Mind & Society (2008) 7:77–94
DOI 10.1007/s11299-007-0044-4

ORIGINAL ARTICLE

Embodiment versus memetics

Joanna J. Bryson

Received: 17 October 2006 / Accepted: 20 September 2007 / Published online: 20 November 2007

Abstract The term *embodiment* identifies a theory that meaning and semantics cannot be captured by abstract, logical systems, but are dependent on an agent's experience derived from being situated in an environment. This theory has recently received a great deal of support in the cognitive science literature and is having significant impact in artificial intelligence. *Memetics* refers to the theory that knowledge and ideas can evolve more or less independently of their human-agent substrates. While humans provide the medium for this evolution, memetics holds that ideas can be developed without human comprehension or deliberate interference. Both theories have profound implications for the study of language—its potential use by machines, its acquisition by children and of particular relevance to this special issue, its evolution. This article links the theory of memetics to the established literature on semantic space, then examines the extent to which these memetic mechanisms might account for language independently of embodiment. It then seeks to explain the evolution of language through uniquely human cognitive capacities which facilitate memetic evolution.

Keywords Embodiment · Memetics · Semantic space · Language evolution · Cultural evolution

J. J. Bryson (✉)
Artificial Models of Natural Intelligence, Department of Computer Science,
University of Bath, BA2 7AY Bath, UK
e-mail: J.J.Bryson@bath.ac.uk

J. J. Bryson
Konrad Lorenz Institute for Evolution and Cognition Research, Adolf Lorenz Gasse 2,
3422 Altenberg, Austria

1 Introduction

There is no doubt that embodiment is a key part of human and animal intelligence. Many of the behaviours attributed to intelligence are in fact a simple physical consequence of an animal's skeletal and muscular constraints (Raibert 1986; Port and van Gelder 1995; Paul 2004). Taking a learning or planning perspective, the body can be considered as bias, constraint or (in Bayesian terms) a prior for both perception and action which facilitates an animal's search for appropriate behaviour. Since this search for expressed behaviour *is* intelligence, there can be no question that the body is a part of animal intelligence. In other words, in nature at least, autonomous behaviour and bodies have co-evolved.

The influence of the body continues, arguably through all stages of reasoning (Lakoff and Johnson 1999; Chrisley and Ziemke 2002; Steels and Belpaeme 2005), but certainly at least sometimes to the level of semantics. For example, Glenberg and Kaschak (2002) have demonstrated the *action-sentence compatibility effect*. That is, subjects take longer using a gesture to signal comprehension of a sentence about motion if the signalling gesture must be in the opposite direction as the motion indicated in the sentence. For example, given a joystick to signal an understanding of 'open the drawer', it is easier to signal comprehension by pulling the joystick towards you than by pushing it away. Boroditsky and Ramscar (2002) have similarly shown that comprehension of ambiguous temporal events is strongly influenced by the hearer's physical situation with respect to current or imagined tasks and journeys.

These sorts of effects have lead some to suggest that the reason for the to-date rather unimpressive state of natural language comprehension and production in Artificially Intelligent (AI) systems is a consequence of their lack of embodiment (Harnad 1990; Brooks and Stein 1994; Roy and Reiter 2005). The suggestion is that, in order to be meaningful, concepts must be grounded in the elements of intelligence that produce action.

The pursuit of embodied AI has lead us to understand resource-bounded reasoning which explains apparently suboptimal or inconsistent decision-making in humans (Chapman 1987). It has also helped us to understand the extent to which agents can rely on the external world as a resource for cognition—that perception can replace or at least supplement long-term memory, reasoning and model building (Brooks 1991; Clark 1997; Ballard et al. 1997; Clark and Chalmers 1998). However, despite impressive advances in the state of artificial embodiment (e.g. Chernova and Veloso 2004; Schaal et al. 2003; Kortenkamp et al. 1998), there have been no clear examples of artificial natural language systems improved by embodiment. No speech recognition, text generation or interactive tutoring system has utilised embodiment to improve its semantic performance—indeed, this idea still seems absurd. If embodiment is the key to semantics, why is it ignored by even high-end language research and entertainment systems?

I believe this is because embodiment, while certainly playing a part in both the evolution of human semantics and its development in individuals, is not in itself sufficient to explain all semantics in either context. We *have* seen neat examples of the embodied acquisition of limited semantic systems (e.g. Steels and Vogt 1997;

Steels and Kaplan 1999; Roy 1999; Billard and Dautenhahn 2000; Sidnera et al. 2005; Hawes et al. 2007). These systems demonstrate not only that a semantics can be established between embodied agents, but also the relation between the developed lexicon and the agents' physical plants and perception. However, such examples give us little idea of how words like INFINITY, SOCIAL or REPRESENT might be represented. Further, they do not show the *necessity* of physical embodiment for a human-like level of comprehension of current natural language semantics. On the other hand, if some abstract semantic system underlies our representation of words such as JUSTICE, it is possible that that the semantic processes for that system may also be sufficient for understanding terms like KICK and MOTHER which originally evolved in reference to categories learned through embodied experience. If so, this might explain why (for example) congenitally blind people use visual metaphors as naturally as those who actually see.

This article does not contest the importance of understanding embodiment to understanding human intelligence as a whole. This article *does* contest one of the prominent claims of the embodied intelligence movement—that embodiment is the only means of grounding semantics (Brooks and Stein 1994). Roy and Reiter (2005) in fact *define* the term GROUNDED as 'embodied', which might be fine (compare with Harnad 1990) if GROUNDED had not also come to be synonymous with MEANINGFUL. The central claim of this article is that while embodiment may have been the origin of most semantic meaning, it is no longer the only source for accessing a great deal of it. Further, some words (including their meanings) may have evolved more or less *independently* of grounded experience, possibly via memetic processes.

In this article, I propose a model consisting of an interaction between a disembodied, memetic semantics and embodied knowledge. The disembodied semantics is *not* the traditional, logic-based symbol system that the embodiment theory originally arose in opposition to. Rather it is another cognitively minimalist representational system similar to other well-established forms of perceptual learning. I claim that humans are capable of recognising and relating patterns in sequential streams of information and can thus derive a form of semantics from their cultural environment of speech streams.

This article begins by redefining some linguistic terms—not because they generally need redefining, but solely for local use in this article for the purpose of elucidating the model. I then review the current literature on the automatic, perception-like acquisition of semantic relationships. Next I present my model of how such a semantic system fits into human intelligence and language use. Finally I examine the implications of the model for the evolution of language.

2 Definitions and claims

Because we are in the process of trying to understand what 'semantics' and 'embodiment' mean, it follows that every article will have slightly different connotations for these and related terms. This section describes special usages of these terms for this paper. These usages are flagged by an *Embodiment vs. Memetics* subscript, e.g. semantics$_{e/m}$. I am not claiming these are the current ordinary usages of the terms;

neither do I mean to suggest they should ultimately become so. These are just usages local to this article that make it easier to state my argument clearly. Where these terms are *not* subscripted, I am referring to the more standard meanings for these words.

First, the basics:

- *semantics*$_{e/m}$: how a word is used (a fuller explanation of this is the topic of the next section.)
- *plant*$_{e/m}$: any part of an agent that might directly impact or be impacted upon by the agent's environment.
- *expressed behaviour*$_{e/m}$: behaviour that impacts the environment (including other agents), and is consequently externally observable.
- *grounded*$_{e/m}$: linked to, part of, or associated with a representation (e.g. neural encoding, sense-act pairings) that determines an expressed behaviour.
- *understand*$_{e/m}$: associate a semantic$_{e/m}$ term with a grounded$_{e/m}$ concept or category.

Embodiment$_{e/m}$ then is just having a plant$_{e/m}$. Notice that by these definitions software agents may be embodied$_{e/m}$, so long as they can impact the environment they are situated in. For example, game agents can create and destroy in their virtual worlds. A web agent would be embodied if it could change as well as respond to the content of the web. It is because this sort of agent/environment interface is so far from the conventional biological sense of 'body' that I use here the more industrial term 'plant'.

Embodiment$_{e/m}$ is really a continuum: having more and richer interactions with a richer environment clearly increases the potential for interesting grounding$_{e/m}$. Thus a virtual reality agent with complex physics and actuators such as those of Noda and Stone (2003), Tu (1999) or Maes et al. (1994) might actually be more embodied than a mobile robot with only infrared sensors and a cylindric, limb-less plant$_{e/m}$. Ziemke (1999) makes a related point that just because an agent is physically embodied (i.e. a robot) does not mean that any trouble has been taken to ground$_{e/m}$ its concepts.

By these definitions one cannot possibly understand$_{e/m}$ anything if one is not embodied$_{e/m}$. I make this definition in deference to the embodiment theorists (e.g. Cangelosi and Harnad 2001; Lakoff and Johnson 1999). But I believe semantics$_{e/m}$ has at least an equal claim to 'meaning' as conventional grounded understanding.

The main claims of my model are that:

- an agent can say useful and sensible things without understanding$_{e/m}$ what they are talking about,
- knowledge and (therefore) intelligent behaviour can be built up this way through social processes and cultural evolution, and
- these points help explain the evolution of language.

3 Semantics without reference

A basic premise of this article is that human-like semantics$_{e/m}$ can be derived without any particular plant$_{e/m}$ or embodiment$_{e/m}$. Fortunately, this is not a hypothesis, it is a demonstrated fact. The demonstrations are in the computational psycholinguistics literature (Finch 1993; Landauer and Dumais 1997; McDonald and Lowe 1998; Wiemer-Hastings 1998; Hofmann 2001; Levy and Bullinaria 2001; Padó and Lapata 2003). This cognitive-science approach uses *semantic space* to define lexical semantics.

The underlying motivation for a semantic space model is the observation that whatever the reason that two words are similar in meaning, that similarity shows up in the distributional profiles of the words with which they co-occur. That is, whether meaning is derived from grounding$_{e/m}$, memetics, logical formalism or something entirely different, the word DOG will share more linguistic contexts with CAT than with FREEDOM, and more with FREEDOM than with FLORID (e.g. "we bought a" or "I petted their" or "she loves her"). A semantic space associates each word with a 'meaning' vector. Each dimension of that vector represents the number of times some particular word occurs in a small window of words around the target word. The context words are a small number of relatively frequent words, e.g. the 75–150th most frequent words in English. If the very most frequent words are used, this method identifies primarily syntax rather than semantics. These most frequent words are generally closed-class, e.g. 'the', 'is' or 'and' (Lowe 2000).

Vectors can also be thought of as arrays of integer values. However, for the metaphor which is the semantic space model, it is more useful to think of each value in the array as describing a distance on one axis of a high dimensional space. The number of context words used defines the dimensionality of the vector. Thus the result of the above process is an object with both a length and a direction. The more distributionally similar words are (that is, the more that they are used the same way), the closer the ends of these vectors will be in this vector space. This similarity can be measured with simple linear-algebra functions. For example, the cosine function will return the similarity of the angle of two vectors.

Semantic similarity measured this way correlates tightly with ordinary human semantic similarity as measured in a wide variety of experimental paradigms, for example the reaction times in semantic priming (see Fig. 1). No other known memory model successfully predicts the choice and reaction time of human subjects using data derived purely from ambient language sources.

The fact semantics can be acquired through purely statistical methods from everyday language should not be surprising. We know, for example, that blind or paralysed individuals typically learn to use words in a manner indistinguishable from able-bodied speakers, despite lacking many of the experiences the standard grounding theory of semantics would seem to require (e.g. Landau and Gleitman 1985). We would never assume that the disabled do not know what they are talking about when they use language, including language implying a physical ability that they lack (e.g. a blind person using a visual metaphor such as "I see" to indicate understanding).

As implied in the earlier discussion of context-word choice, one interesting consequence of the semantic-space model is that syntax is no longer a separate entity from semantics$_{e/m}$. Syntax is also a part of how language is used. Syntactic categories discriminate trivially in semantic space (Redington et al. 1998; Howell and Becker 2001). *Starting* with syntax and trying to solve semantics later is barking

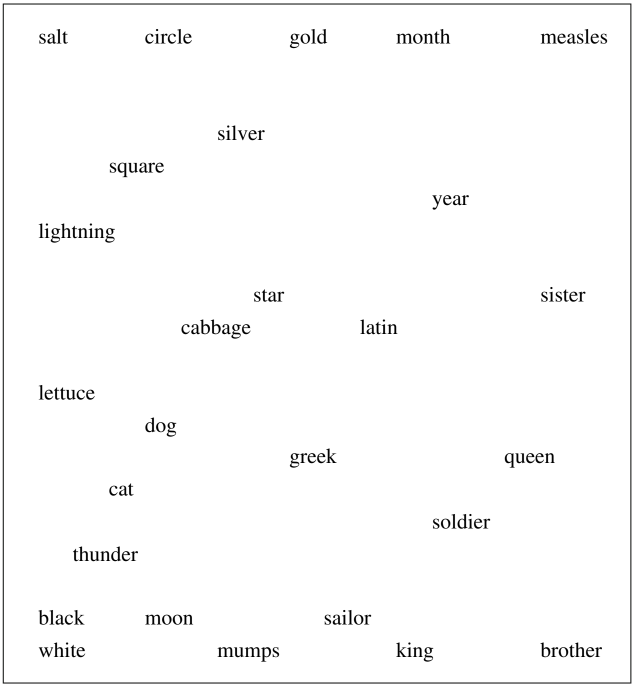


Fig. 1 A two-dimensional projection of a high-dimensional semantic space, after Lowe (1997). The target words are taken from the semantic priming experiments of Moss et al. (1995), whose results Lowe matches. Additional information on nearness is contained in the weights between locations in the 2-D space

up the wrong tree, analogous to the use of logic as a representation in AI. The use of overly syntactically based, semantically under-constrained representations is what lead to the AI frame problem in the first place (Harnad 1993; MacDorman 1999), which in turn lead to the embodiment theory of semantics. Humans and other animals use relatively constrained information in their reasoning (Todd and Gigerenzer 2000; Chater et al. 2003), a fact not surprising considering the combinatorial complexity of an under-constrained representation (Chapman 1987; Wolpert 1996; Bryson 2002). Unifying syntax and semantics is now popular in mainstream computational linguistics as well, see for example categorial grammars Steedman (1998).

4 Individual development: the semantic$_{e/m}$ species

One of the theoretical advantages of statistically acquired semantics is that it simplifies the task faced in language learning by an individual. Learning semantics$_{e/m}$ becomes an unsupervised process much like the rest of perceptual learning and development. Learning to speak becomes like learning to see: a matter of on the one hand discovering and recognising regularities, and on the other finding which regularities are salient to the agent for helping the agent to express behaviours$_{e/m}$ in appropriate times and ways in order to meet its goals. While the salience of a regularity may be encoded as a grounded$_{e/m}$ concept, there is no doubt that humans and other animals have innate predispositions for detecting statistical regularities independent of any external reward or other salience of content (Saffran et al. 1996; Smirnakis et al. 1997; Meister and Berry 1999; Sur et al. 1999; Hauser et al. 2002b; Swingley 2004).

To illustrate this claim, let me contrast my model of language learning with another more standard one (see Fig. 2). Deacon (1997) suggests that humans and some other animals (for example, dogs) are able to learn grounded$_{e/m}$ lexical concepts, and then to learn labels for these concepts through instruction. Deacon believes that what makes humans unique is our ability to then develop a web of relations between these lexicalised concepts. This allows us to perform substitutions and do other forms of symbolic reasoning. The problem with this model is that it is not clear how we learn abstract words such as JUSTICE and REPRESENTS.

What semantic space theory tells us is that humans could very well develop an interconnected web of words *independently* of the process of developing grounded$_{e/m}$ concepts. The process then of learning connections between these two independent representations is the process I refer to as understanding$_{e/m}$. Thus children can learn and use the word JUSTICE before understanding$_{e/m}$ it, perhaps mostly for singing patriotic songs. The fact a *label* for a concept exists in a lexicon can in fact help key the learner to the fact the *concept* exists. This provides an anchor for building an embodied$_{e/m}$ category simultaneously with building an understanding$_{e/m}$ (cf. Davidson 2005; Steels 2003; Vogt 2005). Gradually as the learner gains experience of the complexity of conflicting social goals and notions of fairness, they develop a richer notion of both what JUSTICE means semantically$_{e/m}$ and how and when both the word and the grounded$_{e/m}$ concept can be used in behaviour$_{e/m}$ that furthers their goals.

I want to be clear here: in my model, humans still acquire associations between these two representations, just as in Deacon's model.[1] Embodiment is still very important in my model of human intelligence. The differences from Deacon's model are:

- the ordering. In my model, lexical semantics$_{e/m}$ is learned in parallel with grounded$_{e/m}$ categories and expressed behaviour$_{e/m}$. Some semantic$_{e/m}$ lexicon may even help guide grounded$_{e/m}$ category learning. And perhaps *sometimes* it works the conventional way around—grounded$_{e/m}$ categories lead to word discovery and thus immediate understanding$_{e/m}$.
- the connectedness. Only *some* words become grounded$_{e/m}$ as connections are formed between the semantic$_{e/m}$ and embodied$_{e/m}$ representations (see Fig. 2). Not every word in the lexicon needs to be grounded$_{e/m}$ or fully understood$_{e/m}$ in order to be used. Semantics$_{e/m}$ in itself includes enough information to allow a word to generally be used correctly.

[1] Though note that the grounded$_{e/m}$ behavioural lexicon is almost certainly also modular, so there are probably more than two sets of representations becoming linked.

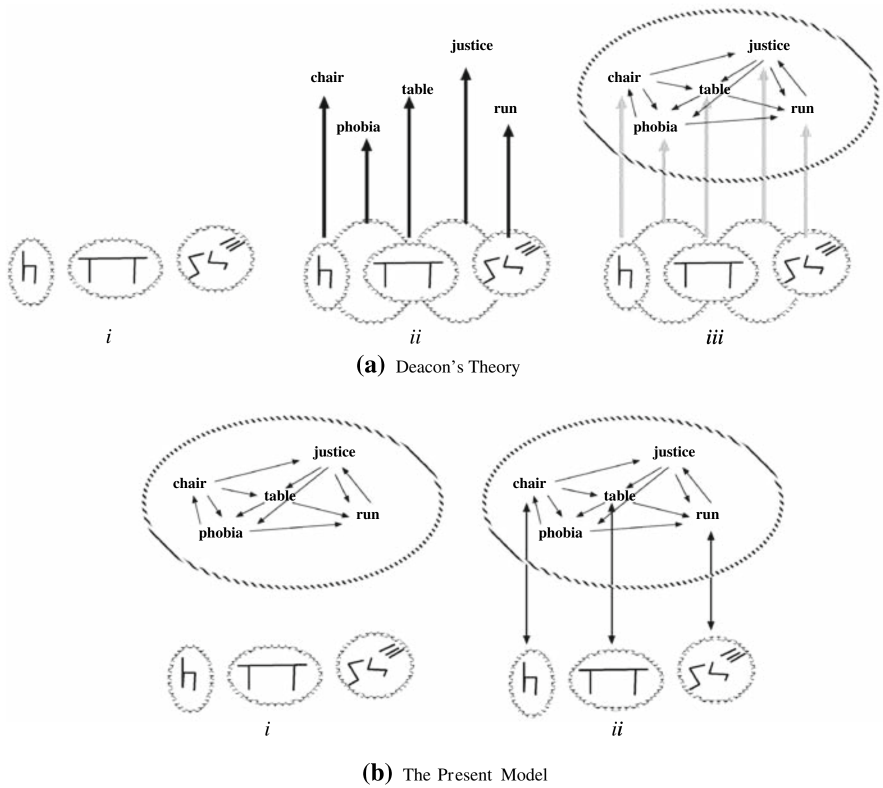


Fig. 2 In Deacon's theory, first concepts are learned [**a**(**i**)], then labels for these concepts [**a**(**ii**)], then a symbolic network somewhat like semantics$_{e/m}$ [**a**(**iii**)]. I propose instead that grounded$_{e/m}$ concepts and semantics$_{e/m}$ are learned in parallel [**b**(**i**)], with some semantic$_{e/m}$ terms becoming understood$_{e/m}$ [**b**(**ii**)]

Understanding$_{e/m}$—the linking between semantic$_{e/m}$ categories and grounded$_{e/m}$ conceptualisations—does not have to be all-or-nothing. A speaker may understand some of the implications of a word without necessarily understanding all of them. They might also be able to express verbally some aspects of their embodied experience, but not all of it.

Nothing in this model undermines the rich theories of embodied intelligence that have been developed in recent decades, notably by Lakoff and Johnson (1980, 1999). For example, Lakoff and Johnson propose that we spend a great deal of time developing the grounded$_{e/m}$ concept of a PATH as infants learning how to achieve a goal. We then use this concept as metaphor which allows us to understand notions like CAREER. My model actually facilitates this theory by explaining how it might work. When a child acquires its web of lexical terms, CAREER and PATH will be close-linked neighbours in the semantic$_{e/m}$ space precisely because we do use the path metaphor frequently in our spoken expressions about career. Thus when the child learns to understand$_{e/m}$ PATH, the linked grounded$_{e/m}$ concept of PATH and its associated expressed behaviours$_{e/m}$ for following and seeking will now also be (slightly more weakly) associated with CAREER. Thus the metaphor is a meme which the child can acquire socially, possibly prior to gaining a full understanding$_{e/m}$ of either PATH or CAREER.

To summarise my argument so far, I have proposed a model of language use based on a combination of memetically acquired semantics$_{e/m}$ and grounded$_{e/m}$, embodied$_{e/m}$ understanding$_{e/m}$. The advantages of this model over more standard ones are:

- it better explains how abstract lexical terms are learned,
- it provides a common representation for all kinds of lexical semantics, and
- it enables a rich representational substrate for insight and analogy, as new linkages can be formed between representations (see also Simon 1995).

In the final section of this article, I return to the main topic of this special issue, which is the evolution of language. My claim is that this model could also support some of the most promising recent theories concerning the evolution of language. However, doing so requires extending it to include data that is so far not as well supported as statistical semantics.

5 Memetics and the evolution of language

As readers familiar with the study of the evolution of language will know, there are two aspects of the question. First, how or why did language evolve in humans, and second, why did it *only* evolve in humans? I take the side of those who argue that language is obviously useful for both communication and thought (e.g. Pinker and Bloom 1990; Nowak et al. 2000; Castro et al. 2004; see Pinker and Jackendoff 2005 for a review). If language is useful, then *why* humans have language is "because it is adaptive". The problem of the evolution of language is therefore reduced to explaining the proximate mechanisms of its evolution in humans. Further, these mechanisms must be unique enough to humans to explain why language has not occurred in other species.

I am happy to take this position for two reasons: first, in other work, Čače and I have already demonstrated that even costly communication is adaptive in the context in which language probably evolved (Čače and Bryson 2007). Second, explaining why other species have not *evolved* language is presumably not much harder than the already necessary task of explaining why other species cannot *learn* language now that it already exists.

I believe that there is actually no single mechanism unique to humans that facilitated language evolution, but rather the unique coincidence of two cognitive representational capacities (see Fig. 3).

5.1 Memetic substrate

First, cultural evolution requires a representational substrate, one capable of recording for extraction the sorts of statistical regularities necessary for the

	2^{nd}-order social representation	no 2^{nd}-order representation
vocal imitation	people	birds
no vocal imitation	other advanced primates	most animals

Fig. 3 Human-like cultural evolution might require both a rich memetic substrate such as provided by vocal imitation, and the capacity for second order social representations

semantic-space construction described earlier. Think for a minute about standard genetic evolution. It took billions of years for replicators to evolve of sufficient complexity and redundancy to reliably record essential mechanisms for life while still allowing unsupervised systems of mutation and crossover to provide sufficient variation and innovation to support the current, relatively recent explosion in diversity and complexity of phenotype. Memetics also requires a complex, redundant representational substrate.

Humans have an ability to precisely remember richly detailed sequences of temporal events of up to 2 or 3 s (Pöppel 1994). These sequences I propose provides a memetic substrate for cultural evolution. Humans are the only species of primate capable of precise auditory replicative imitation (Fitch 2000; Poole et al. 2005; Bispham 2006). Presumably, this special capacity evolved by means of sexual selection, as it seems to have in a variety of bird species and in whales (Vaneechoutte and Skoyles 1998). Vocal imitation implies the capacity for temporally ordered representations of complex, information-rich stimuli. Auditory phrases contain ordered information on a large number of axes, including volume, duration, pitch and timbre.

These results extend to manual as well as acoustic gestures, my hypothesis is agnostic to the gestural versus vocal origin theories of the evolution of language. Indeed, *any* stimuli could become associated with a $\text{semantic}_{e/m}$ entity—a fact that may explain the Glenberg and Robertson (2000) and the Boroditsky and Ramscar (2002) results mentioned in the introduction, as well as other results on implicit semantic binding (e.g. Greenwald and Banaji 1995). If vocal and manual gestures are both present in a ‘speech’ stream, then we have another possible source of both variation and redundancy. Extra redundancy not only provides more robustness in the face of unsupervised innovation techniques, it also improves learnability (Christoudias et al. 2006).

The first cognitive representational capacity I am postulating then is temporally precise behavioural ‘scripts’ with which we can absorb enormous amounts of information from other individuals. This provides the capacity for distinctive and robust (through redundancy) features which can in turn support the evolution of $\text{semantic}_{e/m}$ units. For a description of how such a system could then lead to language as we know it, see Wray (1998), who describes semantic units evolving from extended call phrases. Kirby (1999) offers a demonstration of the evolution of languages expressing elements of Chomsky’s ‘universal grammar’ on simple, resource-limited agents (cf. Zuidema 2005). Note that Kirby’s systems assume some recursive mechanism to enable compositionality, which relates to the second cognitive representational capacity I am postulating, which I address below.

Now that language has evolved, it is possible for it to be represented on a less rich and redundant substrate, such as the one you are currently reading. The vast support of redundancy and diversity needed for human *cultural* evolution is now provided by language itself as the substrate, in addition to many other expressed behaviours$_{e/m}$ and resulting artifacts. But before the existence of language, another substrate had to exist to support *language's* evolution.

This is not to say that language was necessarily the *very* first part of human memetic evolution. We know that a variety of other species do acquire culture (van Schaik et al. 2003; Perry et al. 2003; Perry and Manson 2003; de Waal and Johanowicz 1993; Whiten et al. 1999). By 'culture' I mean behaviour that is reliably and consistently transmitted between individuals by non-genetic means (cf. Bryson and Wood 2005). It may be that the human capacity for temporally precise imitation helped our species far before true language evolved. If we were more likely to know about a greater variety of feeding opportunities, then we could have had a denser population which in turn would increase the probability of acquiring yet further knowledge (Čače and Bryson 2007). In all probability, larger, denser populations than we see currently in non-human apes needed to exist in order for something as elaborate as language to evolve memetically. However, now that language is here, the cultural-evolution process operates on entirely new levels.

5.2 Memetic structure

Having explained why humans differ from other primates, it remains to explain why other species capable of vocal imitation such as birds have not experienced human-like memetic explosion. Perhaps these species miss another computational tool: the capability to form second-order representations.

We as primates have the ability to represent relations *between other agents* (see Fig. 4). Although all social species behave as if they keep records of relations between themselves and their group members (e.g. positive and negative interactions), only primates behave as though they keep tabs on the relations between group members other than themselves (Harcourt 1992; Kudo and Dunbar 2001; though see Connor et al. 2001; Wittemyer et al. 2005). For example, apes will avoid fighting with close associates of dominant animals, and may try to befriend them (de Waal 1996).

For some time, the complexity of primate societies has been hypothesised as the root cause of primate intelligence (Byrne and Whiten 1988; Dunbar 1995). But I am making a very specific claim: that being able to keep track of second-order relationships may have been the intellectual key that enables *compositionality*. Compositionality is a key attribute of language, which means essentially that big semantic chunks can be composed out of smaller ones. In computational terms, compositionality enables a powerful sort of recursive structure which is key to all the computations enabled by context-free grammars (Chomsky 1962).

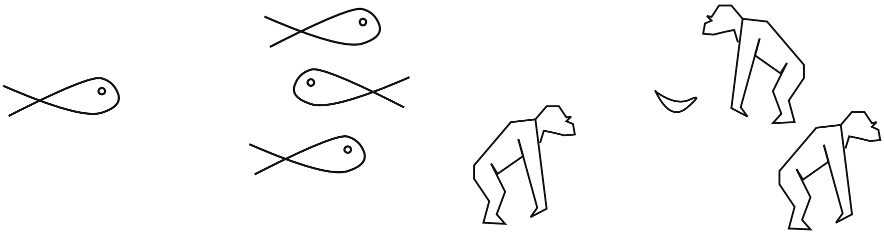
If the link appears tenacious between compositionality and the ability to represent relationships between other group members, consider Fig. 4. Representing tit-for-tat requires only a simple list of associations, with very little state beyond that

	ego
Roy	5
Thelma	2
Eunice	7
Harry	-1

(a) 1^{st}-Order Social Representation

	Roy	Thelma	Eunice	Harry
Roy	–	5	2	-4
Thelma	7	–	8	4
Eunice	-3	8	–	4
Harry	-1	3	5	–

(b) 2^{nd}-Order Social Representation



(c) Guppies **(d)** Primates

Fig. 4 Even guppies seem to keep track of the reliability of former partners (**a**) in risky team behaviour. But only primates compose relationships between self and other (**a**) with relationships between other and other (**b**) when considering social action. (Names are for clarity and are not intended to imply use of symbols by either species.)

needed to recognise individual groupmates. But having a *composable* representation of relationships between other agents requires not only significantly more state, but also the ability to combine elements of that information. Combination is necessary in order to assess the relation between one's own rank or abilities and those of a relevant other. In the presence of a powerful likely collaborator, the other is a different sort of opponent than when encountered alone, or if encountered when the other has recently lost the good will of the potential collaborator. Thus the agent is considering its relationship not just to the other, but also the relationships between self, other, and all other troop-mates present. This is an example of a second-order representation. We might refer to it as *compositional politics*, since the current relative rank between the self and the other is composed of both the agents' absolute ranks and also the ranks of other present agents as weighted by the probability of those agents intervening, which is in turn dependent on the other agents' relationships.

Besides language, compositionality or second-order referencing may also to be key to other cognitive capacities (e.g. Greenfield 1991; Iriki 2006). Indeed, compositionality is a sufficiently useful intellectual skill that it may well have evolved first for some other purpose than social structure; if so this does not significantly alter the main claims of this article. Note for example that this part of my model is similar to the Hauser et al. (2002a) claim about the importance of recursion. The most important difference between these models is not the (fairly small) difference between recursion and second-order representations. It is that the second-order capacity enabling compositional structure would predate language phylogenetically. Also, it is *not* the key difference between us and other primates, although it is the key difference between us and birds (see Fig. 3). The difference between us and other primates is precise temporal imitation, which provides the necessary substrate for representing sufficient cultural information for memetics to operate on a large scale.

5.3 Summary and Implications

My hypothesis then is that our memetic capacities, not our embodiment, have allowed humanity to evolve our uniquely elaborate culture. While our embodiment is unquestionably a part of our intelligence and therefore our language, it is memetics that accounts for the origins of that language. I have hypothesised that what makes us a uniquely memetic species is the conjunction of

- our capacity for precise temporal imitation, which probably has its origins in sexually selected vocal imitation, and which provides us a rich substrate for information transmission; and
- our capacity for second-order representations, which may have its origins in our social reasoning, and which gives our language the computational power of compositionality.

If our memetic representation is a more fertile substrate for supporting unsupervised cultural evolution, then our culture has a richer design space in which to evolve. Darwinian evolution is a form of unsupervised learning—innovation is not necessarily understood or even deliberate, and therefore cannot be done carefully. Thus the substrate itself must support the evolution of protective representational structural mechanisms, such as redundancy, which protect vital structures from unsupervised innovation (Goldberg et al. 1989; Weicker and Weicker 2001; Miglino and Walker 2002). At the same time, evolution also requires a large number of axes of variation, so that new lexical entries can be easily recognisable and expressed unambiguously.

This is not to say that memetic evolution is in every way identical to genetic evolution. This article has stayed carefully agnostic about what constitutes a meme, but note that what *exactly* constitutes a *gene* is also still undetermined, and possibly undeterminable.[2] What we need to know about social transmission is only that *something*, some behaviour, is replicated. Given replication, the only things further required by Darwinian evolution are variation and selection.

The differences between cultural and biological evolution worth discussing then are not about the nature of the meme itself, but rather about the evolutionary processes that exploit it. Unlike the case of genetic evolution, memetic evolution is not *entirely* reliant on random chance for innovation. 'Mutation' in human culture does not come exclusively from copying errors. Humans and other primates change their expressed behaviour as a consequence of their own intelligence and experience—that is, as a consequence of local search conducted by the individual

[2] "It is not easy, indeed it may not even be meaningful, to decide where one gene ends and the next one begins. Fortunately...this does not matter..." (Dawkins 1976, p. 22). See also Dennett (2002).

agent. If something works well and/or the individual who innovated gains higher social status, lives longer or has a larger number of children, then the new behaviour has an increased probability of being passed to others.

Recombination is also different in memetics. Rather than just representing the impact of a mix of two genomes at conception, an individual's culturally acquired behaviour may represent some sort of weighted average of observed behaviours expressed by a large number of conspecifics, or even individuals of other species. These are two reasons why memetic evolution can work so much faster than genetic, and a cultural artifact as elaborate even as language can appear in a relatively short period of time.

6 Conclusions and future work

This article has presented a fusion of two radical—and apparently opposing—theories of semantics: embodiment and memetics. I began by reviewing studies showing that human-like semantics can be acquired purely through latent learning from ambient language. I then presented a model whereby we acquire this knowledge through implicit statistical learning and regularity discovery, and then use this information in concert with the knowledge and categories we discover through physical interactions with the world. This model makes language acquisition less of a special case and more like our other perceptual skills.[3] Forming links between our embodied or 'grounded$_{e/m}$' understanding and our abstract semantic$_{e/m}$ lexicon not only helps us learn to use language, but also provides mutual information that assists concept discovery in both domains.

Finally, I discussed the implications for such a system in accounts of the evolution of language. I suggested that the reason only humans have stumbled on this useful mechanism for ratcheting our culture is because we are unique in having both the capacity for second-order representations (derived from our social capacities as primates) and in having an adequately rich memetic substrate for boot-strapping the evolution of language. This latter derives from our representational capacities as vocal imitators to remember precise and detailed sequences of information for phrases of 2–3 s.

The obvious next step in trying to study and develop this model is to build a working version. This applies both to the idea of having a semantically capable individual with limited embodiment, and to trying to replicate early language evolution. Both problems would take immense computational power; however, we may be approaching the stage where such research is tractable. In fact, search engines already exploit the sort of statistical semantics I have described here (Lowe 2000). It may well be that our culture has evolved to contain so much information that semantic agents with almost no similarity in embodiment to our own could participate in our culture, and contribute to our own knowledge and behaviour. If so, this bodes well for the quest for strong AI.

[3] See also Davidson (2005) for an interesting though somewhat orthogonal discussion of the implications of language as perception.

Acknowledgments This article would not exist if not for many long discussions and a little text from Will Lowe. Not that he necessarily agrees with me. Thanks also for extensive comments on earlier drafts to Gert Westermann, and Mark Johnson (who most definitely does not agree with me), and for the comments of the anonymous reviewers. Earlier versions of this article have been presented to the Evolution of Language conference (Leipzig 2004), to Post-Modular Psychology (Glasgow 2005) and to the AI groups at the universities of Plymouth and Utrecht. Thanks to these institutions also for comments and contributions.

References

Ballard DH, Hayhoe MM, Pook PK, Rao RPN (1997) Deictic codes for the embodiment of cognition. Brain Behav Sci 20:723–742

Billard A, Dautenhahn K (2000) Experiments in social robotics: grounding and use of communication in autonomous agents. Adapt Behav 7:415–438

Bispham J (2006) Rhythm in music: what is it? Who has it? And why? Music Percept 24:125–134

Boroditsky L, Ramscar M (2002) The roles of body and mind in abstract thought. Psychol Sci 13:185–188

Brooks RA (1991) Intelligence without representation. Artif Intell 47:139–159

Brooks RA, Stein LA (1994) Building brains for bodies. Auton Robots 1:7–25

Bryson JJ (2002) Language isn't quite *that* special. Brain Behav Sci 25:679–680. Commentary on Carruthers,"The Cognitive Functions of Language", same volume

Bryson JJ, Wood MA (2005) Learning discretely: behaviour and organisation in social learning. In: Demiris Y (ed) Third international symposium on imitation in animals and artifacts. The Society for the Study of Artificial Intelligence and the Simulation of Behaviour, Hatfield, pp 30–37

Byrne RW, Whiten A (eds) (1988) Machiavellian intelligence: social expertise and the evolution of intellect in monkeys, apes and humans. Oxford University Press, New York

Čače I, Bryson JJ (2007) Agent based modelling of communication costs: why information can be free. In: Lyon C, Nehaniv CL, Cangelosi A (eds) Emergence and evolution of linguistic communication. Springer, London, pp 305–322

Cangelosi A, Harnad S (2001) The adaptive advantage of symbolic theft over sensorimotor toil: grounding language in perceptual categories. Evol Commun 4:117–142

Castro L, Medina A, Toro MA (2004) Hominid cultural transmission and the evolution of language. Biol Philos 19:721–737

Chapman D (1987) Planning for conjunctive goals. Artif Intell 32:333–378

Chater N, Oaksford M, Nakisa R, Redington M (2003) Fast, frugal, and rational: how rational norms explain behavior. Organ Behav Hum Decis Process 90:63–86

Chernova S, Veloso M (2004) Learning and using models of kicking motions for legged robots. In: The Proceedings of ICRA-2004, New Orleans

Chomsky N (1962) Context-free grammars and pushdown storage. Technical Report 65 MIT RLE Quarterly Progress Report, pp 187–194

Chrisley R, Ziemke T (2002) Embodiment. In: Encyclopedia of cognitive science. Macmillan, New York, pp 1102–1108

Christoudias CM, Saenko K, Morency LP, Darrell T (2006) Co-adaptation of audio-visual speech and gesture classifiers. In: ICMI '06: Proceedings of the 8th international conference on multimodal interfaces. ACM Press, London

Clark A (1997) Being there: putting brain, body and world together again. MIT Press, Cambridge

Clark A, Chalmers D (1998) The extended mind. Analysis 58:7–19

Connor RC, Heithaus MR, Barre LM (2001) Complex social structure, alliance stability and mating access in a bottlenose dolphin 'super-alliance'. Proc R Soc B 268:263–267

Davidson D (2005) Seeing through language. In: Preston JM (ed) Truth, language and history, Chap 9. Clarendon Press, Oxford, pp 127–141. Originally published in *Thought and Language*. Cambridge University Press, 1997

Dawkins R (1976) The selfish gene. Oxford University Press, New York. Page numbers from the 1986 revised edition

Deacon T (1997) The symbolic species: the co-evolution of language and the human brain. W. W. Norton & Company, New York

Dennett DC (2002) The new replicators. In: Pagel M (ed) The Encyclopedia of Evolution, vol 1. Oxford University Press, Oxford, pp E83–E92
Dunbar RIM (1995) Neocortex size and group size in primates: a test of the hypothesis. J Hum Evol 28:287–296
Finch S (1993) Finding structure in language. Ph.D. thesis, Centre for Cognitive Science, University of Edinburgh
Fitch WT (2000) The evolution of speech: a comparative review. Trends Cogn Sci 4:258–267
Glenberg AM, Kaschak MP (2002) Grounding language in action. Psychon Bull Rev 9:558–565
Glenberg AM, Robertson DA (2000) Symbol grounding and meaning: a comparison of high-dimensional and embodied theories of meaning. J Mem Lang 43:379–401
Goldberg D, Korb B, Deb K (1989) Messy genetic algorithms: motivation, analysis and first results. Complex Syst 3:493–530
Greenfield PM (1991) Language, tools and brain: the ontogeny and phylogeny of hierarchically organized sequential behavior. Brain Behav Sci 14:531–595
Greenwald AG, Banaji MR (1995) Implicit social cognition: attitudes, self-esteem, and stereotypes. Psychol Rev 102:4–27
Harcourt AH (1992) Coalitions and alliances: are primates more complex than non-primates? In: Harcourt AH, de Waal FBM (eds) Coalitions and alliances in humans and other animals, Chap 16. Oxford, pp 445–472
Harnad S (1990) The symbol grounding problem. Physica D 42:335–346
Harnad S (1993) Problems, problems: the frame problem as a symptom of the symbol grounding problem. Psycholoquy 4
Hauser MD, Chomsky N, Fitch WT (2002a) The faculty of language: what is it, who has it, and how did it evolve? Science 298:1569
Hauser MD, Weiss D, Marcus G (2002b) Rule learning by cotton-top tamarins. Cognition 86:B15–B22
Hawes N, Sloman A, Wyatt J, Zillich M, Jacobsson H, Kruijff G-J, Brenner M, Berginc G, Skočaj D (2007) Towards an integrated robot with multiple cognitive functions. In: Proceedings of the twenty-second conference on artificial intelligence (AAAI-07), pp 1548–1553
Hofmann T (2001) Unsupervised learning by probabilistic latent semantic analysis. Mach Learn 42:177–196
Howell SR, Becker S (2001) Modelling language acquisition: grammar from the lexicon? In: Proceedings of the 23rd annual meeting of the cognitive science society. Lawrence Erlbaum Associates, Edinburgh
Iriki A (2006) The neural origins and implications of limitation, mirror neurons and tool use. Curr Opin Neurobiol 16:660–667
Kirby S (1999) Function, selection and innateness: the emergence of language universals. Oxford University Press, New York
Kortenkamp D, Bonasso RP, Murphy R (eds) (1998) Artificial intelligence and mobile robots: case studies of successful robot systems. MIT Press, Cambridge
Kudo H, Dunbar RIM (2001) Neocortex size and social network size in primates. Anim Behav 62:711–722
Lakoff G, Johnson M (1980) Metaphors we live by. University of Chicago Press, Chicago
Lakoff G, Johnson M (1999) Philosophy in the flesh: the embodied mind and its challenge to Western thought. Basic Books, New York
Landau B, Gleitman LR (1985) Language and experience: evidence from the blind child. Harvard University Press, Cambridge
Landauer TK, Dumais ST (1997) A solution to Plato's problem: the latent semantic analysis theory of induction and representation of knowledge. Psychol Rev 104:211–240
Levy JP, Bullinaria JA (2001) Learning lexical properties from word usage patterns. In: French RM, Sougné J (eds) Connectionist models of learning development and evolution: proceedings of the 6th neural computation and psychology workshop (NCPW 6). Springer, Heidelberg, pp 273–282
Lowe W (1997) Semantic representation and priming in a self-organizing lexicon. In: Bullinaria JA, Glasspool DW, Houghton G (eds) Proceedings of the fourth neural computation and psychology workshop: connectionist representations (NCPW 4). Springer, London, pp 227–239
Lowe W (2000) Topographic maps of semantic space. Ph.D. thesis, Institute for Adaptive and Neural Computation, Division of Informatics, University of Edinburgh
MacDorman KF (1999) Grounding symbols through sensorimotor integration. J Robot Soc Jpn 17:20–24

Maes P, Darrell T, Blumberg B, Pentland A (1994) ALIVE: artificial life interactive video environment. In: Proceedings of the twelfth national conference on artificial intelligence (AAAI-94). AAAI Press, Seattle

McDonald S, Lowe W (1998) Modelling functional priming and the associative boost. In: Gernsbacher MA, Derry SD (eds) Proceedings of the 20th annual meeting of the cognitive science society. Lawrence Erlbaum Associates, New Jersey, pp 675–680

Meister M, Berry MJ II (1999) The neural code of the retina. Neuron 22:435–450

Miglino O, Walker R (2002) Genetic redundancy in evolving populations of simulated robots. Artif Life 8:265–277

Moss HE, Ostrin RK, Tyler LK, Marslen-Wilson WD (1995) Accessing different types of lexical semantic information: evidence from priming. J Exp Psychol Learn Mem Cogn 21:863–883

Noda I, Stone P (2003) The RoboCup soccer server and CMUnited clients: implemented infrastructure for MAS research. Auton Agents Multi Agent Syst 7:101–120

Nowak MA, Plotkin JB, Jansen VAA (2000) The evolution of syntactic communication. Nature 404:495–468

Padó S, Lapata M (2003) Constructing semantic space models from parsed corpora. In: Proceedings of the 41st annual meeting of the association for computational linguistics (ACL '03), Sapporo, pp 128–135

Paul C (2004) Investigation of morphology and control in biped locomotion. Ph.D. thesis, Department of Computer Science, University of Zurich Switzerland

Perry S, Manson JH (2003) Traditions in monkeys. Evol Anthropol 12:71–81

Perry S, Baker M, Fedigan L, Gros-Louis J, Jack K, MacKinnon K, Manson J, Panger M, Pyle K, Rose L (2003) Social conventions in wild white-faced capuchin monkeys: evidence for traditions in a neotropical primate. Curr Anthropol 44:241–268

Pinker S, Bloom P (1990) Natural language and natural selection. Behav Brain Sci 13:707–784

Pinker S, Jackendoff R (2005) The faculty of language: what's special about it? Cognition 95:201–236

Poole JH, Tyack PL, Stoeger-Horwath AS, Watwood S (2005) Elephants are capable of vocal learning. Nature 434:455–456

Pöppel E (1994) Temporal mechanisms in perception. Int Rev Neurobiol 37:185–202

Port RF, van Gelder T (eds) (1995) Mind as motion: explorations in the dynamics of cognition. MIT Press, Cambridge

Raibert MH (1986) Legged robots that balance. MIT Press, Cambridge

Redington M, Chater N, Finch S (1998) Distributional information: a powerful cue for acquiring syntactic categories. Cogn Sci 22:425–469

Roy DK (1999) Learning from sights and sounds: a computational model. Ph.D. thesis, MIT, Media Laboratory

Roy D, Reiter E (2005) Connecting language to the world. Artif Intell 167:1–12

Saffran JR, Aslin RN, Newport EL (1996) Statistical learning by 8-month-old infants. Science 274:1926–1928

Schaal S, Ijspeert A, Billard A (2003) Computational approaches to motor learning by imitation. Philos Trans R Soc Lond B 358:537–547

Sidnera CL, Leea C, Kiddb CD, Lesha N, Richa C (2005) Explorations in engagement for humans and robots. Artif Intell 166:140–164

Simon HA (1995) Explaining the ineffable: AI on the topics of intuition, insight and inspiration. In: Proceedings of the 14th international joint conference on artificial intelligence. Morgan Kaufmann, Montreal

Smirnakis SM, Berry MJ, Warland DK, Bialek W, Meister M (1997) Adaptation of retinal processing to image contrast and spatial scale. Nature 386:69–73

Steedman M (1998) Categorial grammar. In: The MIT encycylopedia of cognitive sciences. MIT Press, Cambridge

Steels L (2003) Evolving grounded communication for robots. Trends Cogn Sci 7:308–312

Steels L, Belpaeme T (2005) Coordinating perceptually grounded categories through language: a case study for colour. Behav Brain Sci 28:469–489

Steels L, Kaplan F (1999) Bootstrapping grounded word semantics. In: Briscoe T (ed) Linguistic evolution through language acquisition: formal and computational models. Cambridge University Press, Cambridge

Steels L, Vogt P (1997) Grounding adaptive language games in robotic agents. In: Husbands C, Harvey I (eds) Proceedings of the fourth European conference on artificial life (ECAL97). MIT Press, London

Sur M, Angelucci A, Sharma J (1999) Rewiring cortex: the role of patterned activity in development and plasticity of neocortical circuits. J Neurobiol 41:33–43

Swingley D (2004) Statistical clustering and the contents of the infant vocabulary. Cogn Pscyhol 50:86–132

Todd PM, Gigerenzer G (2000) Simple heuristics that make us smart. Brain Behav Sci 23:727–741

Tu X (1999) Artificial animals for computer animation: biomechanics, locomotion, perception and behavior. Springer, Heidelberg

Vaneechoutte M, Skoyles J (1998) The memetic origin of language: modern humans as musical primates. J Memet Evol Models Inf Transm 2

van Schaik CP, Ancrenaz M, Borgen G, Galdikas B, Knott CD, Singleton I, Suzuki A, Utami SS, Merrill M (2003) Orangutan cultures and the evolution of material culture. Science 299:102–105

Vogt P (2005) The emergence of compositional structures in perceptually grounded language games. Artif Intell 167:206–242

Waal FBM (1996) Good natured: the origins of right and wrong in humans and other animals. Harvard University Press, Cambridge

Waal FBM, Johanowicz DL (1993) Modification of reconciliation behavior through social experience: an experiment with two macaque species. Child Dev 64:897–908

Weicker K, Weicker N (2001) Burden and benefits of redundancy. In: Martin W, Spears W (eds) Foundations of genetic algorithms 6. Morgan Kaufmann, San Francisco, pp 313–333

Whiten A, Goodall J, McGew WC, Nishida T, Reynolds V, Sugiyama Y, Tutin CEG, Wrangham RW, Boesch C (1999) Cultures in chimpanzees. Nature 399:682–685

Wiemer-Hastings K (1998) Abstract noun classification: using a neural network to match word context and word meaning. Behav Res Methods Instrum Comput 30:264–271

Wittemyer G, Douglas-Hamilton I, Getz WM (2005) The socioecology of elephants: analysis of the processes creating multitiered social structures. Anim Behav 69:1357–1371

Wolpert DH (1996) The lack of a priori distinctions between learning algorithms. Neural Comput 8:1341–1390

Wray A (1998) Protolanguage as a holistic system for social interaction. Lang Commun 18:47–67

Ziemke T (1999) Rethinking grounding. In: Riegler A, Peschl M, Von Stein A (eds) Understanding representation in the cognitive sciences. Plenum Press, New York, pp 177–190

Zuidema WH (2005) The major transitions in the evolution of language. Ph.D. thesis, Theoretical and Applied Linguistics, University of Edinburgh, United Kingdom