Language as a Tool for Interacting Minds

KRISTIAN TYLÉN, ETHAN WEED, MIKKEL WALLENTIN, ANDREAS ROEPSTORFF AND CHRIS D. FRITH

Abstract: What is the role of language in social interaction? What does language bring to social encounters? We argue that language can be conceived of as a tool for interacting minds, enabling especially effective and flexible forms of social coordination, perspective-taking and joint action. In a review of evidence from a broad range of disciplines, we pursue elaborations of the language-as-a-tool metaphor, exploring four ways in which language is employed in facilitation of social interaction. We argue that language dramatically extends the possibility-space for interaction, facilitates the profiling and navigation of joint attentional scenes, enables the sharing of situation models and action plans, and mediates the cultural shaping of interacting minds.

1. Introduction

Sometimes work, vacation, or love take you to another corner of the world where people speak a language unknown to you. Suddenly, you are reminded how linguistic communication pervades even our simplest everyday activities and practices, like navigating downtown or buying a cup of coffee. In some face-to-face situations, you might actually get along pretty well with pointing gestures and other types of non-linguistic communication, though it might still take some effort to reach a common understanding. But once you find yourself depending on more mediated forms of linguistic communication (e.g. written instructions, menu cards or telephone calls) you are in big trouble. You are missing one of the most important tools for interacting with other minds.

The tool perspective on language is not new, finding articulation especially in the work of Lev Vygotsky (1986). Also Herbert Clark (1996), Michael Tomasello (1999) and Andy Clark (1997) have more recently pursued similar instrumental-functional accounts of language. In a continuation of these approaches, the purpose of this article is twofold: first we will consider in what respects language can be considered a tool for interacting minds, and secondly we will review findings from a multiplicity of disciplines that have investigated various aspects of the impact of language on human interaction.

We would like to thank the Danish National Research Foundation for supporting our work. CDF is also supported by the AHRC CNCC scheme AH/E511112/1.

Address for correspondence: Chris Frith, The Interacting Minds Group, Center for Functionally Integrative Neuroscience, Building 10g, 5th floor, Aarhus University Hospital, Nørrebrogade 44, DK-8000 Aarhus C, Denmark.
Email: cfrith@fil.ion.ucl.ac.uk

Mind & Language, Vol. 25, No. 1 February 2010, pp. 3–29.

1.1 Tools for Bodies and Minds

In its literal sense, a tool is a cultural construct that augments and enhances the natural potential of our bodies. Though we can use our hands, a spade is a more convenient and effective means for digging holes in the ground. And though we can walk, a bike brings us faster to our destination. But tools do not only enhance our natural bodily action potentials. The invention of some tools even makes entirely new bodily practices possible. For example, the invention of needle and thread make possible the new activity of sewing (Sinha, 2009). New tools and technologies thus often come to restructure and change the way we act in the material world, by creating new practices and new possibilities for action (Clark, 2006b). In addition to these 'world directed' tools, we also surround ourselves with other kinds of cultural artefacts. These are 'mind directed', in the sense that they enhance and restructure our cognitive abilities. An illustrative example is the abacus, a tool that, when manually manipulated, facilitates complex mathematical calculations (Clark, 1997). Other often-mentioned examples include calendars, notational systems, diagrams, and compasses, all of which, it can be argued, augment our cognitive performance in various respects (Hutchins, 1995; Stjernfelt, 2007). But it has also been suggested that verbal language can be thought of as a tool for solving particular kinds of cognitive problems. Comparable to the way the abacus takes calculations 'out of the head' and thereby allows for perceptual and dexterous manipulation and experimentation, spoken or written words, in a public language, can, by virtue of their simple material encodings (phonetic or graphic symbols), come to reify complex ideas and scaffold action (Clark, 2006a; Roepstorff, 2008). So far, most of the existing literature on this instrumental perspective on language has been preoccupied with the way that internalized verbal instruction (Vygotsky, 1986; Clowes and Morse, 2005), inner speech (Vygotsky, 1986; Clark, 1997), or the manipulation of material symbols (Clark, 2006a; Roepstorff, 2008), come to scaffold *individual* thought processes (important exclusions include Clark, 1996; Tomasello, 1999). Here, however, we intend to extend the language-as-a-tool perspective beyond the scope of individual minds and problem solving to investigate the role of language as mediating *between* minds and thus facilitating human interaction. We will consider four related ways in which language becomes an efficient tool for interacting minds.

1. Language extends the 'interaction space' in space and time: While most forms of non-linguistic, social interaction are bound to the here-and-now of online multisensory coordination between two or more individuals, language liberates social interaction from these immediate contexts. The discrete material (phonetic or graphic) and conventional-normative traits of linguistic symbols enable externalized and mediated forms of interaction that work over distances and times.
2. Language is an elaborate system for profiling and navigating joint attentional scenes (present, past, or fictional): Linguistic symbols (deictic markers, propositions, levels of categories, verb tense etc.) lend themselves as efficient

tools for structuring, profiling and navigating a joint attentional space in a precise way.

3. Language enables interacting people to align and share higher-order situation models and action plans. They are thereby in a better position for making predictions about what others are going to do, which in turn facilitates engagement in joint activities and actions. Sharing models also enables the creation of better models.
4. Language guides our attunement at a cultural level to certain aspects of visual, auditory and spatial perception. In this more global perspective, language-specific ways of profiling and construing the world come to influence and shape non-linguistic cognition.

2. What Language Enables Us to Do

2.1 Language as a Tool for Extending the Interaction Space

There is a sense in which language is both process and object—joint action and symbolic artefact. Though these aspects of language—process and object—may seem contradictory and mutually exclusive, they are in fact complementary and dependent (Raczaszek-Leonardi and Kelso, 2008). The process properties of language relate to the dynamic nature of the subtle online coordination and multi-level alignments especially characteristic for dialogical encounters (Clark, 1996; Garrod and Pickering, 2004). Language is thus a tightly interwoven and almost inseparable part of the interactive process in which individuals coordinate their bodies, actions, and perspectives in a continuous and complementary way to accomplish shared meaning and joint goals (Clark, 1996). Among the essential features of such communicative encounters are turn taking patterns and rhythms, along with a number of language internal and external semiotic resources such as prosody, intonation, stress patterns, gesture, facial expressions, gaze direction and bodily posture, all of which contribute to the unfolding meaning of the discourse (Goodwin, 2000). Process perspectives on linguistic communication thus tend to emphasise the 'here-and-now' situation, where two or more people are maximally present to each other. In other words, such interactions take place in, and are constrained by, what we will here term the *interaction space*. There are natural limits to the interaction space: The interacting parties have to be within a certain proximity of each other, and there must be sufficient illumination for them to register all the subtleties of visually accessible bodily and gestural cues. In a similar vein, the auditory environment (e.g. the volume of background noise) has to allow for the exchange of vocalizations. The interaction space thus seems to share features with the human *peripersonal space* (Rizzolatti, *et al.*, 1997).[1]

[1] In a series of recent experiments it has been shown that humans (as well as monkeys) project around them an immediate multisensory periphery in which aspects of the material environment can potentially be grasped and manipulated (Làdavas, 2002; Holmes and Spence, 2004)

But, as stated initially, language cannot be satisfactorily described only in terms of 'fluent' dynamics. Beyond the local, situated and highly interactive coordination processes, language is also an object. And perhaps the true flexibility and power of language is related to its *material* and *symbolic* aspects (Clark, 1997, 2006a; Clowes, 2007; Roepstorff, 2008). The object-characteristics of language thus relate to stable, recurrent entities such as 'words' and 'grammatical constructions' (hence collectively termed *linguistic symbols*). Unlike many of the process features of communication, linguistic symbols are not only constituted by their local instantiation and negotiated use, but form part of explicit rule-governed, conventional–normative systems (Itkonen, 2008; Zlatev, 2008). They thus transport more stably between quite different usage-situations, over shifts in modality (written versus spoken), and are generally less susceptible to local variation (e.g. a ditransitive construction remains a ditransitive construction whether whispered or yelled, with an Italian or Danish accent, etc.). But most importantly, the elaborate combinatorial system of discrete and parsimonious material encodings (auditory/phonetic or visual/graphic) makes language an efficient tool for putting complex ideas 'out in the open', allowing language users to share meaning and scaffold joint attention and actions.

The special nature of linguistic symbols also has another profound effect on communicative interaction. Analogous to the way that manual tool use has been shown to enlarge the peripersonal space by extending the bodily action potential of arm and hand in space (Farnè and Ladavas, 2000; Maravita *et al.*, 2001; Maravita and Iriki, 2004; Farnè *et al.*, 2007; Cardinali *et al.*, 2009), linguistic symbols liberate human interaction from the temporal and spatial immediacy of face-to-face and bodily coordination and thus radically expand the *interaction space*. While many of the non-linguistic modes of communication critically depend on attending to each other's hand gestures, facial expressions, and gaze directions etc., the auditory/phonetic qualities of speech enable interacting agents to communicate and coordinate while they engage in manual activities that occupy hands and vision (e.g. washing up together). More importantly, speech enables people to interact at a distance even when deprived of other sensory modalities of contact, as when e.g. located in adjacent rooms or talking over the telephone. The graphic encoding of speech (writing) extends the interaction space further, enabling us to interact at a temporal distance, as when we leave notes to each other on the kitchen table[2]. Finally, the introduction of language (from an onto- and phylogenetic perspective) even makes possible new social communicative practices such as narrative (Donald, 2001). Comprehensive complexes of meaning such as Martin Luther King's famous speech from 1963 or Lev Tolstoy's *Anna Karenina* (2008) are hard to imagine without the medium of language (however, this cultural

[2] This is Just to Say//I have eaten/the plums/that were in/the icebox//and which/you were probably/saving/for breakfast//Forgive me/they were delicious/so sweet/ and so cold (William Carlos Williams, 1963).

perspective on human language is beyond the scope of this article and will not be pursued any further).

'Linguistically extended interaction' has been investigated in a number of studies using different methodologies and measurements. In an experiment reported by Clark and Krych (2004), pairs of participants solved a task in which one person instructed another in how to build a model with Lego blocks. In one of the experimental conditions, the visual contact between participants was blocked so they had to rely solely on speech to solve the task. Though participants solved the task faster when the instructor had visual access to the builder's workspace, they were, nonetheless, able to complete the task in the 'language only' condition. Shockley *et al.* (2003) replicated these findings in a similar setup, but with an interesting extension. They measured the subtle, continuous bodily coordination ('postural sway') of participants while they solved a cooperative 'puzzle task'. Overall they found that participants tended to spontaneously coordinate their postural sway when engaged in the cooperative task. Interestingly, this effect was also found when participants could not see each other, but interacted only through verbal conversation. In a follow up study, Shockley and colleagues found that this coordination effect was due to convergent patterns of speech (e.g. aligned stress patterns and speech pace) characterizing cooperative conversation (Shockley *et al.*, 2007; Fowler *et al.*, 2008). Together, these findings suggest that even when interacting agents cannot make use of the rich expressive qualities of bodily gestures, eye gaze etc., but must rely on speech, they still manage to achieve tight coordination and even low-level bodily synchronization.

Another study put this point to a critical test by exploring a different modality of communication. Galantucci (2005) had pairs of participants play a cooperative game in a virtual environment. Participants' task was to find each other in a four-room maze architecture. Participants were seated in separate rooms and could only communicate by means of a digitizing pad. However, properties of the pad (a constant drift on the vertical axis) made it impossible to use any conventional signs such as letters or numbers. Playing against time, the only way to solve the game was to negotiate a new 'language' from scratch that would allow participants to share information about their whereabouts with each other. Despite great differences in speed (under twenty minutes to nearly three hours), most pairs eventually managed to jointly develop stable systems of symbols, which turned a nearly impossible task into a trivial one. The Galantucci experiment thus points to the remarkable efficiency of symbolically mediated coordination in joint action. By sharing and exchanging a minimal set of parsimonious symbols, players were able to coordinate their motions in the virtual space of the game and thereby accomplish a joint goal. In this experimental setting, at least, symbolic communication was not only the most efficient means of solving the task; it appears to be the only way of solving it. Without the symbolic extension (an emergent 'language') of their interaction space, the participants simply could not engage in the joint action of finding one another in the maze (for examples of other types of symbolically extended interactions, see e.g. Tylén *et al.*, 2009a; Tylén *et al.*, 2009b).

2.2 Language as a Tool for Alignment of Attention

It has been suggested that humans are the only species that point to share experiences (Tomasello, 2006). To what extent this depends on biological, ecological or cultural factors is a subject of heated debate (cf. e.g. Leavens and Racine, 2009 for a discussion), however most researchers agree that extensive engagement in joint-attentional activities is a characteristic trait of human sociality. While the practice of gaze-following and pointing has been found to subserve joint attention in pre-linguistic infant-adult interactions (Carpenter *et al.*, 1998), in later stages of development verbal language lends itself as another tool for directing and sharing attention (Tomasello, 1999; Talmy, 2000).

The function of language as profiling and structuring joint attentional scenes, rather than representing them, finds support in the observation that the linguistic contribution to the overall meaning of communicative situations is often largely underdetermined (Raczaszek-Leonardi and Kelso, 2008). Consider the utterance: 'Are you picking them up today?' Though the sentence is composed of high frequency words in a conventional interrogative construction, and at first glance seems perfectly meaningful, in fact the utterance makes little sense when detached from the situational context in which it is spoken (Wallentin *et al.*, 2006). Is it about kids being collected from daycare? Or little pearls that have fallen on the floor during a creative workshop? Or is it about trafficking of illegal drugs? We thus need contextual specification of (at least) the personal pronouns 'you' and 'them', not to mention the verb 'to pick up', in order to grasp the intended meaning of the utterance. To the interacting agents involved in the discourse, the specifications of intended meanings are certainly accessible and present in their shared situational context. What is at stake in the sentence is thus merely a (re)organization and profiling of the already existing state of affairs (e.g. 'is it *me* or *you*?', 'is it *today*?', etc). In this example, the role of the linguistic symbol string in the communicative event is thus not primarily to *represent* meaning, but rather to *structure*, *guide* and *constrain* joint attention and perspective-taking in an already existing, shared meaning space. Or, as put by Sinha (2004, p. 224): 'A symbol [. . .] directs and guides, not the behavior of the organism(s) receiving the signal, but their *understanding* (construal) or (minimally) their *attention*, with respect to a shared referential situation'.

Not only can the linguistic symbol string align the attention of the hearer to higher-order aspects of a situation ('is it *me* or *you*' in the example above), it can also guide the hearer's lower-level perceptual attention as well. For example, in a number of studies (Eberhard *et al.*, 1995; Tanenhaus *et al.*, 1995) it has been found that language comprehension guides attention to the visual context. When participants were shown a set of playing cards and given complex instructions, such as 'Put the five of hearts that is below the eight of clubs above the three of diamonds', they would visually track the information, so that as they heard 'the five of hearts' they would look at each of the two potential referents successively, and upon hearing 'below the' they would fixate on the card above the five of hearts that they were currently monitoring. By the time they heard the word 'clubs', they

would investigate the card above the other five of hearts. They would then identify the target card and remain there until the motor task was executed. Experiments of this type demonstrate that the linear string of language guides the hearer's attention in an online fashion, allowing her to home in on the current referential content during communication.

In a similar vein, Spivey *et al.* (2001) investigated the effect of language on a simple visual search task. A classic finding in the visual search literature is the pop-out effect: a green object, for instance, immediately 'pops out' among a group of red objects. But when one needs to locate an object that is defined by a conjunction of attributes, e.g. both colour and shape, then processing time increases steeply as a function of set size (Treisman and Gelade, 1980). Spivey and colleagues replicated this finding with a setup in which participants were initially asked e.g. 'Is there a red vertical?' and subsequently shown images containing red or green rectangles with either a vertical or a horizontal orientation. But interestingly, when participants were presented with the question at the same time as the image, the increase in response time as a function of how many objects were in the image became much shallower. By the time participants hear the word 'red', they can minimize the set size by eliminating all the green objects, making the job of finding the vertical bar substantially easier. It thus appears that a general function of language is that it allows a speaker to continuously minimize the set of possible attentional foci in the hearer's environment.

The studies reviewed so far illustrate how language can be used to guide the attention of the hearer in an on-line fashion. Before the speaker has finished her sentence, the hearer has already used the linguistic cues provided by the speaker to narrow the range of relevant information in another domain: the hearer's visual field. A series of studies by Richardson and colleagues illustrates that this process is not merely a matter of the speaker directing the hearer's attention; rather, language can be seen as a means of coordinating the attention of the speaker and the hearer.

Richardson and Dale (2005) asked four speakers to look at an array of pictures of characters from two well-known television shows (*Friends* and *The Simpsons*). The speakers were instructed to talk freely about the characters, describing a favourite episode or describing the relationships between the characters. As they spoke, their eye movements were tracked and their descriptions recorded. A portion of their speech was then played for a group of listeners who looked at the same array of pictures, and whose eye movements were also tracked. Listeners' eye movements were coordinated with the speaker's eye movements, such that within seconds of a speaker looking at a particular point, the listener was more likely than chance to be looking at the same point. In addition, listeners' performance on a memory test probing what the speaker had just said correlated positively with how closely their eye movements were coordinated with the speaker's.

The Richardson and Dale (2005) study represents a subtle but important shift from a view of language as an attention-directing device to language as an attention-coordinating device. The speakers and hearers in this study were separated in space and time, and yet their eye-movements were coordinated with astonishing

precision, with only language serving as a guide. Richardson, Dale, and Kirkham (2007) extended this finding to a real-time dialog situation. Using the same stimuli as in Richardson and Dale (2005), they tracked the eye movements of pairs of participants who, although physically separated, were both looking at the same array of pictures of television characters and discussing them verbally. In this two-way dialog situation there was no clear leader or follower. Nevertheless, participants coordinated their eye movements and looked at the same points within seconds of each other.

These results suggest that while pointing is a very efficient means of directing attention when both parties can see one another, language can also serve as a pointing device, guiding people's attention in physical space even when speaker and listener are physically and/or temporally separated from one another. Furthermore, the coordination of eye movements, achieved by means of language, has repercussions for our ability to remember what another person has said. In broader terms, this suggests that language is an important tool for coordinating our joint attention in both physical and interactional space, as well as guiding our unconscious decisions about what we should commit to memory. This common encoding of certain information to memory is likely to have implications for the two parties' ability to act in unison on an object or situation at a later point in time.

By a highly elaborated system of temporo-spatial deictics and prepositions, basic level, superordinate, and subordinate categories, tense and aspect, etc., language is a very efficient means for navigating and profiling a joint attentional and referential space in precise and subtle ways (Talmy, 2000). When we use words to point to objects, relations or activities in a shared perceptual scene, language-internal structures prompt us to specify a certain perspective or construal of that focus. We can for instance refer to the same object as either a piece of furniture, a chair, an armchair, or an 'Arne Jacobson', directing attention to the categorical, functional or connotative property relevant for a particular discourse situation (Tomasello, 1999).

The work of Semin, Fiedler, and others on the role of lexical selection in social descriptions provides an illustration. Their Linguistic Category Model (LCM) establishes a means of categorizing words on a concreteness-abstractness scale. Semin and Fiedler (1988) provide evidence that abstract words carry with them an implication that the trait described is an enduring one. At the same time, traits described with abstract words are difficult to verify or disprove, while the reverse is true for concrete words. Consider a hypothetical situation in which A hits B on the arm with his fist. Using the LCM scale, this same physical action could be described at four levels:

(1) 'A punches B' (a 'descriptive action verb'), most concrete;
(2) 'A hurts B' (an 'interpretive action verb');
(3) 'A hates B' (a 'state verb');
(4) 'A is aggressive' (an adjective), most abstract.

(1) merely describes the immediate situation, says little about A in general, and can be easily verified, while at the other end of the spectrum, (4) draws a largely

unsubstantiated inference from the concrete situation, implies a great degree of enduringness (A just *is* aggressive, and will probably always be so), and is very difficult to disprove.

Maass *et al.* (1989) used this scale to analyse subjects' descriptions of actions performed by in-group and out-group characters in a picture-description task. The subjects came from competing teams in a traditional Italian horse-racing competition. The stimuli depicted members of the subject's own team or the opposing team (as indicated by the team colours on the character's shirt) carrying out either positive or negative actions, such as either helping someone (positive) or littering (negative). Subjects were told that the scenes depicted real events that had transpired over the past two years. Subjects were given four description choices for each picture, corresponding to the four levels of abstraction.

As expected, participants described negative actions performed by members of their own group using the most concrete terms, implying that the action may have been a one-time offence, with no implications for the character of the person in general. The same action, however, when performed by members of the out-group, was described in abstract terms, implying that this was a typical example of the person's generally distasteful character. The reverse trend was seen for positive actions. The effect for positive descriptions was replicated in a free-description task.

Rubini and Menegatti (2008) have taken this line of research a step further, using naturalistic data collected from a university hiring committee. Working with a corpus of written evaluations of applicants for research and professor positions, they found that the hiring committee tailored their language in such a way that positive descriptions of candidates were written using more abstract language than negative descriptions (to avoid being unnecessarily derogative to rejected candidates), and that positive descriptions of successful candidates were more abstract than positive descriptions of rejected candidates. In addition, individual committee members were more concrete in their negative evaluations of applicants with whom they had co-authored papers than in their evaluations of applicants with whom they had no work interdependence. Thus, individual members of the hiring committee also systematically employed language as a tool to further more personal agendas.

Finally, Semin, Gil de Montes, and Valencia (2003) investigated the importance of communicative context on linguistic bias. Participants were asked to describe actions that they were told had been carried out by another person with whom they would either cooperate or compete against in a quiz game. In some cases, they were told that the person would read the description before the game; in other cases, they were told that the person would not see the descriptions. In this way, the researchers constructed a situation in which participants were either supplied with a communicative purpose for the message they were to construct, or with no communicative purpose (the other person would not read the message). The authors found that in the conditions in which participants had a clear audience, they systematically varied their language along the concrete-abstract scale. However, in the condition in which they were told that the person being described would not see the message, they did not vary their word choice in any systematic way.

This final experiment nicely illustrates the tool-like nature of language in pragmatics. Although people do seem to vary their language along the concrete-abstract scale in predictable ways, and according to internal psychological factors, they only do this to the extent that it is pragmatically relevant. The variation of utterances along the concrete-abstract scale is thus a feature of language that is only useful to the extent that it can exert an effect on the receiver.

Research on the Linguistic Coding Model provides an insight into some of the mechanisms involved in word choice that can in turn impact the social perceptions and, ultimately, the actions of others. This can operate at the local level, for instance when participants in the quiz game experiment strategically varied their language in order to reinforce an alliance with a partner, or in the case of the individual university hiring-board members who subtly alter their language along the concrete-abstract spectrum in order to present candidates with whom they have a prior connection in a more favourable light. This same mechanism seems to operate at the group level as well, as in the case of the Italian riding clubs, who systematically varied their descriptive language in order to maintain and reinforce the negative stereotypes associated with the opposing club, and the positive associations within their own group.

As we have seen in the previous sections, language plays an important role in mediating interpersonal activities when these are grounded in a concrete situation. However, the research reviewed here on linguistic bias effects illustrates that the symbolic nature of language also allows it to exert a subtle cohesive or divisive influence at a larger distance. The university hiring committees, for instance, cannot use a smile to lighten the blow of their criticism of unsuccessful candidates. They can, however, modulate their use of abstract and concrete language to deliver the same message while sparing the feelings of the recipient. At the same time, and in the same text, they can also achieve another goal, aimed at another audience: justifying their choice of one candidate over the other, by subtly influencing the reader's perception of the candidate by means of the same linguistic technique.

2.3 Language as a Tool for Sharing Situation Models and Action Plans

Many everyday situations require that two or more people get together to share perceptions, engage in joint collaborative tasks or otherwise coordinate their attention and attitudes to states of affairs in the world. Recent developments in joint action and social cognition research point to a number of processes at various levels of cognition involved in such coordination processes. For instance, it has been found that people tend to spontaneously and unconsciously follow gaze (Zuberbuhler, 2008), synchronize and otherwise influence each other's movement patterns (Richardson *et al.*, 2007; Atmaca *et al.*, 2008; Richardson *et al.*, 2008). However, sometimes a joint task may not only require synchronization of perception and movement, but also fine coordination of *complementary* actions to achieve a common goal (Sebanz *et al.*, 2006; Richardson *et al.*, 2007). To illustrate, imagine two persons moving large furniture down a narrow staircase. In such cases, the

synchronization of action might not help at all. Rather, each participant must take different roles and do quite different things in relation to the overall project, and their success largely depends on well-coordinated, spatio-temporal orchestration of distributed acts in relation to each other, the object they carry, and the physical environment, including the challenging route and the obstacles they may encounter (Marsh *et al.*, 2009). In order to achieve smooth action coordination and timing, it is not sufficient to perceive and understand what the other is doing, we must also be able to make predictions about what she will do next (Frith and Frith, 2006; Sebanz and Knoblich, 2009). It might turn out that we have different conceptions of the task and our roles in it, with potentially catastrophic consequences. A crucial element of such joint activities is thus mutual alignment of expectations and sharing of higher-order situation models, action plans and goals (Jack and Roepstorff, 2002; Roepstorff and Frith, 2004; Tomasello *et al.*, 2005). The alignment of action plans and goals enables interacting agents to optimize their predictive models (Roepstorff, 2004; Kilner *et al.*, 2007) by providing a common ground against which actions and expressive behaviours can be interpreted (Clark and Brennan, 1991). Language is an efficient tool for building up and aligning shared situation models, action plans and goals (Clark, 1996; Pickering and Garrod, 2004)[3]. If we are to move heavy furniture down narrow staircases without causing damage to furniture, fingers or doorways, language offers itself as an effective means for negotiating joint plans and coordinating actions (Bangerter and Clark, 2003; Fowler *et al.*, 2008).

This point is richly illustrated in a study by Bangerter and Clark (2003). They investigated how interacting agents use dialogue to coordinate and navigate various kinds of joint projects similar to the one described above. The general finding is that distinct verbal cues are used between agents to navigate through different phases of a joint activity. In one of the analyzed examples, pairs of participants are to plan the shipment of goods from pier A to B (cf. the Trains task, Gross *et al.*, 1993). The task divides into a number of projects and subprojects (how to get type A or B of goods from one means of transportation to another, and off to the goal destination, etc.). Bangerter and Clark find that a pattern of verbal cues (*project markers)* complements the participants' distributed actions to mark and initiate transitions between phases of the projects. Across extensive corpora of project dialogues, cue words like '*okay*' and '*all right*' are recurrently used to initiate and exit subprojects (marking so-called *vertical transitions*), while '*yeah*', '*yes*', 'yep', '*m-hm*', and *'uh-huh*' are used to ground contributions within a subproject and mark continuation (*horizontal transitions*). The study thus points to the way that subtle linguistic markers are systematically used by interacting agents to coordinate their joint progression through a multistage project plan, marking points of transitions and updating shared situation models.

[3] Bangerter and Clark even suggest that the main purpose of language (or dialogue) is to coordinate joint activity: 'Still, ... basic joint activities are primary, and dialogue is created to manage them' (2003, p. 196)

In a study of another kind of multimodal joint action, a game of hopscotch, Goodwin (2000) describes the rich use of gesture, body posture, and prosody used by a young girl in challenging another girl's move. The argument arises when one of the girls jumps into what the other girl considers to be the wrong square on the hopscotch grid. Physically blocking the girl whose turn it is, and thus preventing her from finishing her turn, she verbally challenges her saying '*Cheater—because this is the four. And you go in the four. Don't go in the fifth*'.[4] Though Goodwin argues that language is only one of many 'semiotic fields' employed by the speaker, it still seems to play a special role in this example. Like many children's games, hopscotch involves normative practices for how to act at particular times and a structured space consisting of fields with different meanings (in this case a configuration of numbered squares). To participate in the game you must know and conform to these practices and scripts for action. Some of the properties of the game environment might be explicit, such as the hopscotch grid drawn on the sidewalk, while other are implicit and inferred by convention. In the referred case, the particular numbers of each field are thus part of such implicitly shared situation models and the dispute between the girls seems to be due to *misalignment* of models resulting in differences in the attribution of numbers to the fields. The offended girl uses gesture, prosody, and body position to efficiently establish that she is dissatisfied and wants to interrupt the game, but the cause of her dissatisfaction—the other girl's apparent violation of the rules of the game—is expressed by verbal means. Language thus seems to be the preferred means by which we negotiate and share normative structures such as game rules and scripts (Wyman *et al.*, 2009).

The same properties of language also make possible efficient forms of instruction and learning. This point has been vividly illustrated in comparative studies of primate cognition. While it takes several months to train and accustom monkey participants to a novel experimental task, human participants often need less than half an hour of verbal instruction to get acquainted with the task (Roepstorff and Frith, 2004). Linguistic instructions can also be internalized and used to guide future actions by 'silent repetition', which has been repeatedly shown to enhance performance in a variety of cognitive tasks (Vygotsky, 1986; Clark, 1997; Clowes and Morse, 2005). Furthermore, humans can be conditioned to associate a stimulus with fear through mere verbal instruction (Olsson and Phelps, 2004). Whereas most species can only learn about the world through direct, embodied sensory-motor experiences (e.g. reinforcement and Pavlovian conditioning), humans complement these with effective socio-cultural styles of leaning e.g. by the language-mediated sharing of predictive models, plans and goals (Tomasello, 1999; Roepstorff, 2004).

[4] Orig.: '*Chiriona—porque éste es el quarto. Y tú vas en el quarto. No vas en el quinto*' (the English translation is taken from Goodwin, 2000)

2.4 How Language Shapes Interacting Minds

The availability of tools can also have large and long-term effects on the human body and mind. For example, the development of scissors and razors has had a dramatic effect on the appearance of human faces. The tool of language can have similar pervasive and long-term effects. Likewise, the spread of literacy and the wide availability of writing materials have meant that we no longer have to develop the memory skills necessary in classical antiquity to maintain the knowledge of the scholar or the speeches of the lawyer (Cicero, 1960).

Language not only helps us share local models of interaction. In certain instances it also adds to the long-term structuring of these models. The fact that language guides attention (see section 2.2) may have long-term consequences. If attention is continuously guided in certain directions and never in others, then this will affect the processing pathways of the maturing mind/brain. In the study reviewed in section 2.2, Richardson and Dale (2005) showed how informants tracked relevant pictures of faces from a well-known television series during the unfolding of a verbal description of an episode. This is of course only possible because a long-term storage of associations between the names of the television characters and their visual characteristics has taken place. Interestingly, this type of association has been documented in a single-neuron recording study (Quiroga *et al.*, 2005). In this study, firing patterns from neurons in the hippocampus were obtained in patients suffering from pharmacologically intractable epilepsy. Some of these neurons were found to fire selectively when the patient watched images of a particular famous actress. But more relevant for this line of argument, the same neuron was also found to fire when the patient read the name of the actress. It did not respond to other known faces or other known names. This is a clear demonstration of how the brain adapts to the input it gets and that symbols, such as proper names, form a vital part of this blend.

However, linguistic norms may also have an impact on the cognitive system, as witnessed by recent findings from language acquisition and from linguistic anthropology. An illustrative example is the way language shapes hearing. During the acquisition of a language, phonetic distinctions become categorical (Kuhl, 2004). Categorical perception is the tendency for speakers of a language to experience sharp boundaries between the sounds used in their languages, showing no sensitivity to intermediate sounds. Certain languages, such as English and Danish distinguish the two sounds /ra/ and /la/ whereas others, such as Japanese, do not. If speakers of English are asked to identify sounds that vary in small equidistant steps from /ra/ to /la/, they experience an abrupt shift from one sound to the other at some point, rather than a continuous change. Japanese speakers, on the other hand, experience all sounds as belonging to the same phonetic category. In other words, our minds adapt to the sounds of a particular language community. Neurophysiological studies with electroencephalography (EEG) and magnetoencephalography (MEG) have revealed language-specific neural markers of phonetic categories in the left hemisphere of the brain, in both adults and in developing infants (Näätänen *et al.*, 1997; Cheour *et al.*, 1998).

Similar observations have been made for vision. For many years the standard assumption in the colour categorization literature has been that there exist six universal focal colours corresponding to prototypical English primary colours (black, white, red, green, yellow, and blue) (Berlin and Kay, 1969; Rosch, 1973). Opposing this has been a more relativistic view according to which cultures are the sole constraining factor for colour categorization (e.g. Ray, 1952). Recently, the latter view has gained more momentum, with studies finding cultures with colour divisions that do not have the 'natural' colours as their focal points (Davidoff *et al.*, 1999; Roberson *et al.*, 2000; Roberson *et al.*, 2005; Winawer *et al.*, 2007). Roberson and colleagues investigated colour categories in Berinmo, a language spoken in Papua New Guinea, and found that categorical boundaries here did not conform to the claimed universals.

In a similar vein, Winawer and colleagues (2007) studied 'Blues in Russian'. Russian makes an obligatory distinction between lighter blue colours ('goluboy') and darker blues ('siniy'). The study showed that Russian speakers were faster at a non-linguistic colour discrimination task when two colours fell into different linguistic categories compared to when they were from the same category (e.g. both of the 'goluboy' type). English speakers did not exhibit a similar differentiation.

Attempting to bring together these apparently conflicting findings, Regier and colleagues (Kay and Regier, 2007; Regier *et al.*, 2007) simulated an optimal division of colour space and compared that to data from the World Colour Survey (WCS). While many languages conform to an optimal or near-optimal division of colour space, Regier *et al.* also found 'many languages in the WCS with colour-naming systems that are not very similar to the hypothetically optimal model configurations'. These divisions, however, were not found to be completely arbitrary, e.g. when comparing the Berinmo colour divisions (see above) to artificial divisions derived from altering the original Berinmo data slightly (by rotating them along the hue dimension) the original data was found to conform better to the optimal than any of the artificial versions. This result could be generalised to all of the languages in the WCS and it suggests that there are universal attractors for carving colour space, but that these leave considerable room for cultural variation, even when this leads to non-optimal categorization. This is consistent with a weak relativism (Kay and Kempton, 1984) where existing linguistic conventions of a society contribute to constraining how colour space is divided. This is not very surprising, given that much of the relevance for colour categories is rooted in the social realm (Wallentin and Frith, 2008).

Even congenitally blind children can acquire and use a colour vocabulary with some sophistication. Landau and Gleitman (1985) and Landau (2000) studied the child Kelli and observed that she produced visual words, such as 'look' and 'red', and that these words were used in a semantically coherent fashion. They therefore argue that much of the semantics of words can be acquired by language internal means, i.e. through distributional evidence and syntactic context. Their assumption is that this is facilitated by innate knowledge about the structure of language. Whether or not the latter is true is beyond the scope of this article (see e.g. Evans

and Levinson, 2009 for a discussion), but it suffices to say that the result again seems to be that language in itself is an important constraining and driving factor in the acquisition of even the most 'perceptual' words.

Written language has also had an impact on our experience of events in the visual world. Chatterjee *et al*. (1999) asked their study participants to make drawings of events unfolding in the horizontal plane, e.g. 'a staggering drunk'. They found that drawings depicted theses events as going from left to right with a ratio of 7:1. In a subsequent experiment participants were shown images of simple stick figures interacting. These pictures had ambiguous interpretations, e.g. an image could be understood either as one stick man pushing the second or as the second stick man pulling the first. Thus action either moved away from the agent (push) or towards the agent (pull). In a picture/sentence matching task participants responded faster both if the agent was placed to the left of the patient and if the action proceeded from left to right. This suggested that participants 'read' images in a left-to-right fashion and this was originally interpreted as an effect related to brain lateralization. Subsequently, however, it was found that speakers of Korean, who were not brought up reading left-to-right, did not exhibit this left-to-right bias (Barrett *et al*., 2002). Pre-school children did not show it either (Dobel *et al*., 2007), and adult readers accustomed to a right-to-left reading system, such as Israelis, exhibited a right-to-left bias (Dobel *et al*., 2007). It therefore seems most likely that these effects are caused by the exposure to written language rather than arising from brain lateralization. It thus constitutes another example of the way that language shapes our minds.

Another example of the mind adapting to linguistic conventions comes from the study of how languages code for spatial relations. Across languages, it has been suggested that languages code spatial relations using one or more of three overall reference systems: relative, intrinsic or absolute (Levinson, 2003; Majid *et al*., 2004; Kemmerer, 2006). Within the relative reference frame, objects are usually situated relative to the speaker's own egocentric point of view, i.e. we might say that 'X is to the left of Y' meaning that from where I stand, Mary is positioned to the left of her car. But if Y has an internal orientation of its own, such as a car with its front and back we may also say that 'X is to the left of Y', meaning that Mary is to the left of her car when seen from her car's perspective. This is called the intrinsic or object centred reference frame. Lastly, we may say that X is to the north of Y and use an absolute bearing. In Western cultures we mainly use the relative and the intrinsic frames of reference, but other languages completely lack these types of reference frames and speakers thus come to rely solely on the absolute system for communicating. Children have been shown to already be sensitive to language specific spatial categories during their second year of life (Choi *et al*., 1999; Levinson, 2003) and it is therefore perhaps not surprising that cultures with a language that supports only one kind of reference frame become highly skilled at this kind of spatial reckoning (Levinson, 2003). Thus, when asked to point towards known non-visible landmarks, speakers of Guugu Yimithirr, an Australian aboriginal language with an absolute referencing system, are remarkably skilled at

this task, while speakers of relative frame languages, such as English or Dutch, have been found to be quite poor at this task (Levinson, 1997, 2003). No matter where a speaker of an absolute language is situated, she needs to keep track of these absolute coordinates in order to be able to communicate:

> I believe that the explanation lies in the facilitative effect of language on cognition. It is clear to children that the absolute system is important in adult production, as shown by really early use of the terminology. Children then work hard to crack the code (Levinson, 2003, p. 311).

In a similar vein, while language is a driving factor for learning due to its importance in social interaction, the specificities of linguistic conventions also focus and narrow the particular form and scope of learning.

3. Language as a Special Tool

In this paper we have reviewed evidence for our assertion that language is a tool for interacting minds. Like a rake, language extends the space within which minds can interact. Like a torch, language highlights the important regions of the space within which minds are interacting. Like a mould, language creates a common space for minds to interact in. Like a pair of spectacles, language alters the way interacting minds see the world. In this final section we will consider briefly some further variations of the idea of language as a tool for interacting minds. First, language can be deliberately used to restrict interaction. Second, language appears to be a uniquely human tool. Third, language appears to be a special kind of tool that affects both the source and the target of action.

3.1 Language as a Tool for Competing Minds

Interacting minds enable a group to perform better than a collection of individuals. However, there is a dark side to the prosocial behaviour of interacting individuals within a group. The altruism that we show to members of our in-group does not extend to members of other groups (Sober and Wilson, 1998; Fehr *et al.*, 2008; Xu *et al.*, 2009).

As any foreigner who has attempted to get along in a new language knows, language is an exquisitely precise signal that someone is a member of another group. This distinction goes beyond accent. By using not quite the right vocabulary a speaker who has dared to cross disciplines, will immediately be recognized as an amateur rather than a professional brain imager, linguist or what ever the discipline of the conference may be. The professionals would justify this use of jargon on the grounds that it is a common language of exact technical terms. However, languages known as *cant* or *slang* are also deliberately developed to prevent understanding by out-groups. Some claim that the word *slang* is derived from *s*ecret *lang*uage. Such

languages are typically used by criminals, travellers and other persecuted groups. Examples include pig Latin, backslang, rotwelsch and Polari ('So bona to vada your old eek'). This is using language as a screen behind which one group of minds can interact without another group knowing what is taking place.

3.2 Language Depends Upon Uniquely Human Cognition

There is ample evidence of tools use in a number of other species including birds (Kirsch *et al.*, 2008), apes (Breuer *et al.*, 2005) and also monkeys (Fragaszy *et al.*, 2004). However, language appears unique to the human species. So what is the unique form of cognition that enables humans to use language as a tool?

We have suggested that language provides a particularly efficient way of sharing our knowledge of the world with others. But before we share knowledge we have to know that we have the knowledge and also that we are sufficiently confident in this knowledge to believe that it is worth sharing. Knowing that we have knowledge and knowing how confident we are in it requires metacognition (Nelson, 1992). There is some evidence for metacognition in other species, but this remains extremely rudimentary compared to humans. For example, there is some evidence that animals know something of how well they can remember an item (Smith, 2009). However, there is little evidence of animals trying to share their knowledge. Gómez *et al.* (1993) suggest that one key difference between pointing in apes and infants is that apes usually only point to request things (food, objects, actions, etc.) from others, whereas infants also point to show and share attention towards things. Along similar lines, Liszkowski *et al.* (2009) have shown that pre-linguistic infants will point to non-existent entities (e.g. the plate where the biscuits used to be) while chimpanzees will not.

The behaviour of the infants in these examples requires an understanding that the sharing of knowledge is possible. So, for example, when pointing to a non-existent entity the infant must recognise that the adult also knows that this is the plate where the biscuits used to be. In other words, that they have common knowledge. It appears to be this uniquely human form of cognition that enabled language to develop as a tool for interacting minds in the first place.

3.3 Closing the Loop: Language as a Tool for Interacting Minds

Unlike other tools, language operates in mental, rather than physical space. The screen, behind which the user of slang hides, does not prevent others from hearing what is being said. The ignorant hearer might even be able to repeat what had just been said, but he would not know what it meant. This is what is special about the language tool. It enables minds to negotiate and share meanings. However, meaning is a slippery concept. So, to put it in other words, language enables us to share our knowledge of the state of the world and also to share our intentions and plans for changing the state of the world. Through such sharing we can align our knowledge and intentions and thereby dramatically improve the efficiency and

success of joint action. Thus, language is not only involved in a one-way action. Other people can answer back; modifications and effects go both ways. People engage in a rapid interaction, the result of which is to co-construct an emergent pattern of understanding, where participants use each other for scaffolding, and end up in places none of them could have reached on their own. In this way, language ultimately 'works' as a tool, not only because it allows one to act on and with other minds, but also because it allows others to act on oneself. We call this aspect of language use 'closing the loop'. In interesting ways, language, ultimately, comes to act on language itself as patterns of usage get inscribed in ways of understanding, which again affects usage.

4. Conclusion

Language is a complex phenomenon that can be approached from a number of perspectives. Some approaches emphasise language as a special kind of knowledge (Chomsky, 1986; Croft, 2001), others its dynamic process properties (Clark, 1996; Raczaszek-Leonardi and Kelso, 2008), while yet others conceive of language as an object (Clark, 2006a; Roepstorff, 2008), or a bodily skill (Arbib, 2005; Zlatev, 2008). Our purpose in this article has not been so much to discuss what language *is*; rather, we consider what language *does*. Thus, we have investigated the contextualized, functional role of language in particular social, interactive settings. Based on evidence from a multiplicity of disciplines, we argue that, in a range of everyday social situations and practices, language can be conceived of as a tool that enables effective and flexible forms of social coordination and interaction. To this end, we have elaborated on four closely-related yet different ways in which language is employed in the facilitation of social interaction.

First, as argued in section 2.1 language liberates interaction from the immediacy of here-and-now, online interaction. While most forms of non-linguistic communication are constrained to face-to-face encounters where the interacting agents are within a certain proximity and visual exposure to each other, the material phonetic and graphic character of spoken and written linguistic symbols dramatically extends the space for potential interaction in space and time. As shown for instance in the studies of Clark and Krych (2004) and Shockley *et al*. (2003), the application of linguistic means enables continuous coordination (including synchronization of postural sway) even when agents occupy their hands and bodies for other manual purposes or move out of immediate visual contact.

Second, by extending manual pointing, language offers sophisticated ways of aligning, profiling, and navigating joint attentional scenes, as argued in section 2.2. Here another aspect of language materiality—the linear, sequential organization of linguistic symbol strings—serves as a means to scaffold visual search strategies, and, together with deictic markers, levels of categorization, prepositional systems, etc., enable interacting agents to home in on shared foci of attention (cf. e.g. Eberhard *et al*., 1995; Tanenhaus *et al*., 1995). In addition, some of the same structural

properties of language can be used to jointly construe situations, social identities and in-group/out-group relations, as e.g. illustrated in the studies of Semin and Fiedler (1988) and Maass *et al.* (1989).

In a third tool-perspective on language, explored in section 2.3, we point to the way that language is used among interacting agents to align and navigate higher-order situation models and action plans. Often joint projects require agents to orchestrate finely timed, complementary actions in relation to each other and an object of mutual attention. In such cases it is crucial to share the same conception not only of goals and sub-goals but also of each agent's roles and contributions at various times in the project. Language is an efficient means to negotiate and share action scripts and establish common ground, which allows interacting agents to anticipate each other's actions and thus achieve smooth coordination. Bangerter and Clark (2003) show how parsimonious linguistic markers allow interacting agents to navigate and monitor their joint progression through a multistage project. Likewise, language is a well-suited medium for the negotiation of normative rules e.g. in children's games. While the girls in Goodwin's hopscotch-example (2000) get along well with a wide array of non-linguistic cues and signs, they seem to need language to get the rules sorted out and their situation models realigned.

But language is more than an instrument applied in an ad hoc fashion to accomplish local coordination. In line with an extensive number of cross-cultural studies on e.g. categorization and spatial referencing, we have argued that the structuring and attention-guiding character of language itself has some long-term effects on (non-linguistic) human cognition (section 2.4). Languages across the world show large variations in the way they draw categorical lines and construe relations and events. As a consequence, when a child growing up in an English-speaking community acquires the normative structures of her mother tongue, she has to attend to other aspects of the surroundings than a child growing up in an Australian aboriginal community. In the end, each of them becomes competent in making the particular perceptual and relational distinctions that their language codes for. Language thus becomes an important medium for the cultural shaping of cognition.

These four properties of language-as-tool may come in play with varying strength in concrete situations. Together they form a powerful repertoire for people, both embodied and mindful, to act together, and to co-construct interaction spaces that extend in time and space beyond the immediate present. This allows for actions that often only with great effort—if at all—could be done by an isolated individual. Hence language is a key tool for creating the advantages that interacting minds have over individual cognition.

Kristian Tylén
The Interacting Minds Group
Center for Functionally Integrative Neuroscience
Aarhus University Hospital
and
Center for Semiotics

Department of Scandinavian Studies
University of Aarhus
Denmark

Ethan Weed
The Interacting Minds Group
Center for Functionally Integrative Neuroscience
Aarhus University Hospital
and
Department of Archeology, Anthropology and Linguistics
University of Aarhus

Mikkel Wallentin
The Interacting Minds Group
Center for Functionally Integrative Neuroscience
Aarhus University Hospital
and
Center for Semiotics
Department of Scandinavian Studies
University of Aarhus

Andreas Roepstorff
The Interacting Minds Group
Center for Functionally Integrative Neuroscience
Aarhus University Hospital
and
Department of Archeology, Anthropology and Linguistics
University of Aarhus

Chris D. Frith
The Interacting Minds Group
Center for Functionally Integrative Neuroscience
Aarhus University Hospital
and
Welcome Trust Centre for Neuroimaging
University College London

References

Arbib, M. A. 2005: From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, 28, 105–24; discussion, 125-67.

Atmaca, S., Sebanz, N., Prinz, W. and Knoblich, G. 2008: Action co-representation: the joint SNARC effect. *Social Neuroscience*, 3, 410–20.

Bangerter, A. and Clark, H. H. 2003: Navigating joint projects with dialogue. *Cognitive Science*, 27, 195–225.

Barrett, A. M., Kim, M., Crucian, G. P. and Heilman, K. M. 2002: Spatial bias: effects of early reading direction on Korean subjects. *Neuropsychologia*, 40, 1003–12.

Berlin, B. and Kay, P. 1969: *Basic Color Terms: Their Universality and Evolution*. Berkeley, CA: University of California Press.

Breuer, T., Ndoundou-Hockemba, M. and Fishlock, V. 2005: First observation of tool use in wild gorillas. *PLoS Biology*, 3(11), e380.

Cardinali, L., Frassinetti, F., Brozzoli, C., Urquizar, C., Roy, A. C. and Farnè, A. 2009: Tool-use induces morphological updating of the body schema. *Current Biology*, 19, 478–9.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G. and Moore, C. 1998: Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, 63(4), i–174.

Chatterjee, A., Southwood, M. H. and Basilico, D. 1999: Verbs, events and spatial representations. *Neuropsychologia*, 37, 395–402.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K. and Naatanen, R. 1998: Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience*, 1, 351–3.

Choi, S., McDonough, L., Bowerman, M. and Mandler, J. 1999: Early sensitivity to language-specific spatial categories in English and Korean. *Cognitive Development*, 14, 241–68.

Chomsky, N. 1986: *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger.

Cicero 1960: *De Oratore, II*. Cambridge, MA: Harvard University Press, Loeb Classical Library.

Clark, A. 1997: *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: MIT Press.

Clark, A. 2006a: Language, embodiment, and the cognitive niche. *Trends in Cognitive Sciences*, 10, 370–4.

Clark, A. 2006b: Material Symbols. *Philosophical Psychology*,19(3), 1–17.

Clark, H. H. 1996: *Using Language*. Cambridge: Cambridge University Press.

Clark, H. H. and Krych, M. A. 2004: Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, 50, 62–81.

Clark, H. H. and Brennan, S. 1991: Grounding in communication. In L. B. Resnick, J. M. Levine and S. D. Teasley (eds), *Perspectives on Socially Shared Cognition*. Washington, DC: American Psychological Association.

Clowes, R. 2007: Semiotic Symbols and the Missing Theory of Thinking. *Interaction Studies*, 8, 105–24.

Clowes, R. and Morse, A. F. 2005: Scaffolding cognition with words. In L. Berthouze, F. Kaplan, H. Kozima, Y. Yano, J. Konczak, G. Metta, J. Nadel, G. Sandini, G. Stojanov and C. Balkenius (eds), *Proceedings of the 5th International Workshop on Epigenetic Robotics*, Nara, Japan. Lund: Lund University Cognitive Studies 123, 101–5.

Croft, W. 2001: *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Davidoff, J., Davies, I. and Roberson, D. 1999: Colour categories in a stone-age tribe. *Nature*, 398(6724), 203–4.

Dobel, C., Diesendruck, G. and Bölte, J. 2007: How writing system and age influence spatial representations of actions: a developmental, cross-linguistic study. *Psychological Science*, 18, 487–91.

Donald, M. 2001: *A Mind So Rare: The Evolution of Human Consciousness*. New York: Norton.

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C. and Tanenhaus, M. K. 1995: Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, 24, 409–36.

Evans, N. and Levinson, S. C. 2009: The myth of language universals: language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, 32, 429–48; discussion, 448–494.

Farnè, A. and Ladavas, E. 2000: Dynamic size-change of hand peripersonal space following tool use. *Neuroreport*, 11, 1645–9.

Farnè, A., Serino, A. and Làdavas, E. 2007: Dynamic size-change of peri-hand space following tool-use: determinants and spatial characteristics revealed through cross-modal extinction. *Cortex*, 43, 436–43.

Fehr, E., Bernhard, H. and Rockenbach, B. 2008: Egalitarianism in young children. *Nature*, 454(7208), 1079–83.

Fowler, C. A., Richardson, M. J., Marsh, K. L. and Shockley, K. 2008: Language use, coordination, and the emergence of cooperative action. In A. J. Fuchs and V. Jirsa (eds), *Coordination: Neural, Behavioral and Social Dynamics*. Dordrecht: Springer.

Fragaszy, D., Izar, P., Visalberghi, E., Ottoni, E. B. and de Oliveira, M. G. 2004: Wild capuchin monkeys (Cebus libidinosus) use anvils and stone pounding tools. *American Journal of Primatology*, 64, 359–66.

Frith, C. D. and Frith, U. 2006: How we predict what other people are going to do. *Brain Research*, 1079, 36–46.

Galantucci, B. 2005: An experimental study of the emergence of human communication systems. *Cognitive Science*, 29, 737–67.

Garrod, S. and Pickering, M. J. 2004: Why is conversation so easy? *Trends in Cognitive Sciences*, 8, 8–11.

Gómez, J. C., Sarriá, E. and Tamarit, J. 1993: The comparative study of early communication and theories of mind: ontogeny, phylogeny and pathology. In S. Baron-Cohen, H. Tager-Flusberg and D. Cohen (eds), *Understanding Other Minds: Perspectives from Autism*. Oxford: Oxford University Press.

Goodwin, C. 2000: Action and embodiment within human situated interaction. *Journal of Pragmatics*, 32, 1489–1522.

Gross, D., Allen, J. F. and Traum, D. R. 1993: *The Trains 91 Dialogues*. Rochester, NY: Computer Science Department, University of Rochester.

Holmes, N. P. and Spence, C. 2004: The body schema and the multisensory representation(s) of peripersonal space. *Cognitive Processing*, 5, 94–105.

Hutchins, E. 1995: *Cognition in the Wild*. Cambridge, MA: MIT Press.

Itkonen, E. 2008: The central role of normativity in language and in linguistics. In J. Zlatev, T. P. Racine, C. Sinha and E. Itkonen (eds), *The Shared Mind: Perspectives on Intersubjectivity*. Amsterdam: John Benjamins.

Jack, A. and Roepstorff, A. 2002: Introspection and cognitive brain mapping: from stimulus-response to script-report. *Trends in Cognitive Sciences*, 6, 333–9.

Kay, P. and Kempton, W. 1984: What is the Sapir-Whorf hypothesis? *American Anthropologist*, 86, 65–79.

Kay, P. and Regier, T. 2007: Color naming universals: the case of Berinmo. *Cognition*, 102, 289–298.

Kemmerer, D. 2006: The semantics of space: integrating linguistic typology and cognitive neuroscience. *Neuropsychologia*, 44, 1607–21.

Kilner, J. M., Friston, K. J. and Frith, C. D. 2007: Predictive coding: an account of the mirror neuron system. *Cognitive Processing*, 8, 159–66.

Kirsch, J. A., Gunturkun, O. and Rose, J. 2008: Insight without cortex: lessons from the avian brain. *Consciousness and Cognition*, 17, 475–83.

Kuhl, P. K. 2004: Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, 5, 831–43.

Làdavas, E. 2002: Functional and dynamic properties of visual peripersonal space. *Trends in Cognitive Sciences*, 6, 17–22.

Landau, B. 2000: Concepts, the lexicon and acquisition: Fodor's new challenge. *Mind & Language*, 15, 319–26.

Landau, B. and Gleitman, L. R. 1985: *Language and Experience*. Cambridge, MA: Harvard University Press.

Leavens, D. and Racine, T. P. 2009: Joint attention in apes and humans: are humans unique? *Journal of Consciousness Studies*, 16, 240–67.

Levinson, S. 1997: Language and cognition: the cognitive consequences of spatial description in Guugu Yimithirr. *Journal of Linguistic Anthropology*, 7, 98–131.

Levinson, S. C. 2003: *Space in Language and Cognition*. Cambridge: Cambridge University Press.

Liszkowski, U., Schafer, M., Carpenter, M. and Tomasello, M. 2009: Prelinguistic infants, but not chimpanzees, communicate about absent entities. *Psychological Science*, 20, 654–60.

Maass, A., Salvi, D., Arcuri, L. and Semin, G. 1989: Language use in intergroup contexts: the linguistic intergroup bias. *Journal of Personality and Social Psychology*, 57, 981–93.

Majid, A., Bowerman, M., Kita, S., Haun, D. and Levinson, S. 2004: Can language restructure cognition? The case for space. *Trends in Cognitive Sciences*, 8(3), 108–114.

Maravita, A., Husain, M., Clarke, K. and Driver, J. 2001: Reaching with a tool extends visual-tactile interactions into far space: evidence from cross-modal extinction. *Neuropsychologia*, 39, 580–5.

Maravita, A. and Iriki, A. 2004: Tools for the body (schema). *Trends in Cognitive Sciences*, 8(2), 79–86.

Marsh, K. L., Richardson, M. J. and Schmidt, R. C. 2009: Social connection through joint action and interpersonal coordination. *Topics in Cognitive Science*, 1(2), 320–39.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J. and Alho, K. 1997: Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385(6615), 432–4.

Nelson, T. O. 1992: *Metacognition : core readings*. Boston: Allyn and Bacon.

Olsson, A. and Phelps, E. A. 2004: Learned fear of 'unseen' faces after Pavlovian, observational, and instructed fear. *Psychological Science*, 15(12), 822–8.

Pickering, M. J. and Garrod, S. 2004: Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*. 27, 169–90; discussion, 190–226.

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C. and Fried, I. 2005: Invariant visual representation by single neurons in the human brain. *Nature*, 435(7045), 1102–7.

Raczaszek-Leonardi, J. and Kelso, J. A. 2008: Reconciling symbolic and dynamic aspects of language: toward a dynamic psycholinguistics. *New Ideas in Psychology*, 26, 193–207.

Ray, V. 1952: Techniques and problems in the study of human color perception. *Southwestern Journal of Anthropology*, 8, 251–9.

Regier, T., Kay, P. and Khetarpal, N. 2007: Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences of USA*, 104, 1436–41.

Richardson, D. and Dale, R. 2005: Looking to understand: the coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science*, 29, 1045–60.

Richardson, D., Dale, R. and Kirkham, N. 2007: The art of conversation is coordination: common ground and the coupling of eye movements during dialogue. *Psychological Science*, 18(5), 407–13.

Richardson, M. J., Lopresti-Goodman, S., Mancini, M., Kay, B. and Schmidt, R. C. 2008: Comparing the attractor strength of intra- and interpersonal interlimb coordination using cross-recurrence analysis. *Neuroscience Letters*, 438, 340–5.

Richardson, M. J., Marsh, K. L. and Baron, R. M. 2007: Judging and actualizing intrapersonal and interpersonal affordances. *Journal of Experimental Psychology: Human Perception and Performance*, 33, 845–59.

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R. and Schmidt, R. C. 2007: Rocking together: dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science*, 26, 867–91.

Rizzolatti, G., Fadiga, L., Fogassi, L. and Gallese, V. 1997: The space around us. *Science*, 277(5323), 190–1.

Roberson, D., Davidoff, J., Davies, I. R. L. and Shapiro, L. R. 2005: Color categories: evidence for the cultural relativity hypothesis. *Cognitive Psychology*, 50, 378–411.

Roberson, D., Davies, I. and Davidoff, J. 2000: Color categories are not universal: replications and new evidence from a stone-age culture. *Journal of Experimental Psychology General*, 129, 369–98.

Roepstorff, A. 2004: Cellular neurosemiotics: outline of an interpretative framework. In J. Schult (ed.), *Studien zur Theorie der Biologie*. Berlin: VWB/Verlag für Wissenschaft und Bildung.

Roepstorff, A. 2008: Things to think with: words and objects as material symbols. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1499), 2049–54.

Roepstorff, A. and Frith, C. 2004: What's at the top in the top-down control of action? Script-sharing and 'top-top' control of action in cognitive experiments. *Psychological Research*, 68, 189–98.

Rosch, E. 1973: Natural categories. *Cognitive Psychology*, 4, 328–50.

Rubini, M. and Menegatti, M. 2008: Linguistic bias in personnel selection. *Journal of Language and Social Psychology*, 27, 168–181.

Sebanz, N., Bekkering, H. and Knoblich, G. 2006: Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, 10, 70–6.

Sebanz, N. and Knoblich, G. 2009: Prediction in joint action: what, when, and where. *Topics in Cognitive Science*, 1, 353–67.

Semin, G. and Fiedler, K. 1988: The cognitive functions of linguistic categories in describing persons: Social cognition and language. *Journal of Personality and Social Psychology* 54(4), 558–568.

Semin, G., Gil de Montes, L. and Valencia, J. 2003: Communication constraints on the linguistic intergroup bias. *Journal of Experimental Social Psychology*, 39, 142–8.

Shockley, K., Baker, A. A., Richardson, M. J. and Fowler, C. A. 2007: Articulatory constraints on interpersonal postural coordination. *Journal of Experimental Psychology: Human Perception and Performance*, 33, 201–8.

Shockley, K., Santana, M. V. and Fowler, C. A. 2003: Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 326–32.

Sinha, C. 2004: The evolution of language: from signals to symbols to system. In D. K. Oller and U. Griebel (eds), *Evolution of Communication Systems: A Comparative Approach*. Cambridge, MA: MIT Press.

Sinha, C. 2009: Semiotics and Cognitive Sciences. 10th World Congress of Semiotics. La Coruna, Spain.

Smith, J. D. 2009: The study of animal metacognition. *Trends in Cognitive Sciences*, 13(9), 389–96.

Sober, E. and Wilson, D. S. 1998: *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.

Spivey, M. J., Tyler, M. J., Eberhard, K. M. and Tanenhaus, M. K. 2001: Linguistically mediated visual search. *Psychological Science*, 12(4), 282–6.

Stjernfelt, F. 2007: *Diagrammetology: An Investigation on the Borderlines of Phenomenology, Ontology, and Semiotics*. Dordrecht: Springer.

Talmy, L. 2000: *Toward a Cognitive Semantics*. Cambridge, MA: MIT Press.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M. and Sedivy, J. C. 1995: Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632–4.

Tolstoy, L. 2008 [1877]: *Anna Karenina*. Oxford: Oxford University Press.

Tomasello, M. 1999: *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.

Tomasello, M. 2006: Why don't apes point? In N. J. Enfield and S. C. Levinson (eds), *Roots of Human Sociality: Culture, Cognition and Interaction*. Oxford: Berg.

Tomasello, M., Carpenter, M., Call, J., Behne, T. and Moll, H. 2005: Understanding and sharing intentions: the origins of cultural cognition. *Behavioral and Brain Sciences*, 28, 675–91; discussion, 691-735.

Treisman, A. M. and Gelade, G. 1980: A feature-integration theory of attention. *Cognitive Psychology*, 12, 97–136.

Tylén, K., Wallentin, M. and Roepstorff, A. 2009a: Say it with flowers! An fMRI study of object mediated communication. *Brain and Language*, 108, 159–66.

Tylén, K., Philipsen, J. S. and Weed, E. 2009b: Taking the language stance in a material world: a comprehension study. *Pragmatics and Cognition*, 17, 573–95.

Vygotsky, L. S. 1986: *Thought and Language*. Cambridge MA: MIT Press.

Wallentin, M. and Frith, C. D. 2008: Language is shaped for social interactions, as well as by the brain. *Behavioral and Brain Sciences*, 31, 536–7.

Wallentin, M., Roepstorff, A., Glover, R. and Burgess, N. 2006: Parallel memory systems for talking about location and age in precuneus, caudate and Broca's region. *Neuroimage*, 32(4), 1850–64.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R. and Boroditsky, L. 2007: Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Science of USA*, 104(19), 7780–5.

Wyman, E., Rakoczy, H. and Tomasello, M. 2009: Normativity and context in young children's pretend play. *Cognitive Development*, 24, 146–55.

Xu, X., Zuo, X., Wang, X. and Han, S. 2009: Do you feel my pain? Racial group membership modulates empathic neural responses. *Journal of Neuroscience*, 29(26), 8525–9.

Zlatev, J. 2008: From proto-mimesis to language: evidence from primatology and social neuroscience. *Journal of Physiology Paris*, 102, 137–51.

Zuberbuhler, K. 2008: Gaze following. *Current Biology*, 18(11), R453–5.