestures', without and before ritualized forms, or simulation and mimicry; the action of x has the goal (either the intention of x or the function) of being detected by y and meaning something to him.* I call this form of communication Behavioral Implicit Communication not to be mixed up with Non Verbal Communication that refers to a repertoire (a 'lexicon') of inborn or culturally specialized signs for communicating.

The presupposition of Communication is 'Signification': the fact that y perceives a given event (smoke) as the sign of something else (fire). To have Communication it is necessary for the sign to be sent on purpose by x to y in order y catches its meaning. Obviously, for human beings other people's behaviors are very meaningful, and they recognize and 'read' those behaviors in a specific way; but not all of them are 'messages' and 'communication'. Mirror neurons provide certain species with the ability to 'read' the behavior of the other, recognizing the intended action and even its goal and intention. The real –and yet missing – issue is that *x can realize and foresee this ability of y of 'understanding' his practical behavior and might rely on it. This is one of the bases of human cooperation*. This form of communication is rarely focused and disentangled from NVC, gesture language, etc. For example, I question Arbib's thesis that intended communication derives from pantomime and symbolization when the action has no longer praxis goals but is only for a communicative use. Long before this stage the agent is able to and intends to act in order the other gets the message of his behavior, and even understands that x intends to communicate.

It seems presupposed in Arbib's analysis some sort of identification between communication systems and conventional signs systems. I consider this view misleading. If we focus on the origin of 'language' we risk to be simplistic as for the origin of 'communication' and in particular of intentional communication. Gestures come from action because actions assume a communicative purpose. And this seems due to mirror neurons:

mirror neurons are crucial for y to recognize x's action and understand x's intentions; but perhaps also for the second crucial step: that x understands that y is looking at and seeing his action, and that y's is able to recognize and understand it. This can be one of the meanings and messages for x of y's imitation of his action; but also an appropriate reaction of y can have the same meaning. For example, I think that the appropriate reaction of the mother (not necessarily her imitation) means to the baby that she understood his intention to reach and to have the 'pointed' object. To summarize: There is a gap in Arbib's explanation since it arrives directly to Protosign via simple and complex imitation, ignoring the non-conventional use of action for communication which exploits the mirror-neurons-based ability of understanding the other's behavior. In particular, what is missing is the idea that communication evolves - thanks to mirror neurons - in parallel to imitation faculty for converging later on a conventional/intentional gesture languages. The missing path is Behavioral Implicit Communication, its later de-contextualization and conventionalization that produces part of the gesture 'vocabulary'. However, I think that Arbib is right in stressing the importance of pantomime in relation to the very important step of creating signs about entities (for example the difference between to telephone and the telephone), which is one of the major problem in the creation of a gesture language implying the differentiation of verbs and nouns.

Protolanguage and Theory of Mind - Chicken and Egg, or Two Chickens?

Michael Arbib
Oct 5, 2004 1:04 UT

I thank Castelfranchi for has made a crucial observation for the overall theme of this Discussion, Coevolution of Language and Theory of Mind. He observes that:

"To have Communication it is necessary for the sign to be sent on purpose by x to y in order y catches its meaning. Obviously, for human beings other people's behaviors are very meaningful, and they recognize and 'read' those behaviors in a specific way; but not all of them are 'messages' and 'communication'. I [thus] question Arbib's thesis that intended communication derives from pantomime and symbolization when the action has no longer praxis goals but is only for a communicative use. Long before this stage the agent is able to and intends to act in order the other gets the message of his behavior, and even understands that x intends to communicate."

and then argues that

"mirror neurons are crucial for y to recognize x's action and understand x's intentions; but perhaps also for the second crucial step: that x understands that y is looking at and seeing his action, and that y's is able to recognize and understand it."

This would be consistent with those accounts that link mirror neurons to Theory of Mind, and the general tenor of accounts of Machiavellian Intelligence (Byrne & Whiten, 1988) in which, e.g., deception is the use of an action for its effect on another rather than its immediate outcome in praxis. In Arbib (2005), I offered the following "placeholder" when discussing the transition from recognition of praxic movements to pantomime:

"All this assumes, rather than provides an explanation, for LR4, the transition from making praxic movement, e.g., those involved in the immediate satisfaction of some appetitive or aversive goal, to those intended by the utterer to have a particular effect on the recipient. As a placeholder, let me offer the following:

"The Intended Communication Hypothesis: The ability to imitate combined with the ability to observe the effect of such imitation on conspecifics to support a migration of closed species-specific gestures supported by other brain regions to become the core of an open class of communicative gestures.

"Darwin [1872/1965] observed long ago, across a far wider range of mammalian species than just the primates, that the facial expressions of conspecifics provide valuable cues to their likely reaction to certain courses of behavior (a rich complex summarized as "emotional state"). Moreover, the F5 region contains orofacial cells as well as manual cells. I thus posit a progression from control of emotional expression by systems that exclude F5 to the extension of F5's mirror capacity from manual to orofacial movement and then, via its posited capacity (achieved by stage S3) for simple imitation, to support the imitation of emotional expressions. This would then provide the ability to affect the behavior of others by, e.g., appearing angry. This would in turn provide the evolutionary opportunity to generalize the ability of F5 activity to affect the behavior of conspecifics from species-specific vocalizations to a general ability to use the imitation of behavior (as distinct from praxic behavior itself) as a means to influence others. This in turn makes possible reciprocity by a process of backward chaining where the influence is not so much on the praxis of the other as on the exchange of information. With this, the transition described by LR4 (intended communication) has been achieved in tandem with the achievement and increasing sophistication of LR2 (Symbolization)."

When Castelfranchi says “communication evolves - thanks to mirror neurons - in parallel to imitation faculty for converging later on a conventional/intentional gesture [proto]language,” he helps us see how to refine my just quoted argument in addressing the issue of intention in communication. Note that I have added “[proto]” too the quote. Even though Castelfranchi thinks I use “a too vague notion of ‘protosign’ and manual-based ‘protolanguage’,” I believe the distinction between language and protolanguage is worth preserving.

Arbib, M.A., 2005, From Monkey-like Action Recognition to Human Language: An Evolutionary Framework for Neurolinguistics, Behavioral and Brain Sciences. (In press.)

Byrne R. & Whiten A. (Eds.), 1988, Machiavellian Intelligence. Social Expertise and the Evolution of Intellect in Monkeys, Apes and Humans, Oxford, Clarendon Press.

Darwin, C. (1872/1965) The expression of the emotions in man and animals. Chicago: University of Chicago Press.

NVC to BIC

Chris Lofting
Oct 5, 2004 2:10 UT

Cristiano Castelfranchi writes:

"In my view, we should reason about the origin of language and the origin of intentional communication keeping in mind a clear notion of a very basic form of communication: when mere practical behavior, simple effective actions are used for communicating, without and before specialized signs or ‘gestures’, without and before ritualized forms, or simulation and mimicry; the action of x has the goal (either the intention of x or the function) of being detected by y and meaning something to him. I call this form of communication Behavioral Implicit Communication not to be mixed up with Non Verbal Communication that refers to a repertoire (a ‘lexicon’) of inborn or culturally specialized signs for communicating."

BIC is dependent on qualities derived from NVC where those qualities, serving as GENERALS, can be particularised through association with a particular context (and so action). The commonest NVC is found in the form of expressions derived from emotional reactions to context but these in turn appear to have a more generic source.

BENEATH the emotions we find a vaguer realm, focused on dealing with context in the form of (a) replacing context with something 'better' or (b) coexisting with the current context.

(a) reflects the activities of identity assertion through replication (develops into sexual love) or erradication (develops into anger). We are dealing here with DIFFERENTIATING, and so pushing AWAY others to assert self.

(b) reflects the activities of recruiting the context to aid in assertion of identity; and from there seems to emerge a focus on group, social, dynamics. We are dealing here with INTEGRATING, a so pulling together others to assert self.

the development path of language, as in Arbib's focus on the spoken/written word, is from a set of qualities hard coded in ALL neuron dependent life forms. If the METHOD used to derive information is X then the PRODUCTS of that method become the sources of qualities used to determine the meaningful from the meaningless.

As such the METHOD acts as a bound within which all meaning is processed. Within that boundary feedback allows for increase in precision through refinements in differentiations and that includes forms of EXPRESSION - be it verbal, visual, olfactory, gustatory, or kinesthetic.

The verbal nature reflects the discretisation of, the digitalisation of, analogue processes and as such an overall dynamic in the brain of analogue-to-digital conversion - as reflected in the passage of data from dendrites to axon in the neuron, then extended into neural networks and 'up' into general expressions of mind and out into social dynamics that in turn feedback in to the neurology.

This conversion process takes the WHOLE in the form of immediate sensations and stretches into a sequence of PARTS that can be analysed, refined, and re-transmitted. This process also allows for the immediate reaction to stimulus, i.e. instincts/habits, to be refined, made more context-sensitive and so increase choices in expression.

From these 'basics' come the LOCAL factor, as in mode of expression. One of these is reflected in 'flocking' behaviour where individuals making LOCAL distinctions feed into an overall behaviour of the group WITHOUT INTENT. That expression of the group falls within the bounds of the method used to derive meaning and as such can be interpreted as vaguely 'meaningful' and so instigate activities that

can be ritualistic; there is no understanding of how the original pattern emerged other than what one was doing LOCALLY and so that is copied such that it may elicit the pattern again. BIC as such has an 'intent free' element that can lead to ritualisation. That intent-free state is rooted in the dynamics of differentiating/integrating applied to individual actions in a group context.

Chris.

Pidgin Creoles and the cultural evolution of language
Viviane Deprez
Oct 7, 2004 22:45 UT

In questioning the plausibility of a cultural evolution for language, Reboul noted that the emergence of Creoles from Pidgin was problematic for this view. In his reply, Arbib dismissed the problem by 1) raising doubts about the credibility of Bickerton's original bio-program proposal. 2) asserting that grammatical features of Creoles mostly resemble that of their substrate (or superstrate) languages.

I was rather inclined to object to Arbib's rendering of Bickerton's original ideas on Pidgins and Creoles. However, since he attributes them to Denis and not Derek the latter's reputation is safe. Arbib's (mis)interpretation that children could import a whole grammar all at once into a pidgin 'from their genomically encoded Universal Grammar' makes you wonder how a genomically encoded grammar could ever be missing from the Pidgin in the first place since obviously, the parents being humans in all respects, must have had whatever genome Arbib presumes UG could be encoded on. Rather than misremembering Denis' old views, it might be refreshing to look at Derek's current views expressed in his recent book with W.Calvin (Lingua ex Machina). There the distinction is reformulated in terms of Proto-language vs. language, and Pidgins are assimilated to Proto-languages in their forms. That there can be humans (the parents) with both language (their own) and a Proto-language is not surprising. Not all communication systems human use in their lifetime have to be language like. But a normal human child with only Proto-language would be a clear evolutionary anomaly. If the child's only linguistic input were a pidgin, surely a controversial assumption in the case of any plantation Creoles (but perhaps less so for the sign language Creole of Nicaraga), then the evolutionary normal child would have no choice but to turn the Pidgin into a full fledged human language, a feat accomplished with whatever means are available to her, including her LAD. According to Bickerton, this involves the imposition of structure, or in linguistic terminology Merge, the recursive function that builds constituents. Note in particular that the words are already there in the Pidgin, all the child has to do is to impose some hierarchical UG compatible structure on them.

Now to come to Arbib's observation that "There seems to be good evidence the grammatical structure of a Creole will bear a strong resemblance (albeit simplified) to one or more of the languages from which it comes", this is quite expected from a UG point of view. If Creoles are human languages (which no one will deny), then obviously they must have grammars that are within the bounds set by UG, just like their substrate and superstrate languages do. Basic grammatical features of Creoles can however differ from both substrate and superstrate languages. Consider for instance word order. It seems rather firmly established that Creoles commonly have a basic SVO order. This could well result from the fact that most of them have either a substrate or a superstrate language with this basic word order. Some particularly interesting cases of Creoles, however, like Berbice Dutch Creole, clearly show that this cannot be the end of the story. In this case, clear evidence exist that both substrate and superstrate languages are SOV unlike Berbice Dutch Creole which is SVO. Such cases, evidently, raise trouble for a cultural inheritance point of view, not for UG, and I do not see how grammaticalization (mostly a name for a problem rather than a solution) could even begin to help with this case. Hence, it would seem, there are established cases for which Reboul's objection clearly stands.

Putting all the strands together. Coevolution of language and theory of mind

Peter Ford F. Dominey (CNRS - Institut des Sciences Cognitives, Lyon) and
Anne Reboul (CNRS - Institut des Sciences Cognitives, Lyon)
(Date of publication: 12 October 2004)

Abstract: We present an outline of the ideas and discussions throughout the web conference, to highlight the points of agreement and disagreement which have emerged and point out perspectives for further thoughts. Some obvious strands are the relation between language and theory of mind as tested by false belief, the kind of primitive theory of mind that might be present at the first stages of language acquisition and the impact of developing language on developing theory of mind and vice versa. On the evolution side, a major question is how far does ontogeny recapitulate philogeny. In other words, can we rely on what we know about language acquisition and theory of mind development to surmise what happened when language and theory of mind evolved? Can the contribution of studies of animal communication help us answer this question?

It is not easy to close a lively conference such as the present one, doing justice to all the contributors, to the discussions and trying to establish general strands without entirely losing sight of the details. We have decided to try and do it through the general following organisation: we will begin by a short review of the aims of the conference, followed by a short precis of each contribution in the order of their publication, trying to take into account the clarifications and precisions that the discussions following them have brought, then turning to an attempt at a synthesis, aiming at discovering general points of agreement as well as controversial main issues, indulging in some speculative thinking,_and, finally, pointing out a few possible perspectives regarding both theoretical and experimental or modelling issues.

The aim of the conference

The project for the conference came from the fact that both of us were engaged in an ESF Eurocores project on the evolution of language and had — possibly different but strong in any case — skepticisms about popular theories. For instance, quite a few theories have hypothesized that language had emerged from social pressures and that its emergence was made possible by previous mind reading or theory of mind skills. However, such hypotheses never made clear what sort of theory of mind was a necessary condition for language emergence (though Tomasello and colleagues have made a serious effort at spelling out the intention reading prerequisites). Similar claims _were made for the precedence of language acquisition. However, the chronological evidence in both theory of mind development and language acquisition did not seem to agree with such hypotheses in their rough and ready form and it seemed clear that some precision would be needed. A possibility, which some at least of the contributors have followed is that language and theory of mind have coevolved, which of course does not preclude that one or both of them should be necessary for the normal acquisition of the other. Thus, the conference.

Individual contributions

Introducing the conference, Reboul tried to shift through the evidence on both theory of mind and language, using both acquisition and comparative data. Her main hypotheses are that language and ToM are specific to human animals, that ToM is not a monolith but rather a collection of possibly modular abilities, that the acquisition of the lexicon, though it may be helped by relatively elementary mind reading mechanisms, begins without them and, finally, that success at the false belief task may well depend on abilities acquired through language rather than otherwise. In an effort to draw a parallel between the underlying mechanisms of language and ToM, Dominey outlined a mechanism for ToM constructions, _as mappings between behavioral scenario structure and the corresponding social/intentional interpretation or outcome (analogous to the form-meaning mappings in grammatical constructions). The ensuing debate identified the point that aspects of ToM (notably the ability to metarepresent) are semantic abilities, likely distinct from the structure mapping capability. Newmeyer goes on to puzzle out the different contributions of cognition-aiding factors and vocal interaction-aiding factors in the origins and evolution of grammar. He concludes that cognitive factors were the first to shape grammars. But with the passage of time, the exigencies of communication came to play an ever-more important role in grammar, and thus human language today therefore reflects _the influence of both types of factors. While Newmeyer concentrates on the evolution of language, Baron-Cohen concentrates on empathy, which he sees as having both a cognitive element (ToM, illustrated by the false belief test) and a more emotional aspect. _Empathy seems to be stronger in females than in males and

the ability to identify speaker's meaning is an strong advantage in language acquisition, explaining why females are more precocious in that regard than males. Origgi and Sperber acknowledge that ToM comes in different and probably successive modular abilities (they are skeptical of the false belief test as a test of the kind of mindreading abilities crucial for language acquisition or evolution), and see it as a necessary condition for the evolution of the language acquisition device, given that it would enable communicators to go beyond the restrictions of an eventual protolanguage to more or less propositional forms, corresponding to the speaker's meaning. As they point out, this is something impossible on a code model of linguistic communication and their view (following Relevance Theory) is that linguistic communication rests on both code and inference, inference being done through a special purpose pragmatic module. Fitch adopts a comparative perspective, pointing to three capacities, which are both fundamental for language and specific to humans as opposed to chimpanzees: speech and especially vocal imitation, recursive syntax and intentional semantic communication. He argues that vocal imitation amounts to non-semantic sensorimotor quoting from which semantic representations can develop and advocates a "coevolutionary relationship between quoting and theory of mind", additionally and strongly conditioned by syntactic recursion. Jill de Villiers _outlines an hypothesis, backed by experimental data, according to which ToM (specifically the false belief test), far from conditioning language acquisition or evolution, is on the contrary dependent on language and, more precisely, on mastery of the recursive structure of complementation evidenced by discourse and mental verbs (e.g., *saying*, *thinking*). According to her, such structures were first developed to express the reliability of signals and of the signalers who produce them. Ruffman argues that children's initial understanding of mind is likely implicit and manifest in their behavior rather than insights that they can verbalize. This understanding could be innate, but learning likely plays a large role. Over time, and as their language develops, children develop a consciously mediated and verbally based theory on the basis of these implicit intuitions. Diesendruck concentrates on the acquisition of words and, without denying that there may several roads to word learning, claims that only one of them, which goes through mindreading abilities, is going to provide the learner with the kind of words which will enable her to take part in normal linguistic communication. In other words, though very young children, trained animals and, possibly, autistic speakers learn words through a training process of a probably associative nature, only the acquisition of words through (possibly fairly elementary) mindreading abilities is going to provide an individual with the opportunity for normal communication. Nazir and Boulenger consider the problem from a perspective of evolutionary pressure to minimize processing requirements. In this context, infants are strongly predisposed to seek for optimal solutions to minimize processing requirements, and words provide a perfect clue allowing them to rapidly form categories, including those related to theory of mind. Arbib outlines a motor-oriented account of language-readiness, based on successive capabilities to recognize and imitate action, leading to protosign, protospeech and language. He identifies the need to investigate whether an account can be given of a shared evolution of "mirror systems" suiting both ToM and complex imitation, hypothesizing that the ancestral mirror system for manual praxis was distinct from the putative mirror system for facial expression of emotion. The former would support pantomime and thence on to multi-modal symbols; and then the availability of symbols could enrich the latter to yield rudiments of ToM.

As can be seen from the previous quick summing up of the various contributions to the conference, the positions taken seem fairly widespread. However, we think that some major points of agreement emerged during the conference, among other things during the discussions, and that differing positions can indicate some interesting areas for experimental and or theoretical debate. It is to them that we now turn.

Putting all the strands together

There are a few questions that naturally come to mind on the topic of the coevolution of language and theory of mind. Here are some of them: Are either language or theory of mind as such the product of evolution and, if so, what are their adaptive features? Are either language or theory of mind monolithic abilities (i.e., should one consider either of them a module in a Fodorian sense)? If either of them is not, how should it be divided and should its resulting components be regarded as modular in the Fodorian sense? And, perhaps the central questions in the present conference, have language and theory of mind, supposing them to have evolved, coevolved? If not, is there a relation of dependence between language and theory of mind and if there is, in which direction does it go?

Let us begin with the question regarding whether or not language and theory of mind are monolithic abilities. Clearly, as regards language - especially if this question is linked to the evolution of language -, one should presumably distinguish between phonology and syntax (including the part of semantics that is heavily dependent on syntax) on the one hand and the lexicon on the other. For exemple, it could make sense to say that syntax has evolved and is more or less innate, but it clearly does not make sense to say that the lexicon as such has evolved and is innate.

This is why it makes sense to speak of a language acquisition device being innate, noting that this does not include the lexicon, which is language-specific. This does not exclude the parallel evolution of general mechanisms such as those described by Boulenger and Nazir, where a new word triggers the acquisition of a new concept or category nor the general role of mindreading in lexical acquisition, as described by

Diesendruck. Regarding this, some contributors (i.e., Origgi and Sperber, de Villiers) clearly consider that language, in more or less the above sense, has evolved as such, while others do not, either because they do not pronounce on the issue or because they think that language did not evolve as such, i.e., some linguistic abilities have evolved but they did not evolve as linguistic abilities (Dominey, Arbib, Fitch, Newmeyer). Regarding theory of mind, there seems to be a more or less general consensus among the contributors who deal with it, that it presumably is not a monolithic ability, though there may be disagreement over what specific distinct abilities make it (i.e., over how we should cut the cake) as well as over whether these distinct abilities are or not modular in the Fodorian sense. Note as well that some of these abilities might be modular while others might not be. None of the contributors have advocated a new division in ToM abilities, though quite a few have taken up Baron-Cohen's suggestions (1995) regarding the existence of an eye direction detection (EDD) and an intentionality detection (ID), followed by a shared attention mechanism (SAM) and, much later by a Theory of Mind Mechanism (ToMM), which enables children to pass the false belief test at around four years of age.

Regarding the relation of language and theory of mind, some contributors (e.g., Origgi and Sperber) have taken the strong view that some of the modules making up ToM are a prerequisite for the evolution of a language acquisition device (LAD). On this scenario, both ToM and LAD have evolved, but the evolution of ToM, in a possibily fairly elementary form, must precede the evolution of LAD, being a necessary (though not sufficient) condition for it. This view is apparently contradicted by de Villiers who claims that language came first and was indeed the condition for ToM (though she concentrates on false belief test). This seems echoed, up to a point, by Fitch, who also points out the intimate relation between ToM and recursive syntax, though he might favor a more coevolutionary scenario, given his insistence on imitation. A more balanced view is proposed by Diesendruck who insists on the necessity of ToM for lexicon acquisition, but allows for two different learning processes (possibly successive in normal development), only one of which would allow the learner to successfully communicate through language. On Diesendruck's view, some ToM abilities, which are roughly in place by two years of age (note that these might include SAM, on Baron-Cohen's division of ToM abilities), spur language acquisition on. These would be enough for normal language acquisition, though they would still fall short of passing the false belief test. On this count, Diesendruck's and de Villiers' view are not as incompatible as they seem: granted Diesendruck says ToM first, while de Villiers says language first, but they do not mean exactly the same thing by ToM. Diesendruck does not include false belief ascriptions (FBA) in the ToM abilities that are taken to precede language, while de Villiers is only interested in FBA. This might lead to a mixed model under which, though some pretty limited lexicon acquisition would take place before 18 months of age, ToM modules (not including FBA) would then kick in, spurring language acquisition on and allowing normal linguistic communication; language acquisition would proceed, allowing the child to develop more sophisticated and highly language dependent ToM abilities, such as FBA (allowing her to pass the false belief test) and, ultimately, the ability to detect opaque contexts (allowing her to pass, about a year later, _the opaque context tests). By the way, it should be clear that such a scenario might ultimately solve the ToM as theory versus ToM as non-theoretical (either emotional empathy or simulation) debate by going more or less in the direction indicated by Baron-Cohen (2003) through his distinction between the affective and the cognitive components of empathy: one could suppose that the affective (non-theoretical) components would more likely be innate, though susceptible of degrees (hence, the <u>average</u> differences which he describes between males and females), and would condition language acquisition, while the cognitive components (as tested through false belief and opaque contexts tests) would be language dependent and learned (and theoretical).

This scenario makes sense if envisaged relative to Newmeyer's insistence on the importance of cognition for shaping language. On Newmeyer's scenario, the fundamental structure of language is determined by cognitive constraints (though it should be noted that his view is compatible with a concepts first and language as a means of expressing them perspective), though some linguistic variation should be accounted for through communicative constraints. This might lead (introducing ToM, which Newmeyer does not take into account _and giving to language not only cognitive constraints, but a cognitive role) to a view in which language first evolved as a cognitive tool, to organize information (a view very near in some respects to Nowak's and his team (2000, 2001, 2002 - though one would not necessarily accept Nowak's self-organisational outlook), and only then evolved as a communicative medium. In that scenario, the advantage for which language first evolved was cognitive, not communicative, though the size and social complexity of hominoid and early human groups (as well as the long period during which children are dependent on adults, see Baron-Cohen 2003) led to the evolution of mindreading (empathy) devices, facilitating language transition from a cognitive to a communicative tool, _which in turn led to more sophisticated ToM, leading to FBA and ultimately to the detection of opaque contexts and allowing communicative constraints to slightly change the previous structure of language. This would explain both the syntactic difference between language and animal systems of communication, but also its communicative differences with them, well described by Fitch: animals do not seem to have recursive structures in their communicative system neither do they have anything like Gricean cooperation in their communicative use of these systems. It would, as well, explain why linguistic communication cannot be made sense of through a code-only perspective and why a mixed view (both code and inference) is needed. Finally, to go further, if language primarily emerged

(either through evolution, self-organization or exaptation from systems evolved for other reasons) as a cognitive tool, and if it emerged with the *Sapiens Sapiens* lineage, it might explain why that lineage had better evolutive success than did the Neanderthals with whom it co-habited for some seventy thousand years (from the apparition of Sapiens about a hundred thousand years ago to the disparition of Neanderthals about thirty thousand years ago) and may go some way to explain the so-called cultural explosion which occurred some forty to fifty thousand years ago. This is mysterious, because *Sapiens Sapiens* does not seem to have undergone biological (and especially cerebral) evolution since its first apparition. Thus, the big question is: why did it take *Sapiens Sapiens* some fifty thousand of years to radically modify its behavior, inventing art and dramatically improving its technology? One frequent answer is through language. This however would be in direct contradiction with the notion that language is (the product of) a biological evolution. But innovations are in some ways very cultural human manifestations - a fact acknowledged by Newton when he said "If I have seen further it is by standing on the shoulders of giants" - and one very obvious means of cultural transmission is through communicative language. A way to reconcile the idea that language may be a biological evolution is to see *Sapiens Sapiens* as possessing a cognitive language (possibly inherited from a common ancestor with Neanderthals), but going on to use it communicatively, through both better empathy and a radically different vocal tract. On this view, language *qua* cognitive tool would be a biological evolution, but linguistic communication may have emerged culturally. And one might go the way of Baron-Cohen and beyond here and speculate that the transition of language from cognitive tool to communicative tool might have been very much a female affair…

The conference has raised several points that will be of practical use in the study of ToM and language via simulation and robotic experiments.
One of these concerns the language and ToM capabilities that are employed in code vs inferential communication. Several authors indicate that linguistic communication is impossible at the code level - Diesendruk, Origgi and Sperber -, though it may be that initially — both for development and evolution — the system begins in a state that is more code level (or associative) than inferential — Reboul, Diesendruck. This suggests a point of departure for simulation to begin with a code level system, and then to introduce a pragmatic function that may either be an innate "module" - Sperber - or something that can be at least in part developped - Ruffman. In this context de Villier's suggests that recognition of one's own false beliefs may provide the bootstrap for a more generalized ability to attribute mental states in a recursive manner. Though de Villiers certainly does not endorse construction based learning mechanisms, it should be noted that her position on ToM (and FBA) is consistent with the construction based learning mechanism proposed by Dominey and the statistical learning mechanisms cited by Ruffman. The scenario would then be to consider systems with an initial representation of the own and other's goals and intentions, and a code level communication capability, with the potential to make the transition to communication itself becoming part of the action, goal, intention system.

Conclusion

In conclusion, as co-organisers of the present conference, we would like to thank all of the contributors and discussants for having made it lively, intellectually stimulating and great fun. We can only hope that all the participants have enjoyed it as much as we have and that we have not made any major error in the above summary of contributors' positions.
Special thanks go to Gloria Origgi, who, as administrator of the web conference, has made this event possible and is, as such, in great part responsible for its success.

References

Baron-Cohen, S. (1995), *Mindblindness: An essay on autism and theory of mind*, Cambrige, MA, The MIT Press.
Baron-Cohen, S. (2003), *The essential difference*, London, Penguin Books.
Fitch, T. and Hauser, M. (2004), "Computational constraints on syntactic processing in a nonhuman primate", *Science 303,* 377-380.
Nowak, M.A. and Komarova, N.L. (2001), "Towards an evolutionary theory of language", *Trends in Cognitive Science 5/7*, 288-295.
Nowak, M.A., Plotkin, J.B. and Jansen, V.A.A. (2000), "The evolution of syntactic communication", *Nature 404*, 495-498.

Nowak, M.A., Komarova, N.L. and Nyogi, P. (2002), "Computational and evolutionary aspects of language", *Nature 417*, 611-617.