Topics in Cognitive Science 1 (2009) 523–546

ISSN: 1756-8757 print / 1756-8765 online
DOI: 10.1111/j.1756-8765.2009.01033.x

Modeling the Emergence of Language as an Embodied Collective Cognitive Activity

Edwin Hutchins, Christine M. Johnson

Department of Cognitive Science, University of California, San Diego

Received 16 July 2008; received in revised form 1 April 2009; accepted 21 April 2009

Abstract

Two decades of attempts to model the emergence of language as a collective cognitive activity have demonstrated a number of principles that might have been part of the historical process that led to language. Several models have demonstrated the emergence of structure in a symbolic medium, but none has demonstrated the emergence of the capacity for symbolic representation. The current shift in cognitive science toward theoretical frameworks based on embodiment is already furnishing computational models with additional mechanisms relevant to the emergence of symbolic language. An analysis of embodied interaction among captive, but not human-enculturated, bonobo chimpanzees reveals a number of additional features of embodiment that are relevant to the emergence of symbolic language, but that have not yet been explored in computational simulation models; for example, complementarity of action in addition to imitation, iconic in addition to indexical gesture, coordination among multiple sensory and perceptual modalities, and the orchestration of intra- and inter-individual motor coordination. The bonobos provide an evolutionarily plausible intermediate stage in the development of symbolic expression that can inform efforts to model the emergence of symbolic language.

Keywords: Computational simulation models; Emergence of language; Embodied cognition; Bonobo chimpanzee

1. Introduction

Language and language-like behavior are canonical examples of collective cognitive activity. Language is a collective activity both in the sense that learning a language is an entry point into the collective cognitive life of a community, and because shared language

Correspondence should be sent to Edwin Hutchins, Department of Cognitive Science, University of California San Diego, 9500 Gilman Drive, La Jolla CA 92093-0515. E-mail: ehutchins@ucsd.edu

must have emerged over time from collective activity. No individual has ever invented a natural language, but virtually all humans learn to speak one or more in a lifetime. Speaking both requires thinking and can be seen as a form of thought. The principal goal of attempts to construct computational models of the emergence of language is to shed light on the kinds of processes that may have led to the development of such phenomena as shared lexicons and grammars in the history of the human species. As Hutchins and Hazlehurst described it:

> In this line of work, researchers ask, ''What sort of process could lead to the development of a shared language?'' Clearly, some historical process led human ancestors from the condition in which there was no shared language to the condition in which a shared language exists. It is assumed that language, and many other aspects of culture, develop without any central control. That is, there could not have been a ''teacher'' who knew the language first and then taught it to others. Rather, a shared language should be expected to emerge somehow from the interactions among the members of a community who must communicate. Since we have no direct access to the historical events that led to the development of language, a common strategy for addressing this question is to construct a computational simulation model. Such a model begins in a state in which a shared language clearly does not exist. The model is then run and eventually reaches a state in which a shared language does exist (Hutchins & Hazlehurst, 2002).

It is important to stress at the outset that such models can reveal possible evolutionary trajectories, especially ones that had not previously been considered, but they cannot explain why one species followed a particular trajectory and another did not.

Modern humans are prodigious symbol processors. But we are also embodied beings. How do we explain the origin of symbolic communication and symbolic thought in a species whose thinking continues to this day to be embodied and situated?

According to Gibbs (2006) the embodiment premise can be stated as follows:

> People's subjective, felt experiences of their bodies in action provide part of the fundamental grounding for language and thought. Cognition is what occurs when the body engages the physical, cultural world and must be studied in terms of the dynamical interactions between people and the environment. Human language and thought emerge from recurring patterns of embodied activity that constrain ongoing intelligent behavior. We must not assume cognition to be purely internal, symbolic, computational, and disembodied, but seek out the gross and detailed ways that language and thought are inextricably shaped by embodied action (Gibbs, 2006, p. 9).

What does this embodiment premise imply for modeling the emergence of language as a collective cognitive activity? This paper is organized as follows. We will first describe a general framework for modeling the emergence of language that encompasses virtually all efforts in this area. We will use this framework to present and discuss three approaches to

modeling the emergence of language, noting both strengths and weaknesses. We will then examine some recent findings concerning spontaneous communication among bonobo chimpanzees (*Pan paniscus*) and attempt to map the bonobo behavior into the previously described modeling framework. The features of the bonobo communication system, as interpreted through the simulation-modeling framework, suggest some ways that constructing more embodied models could improve the modeling efforts. We conclude by revisiting the elements of the shared computational modeling framework and recommend some new directions for modeling the emergence of symbolic language as an embodied collective cognitive activity.

Judging what progress has and has not been made on the problem of the emergence of symbols requires a typology of the referring functions of signals. In the discussion that follows, we will use one of the many typologies developed by Peirce (1958). This typology distinguishes three kinds of sign: icon, index, and symbol. An icon is a form that denotes an object by virtue of resembling the object. Icons and the objects to which they refer share structural properties. An index is a form that denotes an object by virtue of some causal connection to the object. Pointing gestures, for example, are indexes because they may direct attention to the object. Finally, a symbol is a form that denotes an object by virtue of a conventional agreement. This means that the relations between symbols and their referents can be arbitrary. For example, a typical word in spoken language has neither a structural nor a causal relation to its referent.

Icons and indices are tied to local context. The importance of achieving symbolic representation is that symbols make escape from local context possible, a necessary condition for the formulation of general statements. The power of language to refer to that which is absent or abstract derives in part from its symbolic nature.

2. A schema for modeling the emergence of language

In the late 1980s, researchers began to use computational simulations to explore systems in which language-like features emerge from interactions of simple virtual agents. Through the 1990s and the early years of this decade, most researchers used a coherent set of assumptions to shape the simulation of the emergence of language-like systems. This line of research achieved some success, notably in understandings of processes that can lead to convergence among the members of a community. Understanding how symbols could arise in a system that did not assume symbols at the outset was a more elusive target. We will not discuss the technical details of the simulations in this paper. Our goal here is to describe the background assumptions on which the modeling efforts were based.

All of the simulation models of the emergence of language discussed below assume that change in the system occurs incrementally in a series of encounters among pairs of agents in a community. How we imagine those encounters sets the stage for nearly everything that follows. All of the models assume that language emerged in the context of inter-agent communication rather than simply in the workings of internal mental processes. It is because of

this assumption that models of the emergence of language are considered relevant to collective cognition. Communication in its simplest form is modeled as message passing. Typically a community of virtual agents is created. There may be as few as two agents, and as many as hundreds. A program that controls interactions among agents brings pairs of agents together and sets the conditions under which messages are passed between the agents. There are usually two social roles in the interaction: speaker or sender of the message and hearer or receiver of the message. Messages are modeled as forms (words) that carry meanings. The various models show how language-like organization can emerge in the forms, in the meanings, and in the mappings between forms and meanings. Forms may be atomic (e.g., Batali, 1998; Cangelosi & Parisi, 1998; Hurford, 2002; Hutchins & Hazlehurst, 1995; Kirby, 2000; Kirby & Hurford, 2002; Oliphant, 1997; Oliphant & Batali, 1997) or compositional (e.g., Hazlehurst & Hutchins, 1998), corresponding to single words (lexicon) or to sequences of words in which sequential order affects meaning (syntax). In an effective communication system, the forms must be distinguishable from one another and the mapping between forms and their associated meanings must be shared among the members of the community.

One of the key accomplishments of this research tradition is the demonstration that unanticipated structure can emerge from a particular topology of interaction coupling. If each agent models its own behavior on the behavior of others, a community can converge on shared patterns of behavior. This is a specific implementation of a more general principle of modulated positive feedback. Modulated positive feedback is positive feedback with a resonant filter that favors some signals in the loop and causes others to dissipate. This principle underlies many kinds of auto-organizing processes, including those that produce bio-convection, the formation of branching corals and many other animal-built structures as well as physical structures such as waves at the interface of fluid media and even galaxies (Resnick, 1994; Turner, 2000). While the resonant filter in a modulated positive feedback loop can be implemented by a wide variety of mechanisms, all of the simulation models discussed below exploit imitation mechanisms for convergence. The resonant filter of imitation preserves and strengthens instances in which the behaviors of two agents match, and causes instances in which the behaviors of agents mismatch to dissipate.

The idea that humans are, and that their ancestors were, prodigious imitators is an important theme of contemporary studies of primate behavior (Tomasello, 1999). The mutual and reciprocal targeting of behavior creates a positive feedback loop. Once a behavior enters the repertoire of one agent, for whatever reason, it is likely to enter the repertoires of others, which makes it even more likely to enter the repertoires of still others, and so on. In order to produce a shared lexicon, the positive feedback loop created by mutual and reciprocal behavioral targeting must be modulated or filtered in some way. The solutions to the modulation problem vary depending on the assumptions on which the model is built and on the choices made regarding the representation of the various elements of the model. One of the contributions of a consideration of bonobo communication below is to show how the embodiment framework implies complementary behavior in addition to imitation as a mechanism of modulation.

2.1. Three approaches to modeling the emergence of language

We will not attempt here to review the rich and complex literature on modeling the emergence of language. Excellent overviews of this research area are available (Cangelosi & Parisi, 2002; Hurford, Studdert-Kennedy, & Knight, 1998; Hurford, 2002; Kirby, 2002). Instead, we will use the modeling schema to distinguish three major approaches to the emergence of language-like processes. The goal of this discussion is to bring into clearer relief the assumptions that underlie the modeling efforts.

2.1.1. Symbolic forms with symbolic meanings

One major class of models of the emergence of language uses predetermined symbolic forms and symbolic meanings. In these purely computational models the agents are typically implemented as computer code that can select a message for transmission or receive a message that has been transmitted (Batali, 1998; Cangelosi & Parisi, 1998; Hurford, 2002; Kirby, 1999; Kirby & Hurford, 2002; Oliphant, 1997). Most models assume a conduit model of communication in which the sender packages content in a message and transmits the message through a conduit or communication channel to the receiver. The receiver then makes use of the message content in some way. In a comprehensive review article, Kirby (2002) describes the core of the communication model as follows:

> An agent's communication system is described in terms of two probability functions: s ($\mu \in M$, $\sigma \in S$), the transmission function, and r ($\sigma \in S$, $\mu \in M$), the reception function. s (μ, σ) for a given meaning and signal gives the probability that the agent will produce the signal σ for the meaning μ. Conversely, r (σ, μ) gives the probability that the signal σ will be interpreted as the meaning μ by the agent (Kirby, 2002, p. 190).

In most of the simulations, the agent is simply a program that implements these transmission functions. Utterances in these models are generally taken in a very abstract sense to simulate the production of words. Utterance elements are typically modeled as symbolic tokens, although in some models the utterance elements have continuous values. Meanings are typically modeled as symbolic representations that are assumed to be held in the minds of the agents. Often meaning is assumed to be propositional symbolic structure. In many of the models, the researchers create a fixed set of predetermined meanings and a fixed set of predetermined signals (Batali, 1998; Oliphant, 1997). The problem of semantics in that case is to reach agreement on a systematic symbolic mapping of form-meaning pairs. The process that creates the symbolic tokens guarantees the distinguishability of forms.

These simulations demonstrate a variety of robust mechanisms for achieving the alignment of form-meaning mappings. All of these are based on some version of imitation-modulated positive feedback in which form-meaning pairs that gain some popularity in the community replicate more quickly than others, and thus gain even more popularity.

Although researchers in this tradition often refer to the senders of messages as speakers, no attempt is made to model the internal structure of the forms. They are ''atomic'' symbols that are not subject to decomposition. No effort is made to model prosody, intonation

contours, or any other property that distinguishes speech from other forms of language production. The agents in these models are completely disembodied. There is no sense in which the agents perceive the signals or their meanings. Signals and meanings simply appear in the virtual agents as dictated by the interaction protocol. This disembodied conduit metaphor for communication is easy to model and captures some crucial aspects of communication. These simulations demonstrate the emergence of alignment of form-meaning pairs, but since they assume the prior existence of symbolic tokens, they offer no insight into the origins of symbolic forms.

2.1.2. Symbolic forms with embodied meanings

A number of researchers were aware of the limitations of disembodied agents and set out to model the emergence of language-like properties in communities of embodied agents. Using robots as agents is one way to guarantee a role for embodiment in the interactions because the robots interact with each other and their world via sensory and motor mechanisms (Cangelosi & Harnad, 2000; Nolfi, 2002, 2005; Steels, 1996, 2001, 2003; Steels & Kaplan, 1999, 2002; Steels & Belpaeme, 2005; Steels, Kaplan, McIntyre, & van Looveren, 2002).

The robotic explorations of Luc Steels are the best representatives of this class of model. Steels' robots interact with one another in a ''guessing game.'' Two robotic agents encounter a shared world that contains an array of objects. The first agent looks at a scene and segments the scene using its own developing visual routines. It then chooses an object in the visual scene and produces a symbol to represent what it is seeing. If it has already learned a symbol for the chosen object, it produces that symbol. If not, it mints a new unique and discrete symbol. The second agent looks at the same scene and guesses what in that world the symbol produced by the first agent refers to. The second agent can also see where the first agent is looking, and this provides a weak constraint on what the word chosen by the first agent might have meant. If the second agent guesses correctly, then it may imitate the use of that word for that object in subsequent interactions with other agents.

The word forms in these models are discrete symbols, created whenever an agent encounters an as-yet-unlabeled experience. Each new form is atomic and unique, and the process of symbol creation guarantees that symbols are always distinguishable from one another. The meanings consist of the distinctions the agents make in their perceptual experience. These distinctions arise because of the interaction of the robot's perceptual apparatus and the collective process of naming objects. This production of embodied concepts is an important achievement. The guessing game produces alignment of meanings by keeping instances of inter-agent agreement in the system while excluding instances of disagreement. This is modulated positive feedback. The more agreement there is on a particular form-meaning pair, the more likely the use of that pair is to spread to other agents.

Steels' embodied robots also use a form of gesture to help disambiguate the meanings of terms. Robots are able to sense each other's direction of gaze and can use that information to reduce the possible referents of the exchanged symbol. This access

to direction of gaze models something like a pointing gesture. The introduction of gesture is important. It raises questions concerning the possible kinds of gestures agents can make, the roles that those gestures play in processes through which language might emerge, and the evolutionary sequence in which various kinds of gesture might appear.

Steels and his colleagues developed models in which the meanings were created by the agents in interaction with a shared world. From the perspective of embodiment, there are two important innovations here: (a) the separation of dedicated modalities for experiencing the world and communicating about it; and (b) the use of real perceptual processes for engaging the world constrains the possible conceptual structures that can emerge. Earlier models using both symbolic forms and symbolic meanings distinguished the ''transmission'' of forms from the ''experience'' of meanings, but there were no perceptual mechanisms that might constrain the possible conceptual representations. Steels' models are a major step forward because the meanings are embodied in the sense that they are based on bodily processes. The signals, however, remain completely disembodied, discrete symbols. Gaze mediates the association of these symbols with perceptual experiences, but it is not a part of any emergent signaling process. Like the models described in the previous section, these models build in the creation of symbols, and, therefore, cannot explain the emergence of symbols.

2.1.3. Embodied forms with embodied meanings

A few modeling efforts have attempted to construct systems in which both forms and meanings are elements of embodied experience. This is an important step because it raises a number of interesting new issues. What are the relations between the embodied processes that interact with the world of experience and the embodied processes that interact with the world of signals? Must these be different processes, and must agents be constructed so that they automatically or anatomically distinguish the experience of signals from other kinds of experience? Or might signals just be a subset of the things that are experienced in the world?

2.1.3.1. Emergent forms with embodied meanings: Hutchins and Hazlehurst (1995) attempted to move in the direction of full embodiment of forms and meanings. Their interaction protocol is much like that employed in Steels' guessing game. Two agents are chosen from the population and both agents receive identical sensory input from the world. Each agent then generates a signal in a dedicated communication channel or medium. The agents try to learn an efficient encoding of the structure of the domain from their own repeated experience with the situations in the world. This concept-formation process is internal to the agent. Simultaneously, though, they also try to tune their signal production process to increase the probability that they will in future match their signals to those used by other agents when encountering the same situation. The signals associated with environmental structure (objects) and the mapping between the objects and signals both develop gradually through repeated interactions among pairs of agents and their shared world. While the meanings can be seen as embodied products of the interaction of the agent's perceptual

mechanisms and the structure of the world, signals were transmitted in disembodied form from agent to agent as short real-number vectors.

While the signals in these models were not embodied, they were also not given as fixed discrete forms. The forms begin as undifferentiated patterns of small random numbers. They gradually become distinguishable as a consequence of a process that produces an internal efficiently encoded representation of the set of objects in the world. Imitation tuning produces the alignment of form-meaning mappings in these models performed by each agent in every interaction. Every agent always adjusts its own behavior to increase the probability of matching the behaviors of others in the future.

2.1.3.2. Embodied forms and meanings: Two simulations of the emergence of language as a collective cognitive activity address the problem of the emergence of structure in the production of forms in coordination with action. Hazlehurst and Hutchins (1998) attempted to do this in a simulation framework while Marocco and Nolfi (2007) took a robotic approach. As in all of the other models, these models assume a dedicated communication channel that is anatomically separated from the other sensorimotor channels.

Hazlehurst and Hutchins (1998) created a simulation in which pairs of virtual agents interact with each other and with a shared world. As speakers, the agents see the world and act in it while ''speaking'' about their actions by producing ''words'' as in Hutchins and Hazlehurst (1995). However, in the later model, the world is larger and words refer only to an attended part of the world. Like Steels' eye gaze, the agent's actions represent shifts in attention. Listeners see the world and the actions of the other while hearing the speaker's words. Both a structured vocabulary and a set of conventions for creating sequences of words emerge from interactions. The agents used language to coordinate their actions in the shared world. This requirement led to the emergence of shared conventions for the predication of spatial relations among objects.

> For example, the basis for perceiving an object as ''on'' another is not given in the perceptions of objects themselves but, rather, by their positioning relative to each other and relative to a frame of reference. As such, the exact same arrangement of objects in space can be the basis for any of a large number of different relations. Therefore, the particular relation chosen to predicate an arrangement must be imposed by the speaker and, if it is to be understood correctly, also by the listener (Hazlehurst & Hutchins, 1998, p. 380).

Syntactic categories and shared principles of sentence construction also emerged from the interactions:

> The words constructed and employed by agents as constituents of verbal sequences differentiate into two types, representing objects and actions, respectively. The object-type tokens develop and come to represent objects in the world. The action-type tokens come to represent shifts in agent focus of attention as agents build internal structure that reliably maps such tokens onto motor commands for carrying out those actions (Hazlehurst & Hutchins, 1998, p. 414).

In this simulation the structure of the forms, the meanings of the forms (grounded in perception and action), and the form-meaning relations all emerge together. The models developed by Hazlehurst and Hutchins and those developed by Steels and his colleagues include indexical actions. The signals used in both models are symbolic in that they bear arbitrary relations to their referents.

In the model of Marocco and Nolfi (2007), the robot agents have a motor system and a perceptual system to detect conditions in the environment. Each robot also has a communication sender that produces sounds at various frequencies, and a dedicated directional listening system to detect signals produced by other robots. The communication system's input and output are anatomically separate from the sensorimotor system, although there are internal cross connections between the systems. Forms, perceptually grounded meanings, and form-meaning mappings all emerge from interactions. The task is to learn to behave a particular way in a shared world with other robots. Teams of four robots develop ''the ability to find and remain in the two target areas by subdividing themselves equally between the two areas'' (Marocco & Nolfi, 2007, p. 56). The robots move around in space and must arrange themselves so that there are two robots in each of two designated areas. The simulation model does not specify how the robots should use the sounds they make, but the robots nevertheless developed a solution that uses the sounds effectively.

The forms in this system are vectors of floating point values that drive the sound producing system. Beginning from no discernable structure, the robots develop ''a sort of lexicon (including four to five different signals)'' (Marocco & Nolfi, 2007, p. 59). Meanings are modeled as the sensed states of the world. Much like the agents of Hazlehurst and Hutchins, these robots learn perceptually grounded categories that are represented by emergent forms. The robots implicitly learn form-meaning mappings that are manifest in their actions. They learn to shape their behavior to the signals they detect and produce signals that reflect their actions.

2.2. *Summary of advances*

In all of the models considered in this article, the initial condition of the model includes structure in the architecture of the agents and structure in the interaction protocol. As noted above, some of the models begin with structure in a set of candidate meanings. Some of the models also begin with structure in a predetermined set of possible forms. Other models begin without structure in the forms, but include structure in the world with which agents interact. The models produce language-like phenomena through the transformation and propagation of these structures into new, sometimes surprising, patterns inside or between the individual agents. The most compelling models are those in which the emergent structures cannot be anticipated from an inspection of the initial conditions.

The simulation models based on symbolic forms and symbolic meanings demonstrate the power of a variety of imitation mechanisms to produce well-structured form-meaning mappings. The models that deal with symbolic forms and embodied meanings demonstrate the ways that emerging shared language conventions shape, but do not determine, embodied conceptual structures. The key advance in the last set of models is that the processes from

which the organization of forms emerge as well as the processes from which meanings emerge are embodied.

In all of these models, the mechanisms for the input and output of signals are anatomically distinct from the sensorimotor systems. This distinction is probably a legacy of the seeming clean separation of acting and communicating in idealized human symbolic interaction. By accepting this assumption, the modeling community has avoided the possibility of signals inhabiting the same domain of sensorimotor experience as other objects and events in the world. This makes the modeling problem simpler because one does not have to address the problem of how signals could be distinguished from other sorts of action. An unintended side effect of this assumption is that the modeling community has ignored the production and use of meaningful iconic representations. This seems odd, given that many popular narratives about the evolution of language grant iconic representations a central role (Arbib, 2005; Corballis, 2002; Deacon, 1997; Donald, 1991).

Two decades of simulation modeling have provided a good understanding of how imitation-modulated positive feedback processes can produce convergence of practices in a community. It shows how the development of conventions for signal use can shape the emergence of shared perceptually grounded categories. Structure emerges in the set of symbols, to be sure, but the symbolic status of the words was built into the relation they have to events and objects. What makes a pattern a symbol is the role it plays in a larger process. Once a symbolic capacity is constructed, some of the models can show how it can acquire useful structure.

The development of the models revolves around two conceptual pairs. The perception/production pair concerns the ways that the agents interact with their environments. The world/signals pair concerns the way that signals are distinguished from other sorts of phenomena in the experience of the agent. Some of the early models have no world. They consist only of relations between meanings and signals. The early models that include the experience of a world make a clear distinction between world and signals in the architecture of the model. Communication is described by and implemented as two probability functions, a transmission function and a reception function (Kirby, 2002). Signals are no more than the output of one function that is handed to another function whole, unaffected by transmission, and without loss or noise. That it will be interpreted as a signal (and not as some other kind of phenomenon in the world) is guaranteed by the fact that it is delivered directly as input to an interpretive function.

The next generation of models began to explore perception of the world and production of signals. Different models put the emphasis in different places. In the model created by Hutchins and Hazlehurst (1995), for example, experience of the world is implemented by imposing patterns onto a dedicated sensory surface, and signals are implemented by imposing other patterns onto a different anatomically distinct dedicated sensory surface. Both the internal representation of the world and the organization of the external symbols emerged gradually over the course of many interactions between pairs of agents.

In the next generation of models, something fundamental happened with the relation between the agent and the world. Agents now ''attended'' to parts of the world and disregarded other parts. The index of one agent's attention was made available to the other agent

(all dyadic interactions) via eye gaze (Steels, 2001, 2003) or via pointing fingers (Hazlehurst & Hutchins, 1998). Steels did not model perception or production of signals, but his model did learn ''embodied'' categories and conventions for naming those categories with shared symbols. Convergence on the features that define the categories and on the use of symbols arose in parallel. Hazlehurst and Hutchins (1998) had fixed wiring for producing gestures that indicated the speaker's focus of attention and that permitted the listener to interpret the speaker's gesture. Their model allowed learning in the categories of objects and events in the world, while simultaneously learning both a shared lexicon and conventions for the production of sequences of lexical items to denote paths of attention shifts in the shared world. Hazlehurst and Hutchins (1998) attempted to induce emergent embodied learning both in the realm of signals and in the realm of experience of the world. There are still separate and dedicated sensory surfaces onto which particular kinds of experience are imposed.

At this point in the development of the models, the separation of world from (virtual verbal) signals is retained. An embodied indication of focus of attention mediates the problem of determining which aspects of the experienced world the words are meant to be associated with. Imitation provides the modulation of the positive feedback loop. The models still rigidly enforced the anatomical coupling of sensory surfaces to particular classes of phenomena in the world. That is, the world was experienced as a pattern pressed onto a particular sensory surface. The finger position (gesture) of self and other are pressed onto separate dedicated sensory surfaces, and of course, a speaker's words are directed to the listener's ears. This neat alignment of sensory modalities with particular kinds of phenomena builds ''symbolness'' into the architecture of the models. A particular pattern is taken to model the experience of a symbol only because the pattern occurs in the part of the model that implements the movement of structure from one agent to another.

The perception/production distinction has a carefully preconstructed alignment with the world/symbol distinction. World and signals are perceived on distinct dedicated sensory surfaces. Signals are patterns that arise in the speaker that are conveyed directly to the signal-receiving sensory surface of the listener. Since all of the models build the symbolic relations into the architecture of their agents, none of them can account for the emergence of the symbolic capacity. They model the emergence of structure in a symbolic medium, but not the emergence of a symbolic medium. None of these models sheds any light on the emergence of the *possibility* of symbolic relations.

Looking at the trends in the development of the models, it is possible to discern a direction and perhaps if not an endpoint, then at least a future. In this future, both of these careful distinctions must be collapsed. First, real higher animals engage the world with many modalities of action and perception simultaneously (Gibson, 1986; Noe, 2004; Thompson, 2007). Consequently, animal experience is richly multimodal. Second, the experience does not come prelabeled as ''world'' and ''signal.''

Not only are potential signals mixed in with other phenomena in the world, an animal's engagement with the world is multimodal. It is simply not the case that there is a dedicated sensory surface on which signals (and only signals) appear. Animals have multiple sensory and motor systems in which correlated patterns appear. At the heart of the origin of symbols lies the problem of determining which aspects of experience are signals and which are not.

Accepting this fact puts things in proper evolutionary order and calls for the rejection of the assumption that signals could be symbolic or language-like to begin with. Processes of multimodal meaning making must have been conceptually primary and historically prior to the unimodal meaning making that is seen in modern human symbolic language and thought.

Seen in this light, the modeling efforts appear to be backtracking in a historical continuum. The efforts began with idealized architectures that presume a separation of modalities and the alignment of modalities with experience types that are actually seen only in modern socio-technical systems. Each innovation moved the modeling horizon closer to a biological reality. The relations between agents and their worlds became more complex, and eventually even the production of signals incorporated elements of embodiment. Some of the architectural scaffolds of the early models were gradually removed.

However, even in the most complex robotic models, the problem of how to discern the signals in experience was still solved by fiat. Unfortunately, models that solve this problem by building the solution into the architecture of the modeled agents will never help us understand this fundamental development in cognitive history. What sort of model might help here? Well, one in which the starting state contains a temporally continuous flow of multimodal experience. And how could anything ever come to be experienced as a representation (as opposed to being experienced simply as a thing in itself) in such a world? That would be a matter of how the agents produced the forms and what they did with them. This would have several interesting implications. In such a system, what made something a representation would be a matter of (bio-behavioral-) cultural practice rather than a matter of anatomy. Representations would be behavioral phenomena rather than mental states.

3. Bonobo carry activity

One of the most difficult aspects of modeling the emergence of language is to locate the modeled phenomena in some sort of plausible evolutionary sequence. Since language leaves no fossil traces, we have no hard evidence concerning the development of language in the 6 million years since our line diverged from that of the chimpanzees. And while chimpanzees have surely also changed since our lines diverged, the behavior of contemporary chimpanzees suggests what a moment in our own evolutionary history may have been like. Chimpanzees can be enculturated into some kinds of human symbolic activity (Matsuzawa, 2003; Savage-Rumbaugh, 1986; Savage-Rumbaugh, Romski, Hopkins, & Sevcik, 1989), and studying what chimpanzees can and cannot learn to do with symbols sheds light on how their mental capabilities differ from those of humans. Another strategy is to study what chimpanzees actually do in their own activity systems. Careful examination of interactions among captive, but not human-enculturated, bonobos reveals that they spontaneously produce embodied forms that have some, but not all of the features of language. We will argue that this situation provides both a reasonable starting place for the simulation of the emergence of symbolic language and some clues about the kinds of processes that might be involved in this historically elusive transition.

As a proxy for a moment in this history of our species, consider interactions among bonobos (*Pan paniscus*). Bonobos are our nearest genetic kin and, like us, they are highly social animals. They greet, groom, carry and are carried, eat, play, threaten, and relax together. In this paper, we explore the microstructure of the development of an activity in which bonobo mothers and infants coordinate their actions so that mother carries infant. The carry activity is a mode of joint locomotion in which mother and infant move through space with the infant clinging to the body of the mother. Infants gradually learn how to position their bodies in ways that facilitate and seem to encourage being picked up and carried by their mother.

Mothers and experienced infants come together for the carry activity in a very fluid way. The transition from other activities to the carry is an almost ballistic event. Mothers often sweep up infants and move off while looking at their destination. A mother can pick up an infant without looking directly at the infant because the infant simultaneously moves its body and hands in ways that fit and take advantage of the mother's motions. Mother and infant just come at one another, interdigitating (grab, climb on, lift, etc.) mainly by feel. Bonobo mothers experience most carries as tactile and proprioceptive events rather than as visual events. Mothers and infants coming together for a carry is an oft-repeated trajectory, a shared practice with distinctive roles that tends to unfold in regular patterns.

The ways of entering a carry range from direct ''enactments,'' such as the infant climbing on as the mother lifts the infant, to something much more interesting that we will call ''gestures.'' These gestures take the form of frozen fragments of previously enacted trajectories that have been part of the carry activity. For example, a common part of the infant's role in establishing a ventral carry is to lean back and reach out and up. Infants assume this pose and hold it as a solicitation to the mother to pick up the infant and carry it. The frozen gestures are produced in a complex activity field that includes other attention-getting actions such as the infant touching mother's knee to get her to attend or orient in a particular direction.

Such frozen poses are nearly always fragments of the infant's embodied role in the carry. However, they are not excerpts from the accomplished carry, because they do not and cannot enact the mother's role. For example, an infant will do a ''lean back''—similar to its being tipped backward during a ventral carry—but a mother never will. Similarly, a chimp infant will never do a ''present back'' to the mother (which the mother does to ''invite'' the infant to get on her back), although the youngster will eventually do a ''present back'' to a younger infant, to likewise invite it to ride.

The gestures made by the infant can at best suggest the shape of actions by the mother that could coordinate with the displayed fragment. In fact, what the mother experiences when the infant makes such a solicitation is not something that the mother typically experiences in the cooperative carry activity itself. An infant, who is in contact with her mother, and thus in position to enact her part of a carry, may sometimes move away from mother's body and into mother's visual field, turn to face mother, lean back, and extend her arms (see Fig. 1). The infant's gesture is made available to the mother as a visual experience, yet it seems to refer to an activity that consists primarily of tactile, motor, and proprioceptive experience.

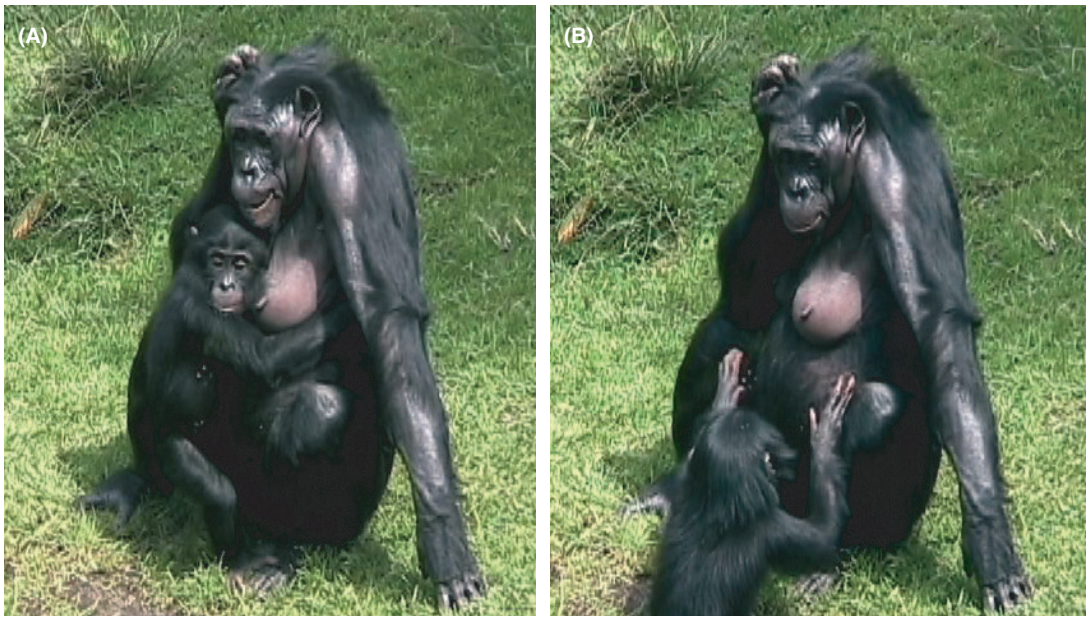

Fig. 1. The infant Kesi clinging to her mother, Lana (A), moves away, then turns and freezes in a pose that invites being picked up (B). (Photos courtesy of Christine Johnson.)

3.1. Interaction protocol

These interactions among bonobo mothers and infants are richly embodied. They are multimodal, involving tactile, proprioceptive, and visual experience. Each animal coordinates the motions of its own body parts and coordinates those with the motions of the body parts of the other. Because both intra-individual and inter-individual coordination are involved, both intra- and inter-individual cognitive processes can affect the organization of the actions.

The primary form of coordination in this activity is the production of complementary rather than matching (imitation) behavior. Complementary coordination provides a different sort of modulation in a positive feedback loop. Actions that ''fit'' the actions of the other are favored. Actions that do not fit the actions of the other are dissipated. This modulation relies on the affordances of adult and infant bonobo bodies and changes as the affordances of the infant body change, and as both mother and infant learn about each other's abilities and preferences.

Thus, the pattern of the bonobo carry as a coherent activity with predictable structure emerges from a complex set of biological constraints. Because the patterns are regular, the animals recognize these oft-repeated trajectories. There is no requirement that they label the activities, but once the activity structure is engaged, the two animals take it smoothly to completion, providing each other the experience of ongoing consummation of expectation.

Because these activities have coherent shape over time, things unfold as expected, and expectations are met and adjusted on a moment-by-moment basis.

Once having emerged, this patterned activity not only has instrumental value, it becomes a resource that is exploited in a new way as a source of raw materials for meaningful gestures. The emergence of structure in interaction (the emergent pattern of complementary movement) is later appropriated by individuals for other purposes (the frozen gestures as indices of the suggested activity). This process of emergence of structure in interaction followed by appropriation of that structure for other purposes provides a source of increasing complexity in social relations that is not driven by genetic changes in the brains of the animals (Strum, Forster, & Hutchins, 1997).

3.2. Transformations of action yield iconic forms

In the language of the general modeling schemas, the pose gestures are forms. Each exists first as a phase of a fluid trajectory performed by the infant in the collaborative accomplishment of a carry. Throughout the period of each act's transformation, the mother–infant dyad continues to engage in seamless iterations of this trajectory, providing the basis for the interpretation of the emerging forms. Over time, acts like lean back and reach differentiate from the normal flow of events and are displaced in space and time. They are performed not in the midst of a carry, but in its absence. With repeated use, they are transformed into phases of new routines that often show simple sequential compositionality. For example, an infant might make contact with the mother to draw her attention, and then enact a lean back in her field of view. Within these new routines, aspects of the original activity stream are preserved, allowing researchers to label them as ''lean back'' and ''reach,'' and presenting the mother bonobo with an embodied piece of the carry trajectory. Thus, for these iconic forms, the carry is both source and referent.

In all of the models discussed above, the symbols are taken to be words that refer to things that can be denoted ostensively by pointing gestures. The bonobo carry phenomena suggest a different scheme, one in which nonpointing iconic gestures refer to patterns of activity by virtue of a similarity between the experience of the form and the experience of the thing to which the form refers. Not only do these forms make reference to such interactions, they are also performed to provoke them. The gestures that emerge are role specific, reflecting the complementarity inherent to these asymmetric interactions. The link between icon and referent promotes the interaction by demonstrating the gesturer's readiness to engage with the complementary moves that typify the shared trajectory.

3.3. Negotiated activity-based meanings

Since the bonobos' iconic forms are built from the embodied repertoire of actions available to primates, and are subject to the constraints of multibody coordination, the same action often plays a role in multiple interactions. A reach by the infant toward the mother, for example, is prelude both to being carried and to being hugged. As a solicitation for a carry, the reach must be configured in a way that helps disambiguate which activity frame is

being suggested. This configuration is structured in both the sequence of actions produced by the infant and in the concurrent behavior of the mother. A reach that actually brings the infant into contact with the mother's ventrum is consistent with a hug if the mother is seated, but with a carry if she is already up on her feet. As a result, a reach performed a short distance away from a seated mother, while still in a place of easy access, is easily differentiated from a request for a hug. The distance and orientation of such a gesture require a mother interested in engagement to physically rise to meet the posing infant, an act more consistent with the carry trajectory than that of the stationary hug. By situating the gesture pose in reference to the mother's position and state, the infant instantiates the activity frame of a carry and in doing so helps to disambiguate the meaning of its gesture.

Another source of disambiguation lies in the conditions under which gesture-mediated interactions recycle. On occasion, the mother responds to the infant's carry gesture by actively embracing her but then remaining stationary. This response establishes a new activity frame. A persistent infant, however, will eschew the next move in the hug trajectory. Instead of grasping its mother's body, it will pull away and reposition itself for yet another gesture presentation. This behavior fails to complement or fit the integrated embodied activity that the mother has initiated. It destabilizes the hug trajectory, further decreasing its likelihood relative to the likelihood of the carry trajectory.

The emergence of carry gestures in the infant does not necessarily depend on the subsequent accomplishment of a carry, as long as that outcome does at least sometimes occur. The long and continuing history of collaborative carries not initiated by such forms helps maintain a statistical regularity to the trajectory that supports the icon's interpretation and effectiveness. This shared history provides many repetitions of the behavioral sequences of the carry activity. The familiarity and complementary expectations that result from the repetitions are needed for the activity to be both the source (a robust pattern to draw on) and referent (distinguishable from other activities) of the iconic gesture. Thus, the form-meaning mapping here does not depend on a successful outcome, but only on establishing the activity frame of a carry. The multistage, multibodied nature of the carry trajectory allows researchers to call even failed attempts ''carryesque'' and provides the animals with a framework within which their negotiated moves have meanings independent of the outcomes of the interaction.

The regularity of the oft-practiced carry trajectory, and the familiarity and affordances of the iconic form, help establish a robust system that is tolerant of innovation and even contradiction. An infant clinging to its mother's ventrum, in the very position it needs to achieve in an accomplished carry, may disassemble this aspect of the configuration and climb off the mother to perform its gesture pose. Further, as time goes on, the infant will come to perform its gesture routine in a fairly standardized fashion whether the mother is laying, sitting, or standing. Because of the participants' shared history with that routine, it takes on a life of its own. As long as some basic criteria are met, such as the gesture being performed in the mother's line of sight, it will function as a solicitation under increasingly variable conditions. As a result, such negotiated iconic forms continue to make reference to the trajectory from which they emerged even as they grow increasingly removed from it.

3.4. Four kinds of remove between forms and meanings

How might the analysis of bonobo carry gestures contribute to our understanding of the emergence of symbolic expression? The bonobos show how meaningful action can be transformed into meaningful iconic gesture that refers to action. This is a small step in the direction of symbolic thought. It provides an evolutionarily plausible starting point for the exploration of processes that can bring about the transformation of meaningful iconic gesture into symbolic forms that bear arbitrary relations to their referents. As noted in the introduction, iconic and indexical forms are bound to local context. Even so, the bonobo carry gestures exhibit four dimensions of removal from the activity they suggest. We will call these dimensions of decontextualization.

1. The iconic pose gesture act is removed in time from the normal course of the carry activity. The gesture happens, in its role of solicitation, both before the activity, and after the many repetitions of the activity that is the source of its structure. It is what Murphy (2004) called ''action in the subjunctive mood.'' It is also an example of prolepsis, a signal in the present that uses the past to refer to the future.
2. The frozen poses are performed out of spatial context. Carries happen in the contact between the infant's body and the mother's body. The gestures are generally performed in the visual field of the mother, but not in contact with her body.
3. The gestures have a different temporal dynamic from the activity itself. They are frozen moments. A stop in the flow of a familiar, normally smooth-changing trajectory is very eye catching—and is detectable to touch as well, when it is performed in contact with another (like an infant in its mother's lap). The altered dynamic serves a cognitive function of increasing the perceptual salience of the display.
4. The shift of modality of presentation from proprioceptive/tactile to visual is another key dimension of decontextualization. This relies on the mother being able to see the pose as a reference to something she experiences in a different sense modality. The possibility of intermodal perception is very important. The typical description of mirror neuron processes, for example, claims that one animal sees another animal acting and experiences some of the motor activation that would be required to perform the observed action. That fits a mechanism of convergence based on imitation, but it is not what seems to be happening when one recovers the feel of another's body from vision. In this case, one animal sees another animal acting and experiences some of the motor activation that would be required to move in ways that complement rather than replicate the observed action.

There are relationships among these dimensions of decontextualization. Removal in space drives the shift in modality of perception from tactile (which requires spatial co-location) to visual (which works best at a small remove in space). Removal in time, which is the basis of the expectation about future activities, makes possible the altered dynamics from continuous motion to frozen pose. The frozen pose dynamic is a way of highlighting the gesture, making it perceptually salient for either tactile or visual perception. In fact, the

production of the freeze action requires a pattern of muscle tension that is not characteristic of other unmoving postures, such as rest. This aspect of the signal is very distinctive, being characteristic of neither the carry nor other activities. Perhaps seeing the frozen pose also activates muscle tension in the viewer that adds to the perceptual salience of the signal.

4. Embodied models of the emergence of symbolic language

How can the analysis of bonobo carry gestures inform efforts to model the emergence of symbolic language? Embodiment should not be a goal of modeling for its own sake, but increasing embodiment in the models brings with it relationships and mechanisms that may be steps on an evolutionary trajectory that leads to symbolic language. Bonobo carry gestures are not fully symbolic forms. But as iconic representations, the pose gestures are a plausible intermediate step between purely instrumental action and symbolic expression. They give us a way to think about the processes that might push things along the act/symbol continuum. Their existence highlights distinctions that have been overlooked or obscured in the current generation of modeling efforts.

4.1. Complementarity and imitation in the interaction protocol

Imitation has been shown to be a powerful engine of inter-agent convergence on emergent behavior patterns. It is an excellent process for filtering signals in a modulated positive feedback loop. But the creative powers of the imitation process are limited by the very nature of imitation mechanisms. To the extent that an imitation mechanism succeeds in creating similar behavior, it fails to create new behavior. When imitation is the engine of convergence, creative change happens only through failure or noise. However, imitation is not the only form of behavioral coordination among animals. Complementary behaviors are also important and may be a more powerful engine of change than imitation.

The complementarity of intra-individual behavior patterns is provided by the constraints (affordances and mechanisms) of animal anatomy and physiology. The dynamics of the body–brain system constrain the patterns. When multimodal behavior is produced, the entrainment of intra-individual modalities is a natural outcome. But entrained behaviors in distinct behavioral media are never identical. Coordinating vocalizations with limb movements, for example, necessarily results in complementary patterns rather than in matching behaviors because the vocal system and the limb movement systems have different dynamics. Embodiment thus provides a natural source of decontextualization of signals *within* each body.

The complementarity of inter-individual behavior patterns emerges from the dynamics of brain–body systems in social interaction and in some animals is further constrained by the jointly understood activity structure in which the behaviors are situated. This complementarity of action in joint activity is what produces the appearance of distinct ‘‘roles’’ in

activities. Coordination of embodied social behavior thus provides a natural source of decontextualization of signals *between* bodies.

4.2. Two more dimensions of remove between form and meaning

The discussion of bonobo carries identified four dimensions of removal between activity and form. In combination, these produced iconic gestures. Symbolic representations require at least two more dimensions of decontextualization.

4.2.1. Noniconic forms

First, the internal structure of a symbol should bear no simple mapping to the internal structure of the concept to which it refers. The pose gesture is iconic because it has an internal structure that maps directly onto the structure of some part of the thing it refers to, the carry activity. One way to move beyond iconic representations is to insist that the medium in which the form is produced and experienced be different from the medium in which the meaning is produced and experienced. All of the simulation models enforce the separation between forms and meanings by constructing a dedicated medium or channel for communication. They build the separation between experience and signals into the architecture of the simulated agents. Even using embodied channels, this creates the possibility of arbitrary form-meaning mappings because the distinct media develop internal structure in response to different processes. In embodied multimodal agents it might be possible for this separation to emerge rather than having it built into the agents' architecture. With multiple, simultaneously active, partially entrained modes of experience and action, agents might learn that in certain contexts, some but not all particular actions in particular modes are reliable signals. Just as the distinguishability of a bonobo iconic pose gesture depends on the field of relations among the elements that constitute the behavioral repertoire of the bonobos, the distinguishability of any form depends on the field of relations among the structures of the entire inventory of forms. Different modalities imply different fields of relations. Once a form takes its place in the ecosystem of forms, it acquires relations-in-use to the other forms. In this cognitive ecosystem it is subject to a complex set of constraints and forces that can alter its shape. For example, if the modality of representation requires compression of information, a form's structure may be driven away from the structures of similar forms (Becker & Hinton, 1992; Hutchins & Hazlehurst, 1995). The significance of embodiment in this discussion is that the functional differences among embodied media guarantee that the structures that emerge in different media will be subject to different forces. Embodiment provides variability on which emergent processes can act.

Complementary representations in different modalities may have different internal structure and may also refer to different aspects of an activity. For example, among agents modeled on the bonobos, a lean-back gesture could refer to the carry activity while a complementary vocalization could refer to the destination of the carry. A compositional form could arise from an ellipsis of a compound action. Among bonobos, compound actions have already been observed in which the first action is an orientation in a particular direction, Ao, and the second action is an iconic solicitation for carry, Ac. The destination of the

carry is implied by the conjunction of the orienting action with the iconic gesture, Ao + Ac. Now imagine agents who produce congruent vocalizations with each of these actions. In addition to the orienting action Ao and the gesture Ac there might be a vocalization with the orienting action Vo, and another vocalization with the lean-back carry solicitation gesture Vc. Once the vocalizations become associated with the actions they accompany, there will be redundancy in the complete compound vocal and action signal, ((Ao + Vo) + (Ac + Vc)). Because the signal is redundant, some elements can be dropped without loss of information. Imagine that the now redundant orienting action, Ao, and the redundant carry vocalization Vc are dropped from the performance. What remains would be Vo, the vocalization for the destination, and Ac, the iconic gesture for the carry activity. This is a compositional structure (Vo + Ac) in which the gesture and the vocalization are complementary. The implication of this thought experiment for modeling efforts is that embodied multimodal compositionality can lead to new possibilities for the kinds of decontextualization that are needed to get from meaningful actions to symbols.

4.2.2. Impersonal forms

A second additional dimension of decontextualization is required by symbols. The word ''carry,'' as it is used in human adult conversation, is not about the experience of the speaker or the listener; it is about the activity in the abstract. There is a schema for the activity with two roles. For the developing child, the word may actually be associated with the first-person experience of being carried. Early on, what ''carry'' means to mother is probably not at all what ''carry'' means to the infant. But the different meanings have complementary behavioral consequences. The word can be used by either party to facilitate the initiation of the carry activity.

It is clear that many cultural elements are transmitted via the emergence of complementary action in joint activity rather than by imitation. The person-neutrality of shared symbols that have the same meaning for speaker and listener is made possible by the separation of the form from the first-person experience in which it was originally grounded. The shift from first-person experience to shared impersonal forms also implies a change in the mechanism of coordination from the tuning of complementary action to the tuning of similar action through imitation.

4.3. Implications of embodiment for modeling efforts

The available data on bonobo carry solicitations do not say much about collective behavior. We do not know how such behaviors might spread through a community over time. However, the alignment of behaviors is the aspect that has been most thoroughly explored in the simulation models. The big change here would be accounting for the alignment of behaviors via the generation of complementary action rather than imitation. This is an additional challenge for future modeling. The analysis of the bonobo carries provides insights into an aspect of the emergence of symbolic language that our computational models have not yet been able to address. Taking decontextualization as a path from action to symbol does not eliminate the embodied, embedded, situated aspects of cognition. Rather, it

describes the use of these processes to create separation between forms and their meanings: separation in time, space, dynamics, modality, structure, and personal involvement. Over the course of phylogenetic history, processes of decontextualization create new environments for thinking in which different skills are embodied and actors are embedded and situated in different cultural activities.

The principal challenges in modeling the emergence of symbolic language as a collective cognitive process are conceptual rather than technological. The technology needed to create fully embodied robot agents already exists. The analysis of bonobo carry gestures highlights a number of conceptual relations that are ripe for exploitation in computational and robotic models:

1. Agents should be more fully embodied and possess multiple sensory and motor modalities for interacting with their environments.
2. Embodied multimodality provides the possibilities and challenges of intra-agent coordination and for much more complex forms of inter-agent coordination.
3. Multimodality also creates the possibility of cross-modal compositionality of signals. This may be essential to achieving the arbitrary relation between forms and meanings that is a sine qua non of true symbolic representation.
4. Signals should appear as patterns in the same world of experience as the things to which they refer. This opens the door to as-yet unexploited iconic representations, which must have played an important role in the subsequent development of non-iconic forms of representation. Furthermore, distinguishing communicative action from other kinds of instrumental action is an important theoretical challenge in its own right. Meeting this challenge is a key part of understanding the origins of symbolic representation.
5. Putting signals into the world of action also creates the opportunity for the reuse of emergent structures as communicative forms. The appropriation of emergent structure is a valuable source of increased complexity in evolving systems.
6. Imitation is essential for community convergence on shared behaviors. Complementarity is another important, but often overlooked, mechanism of social learning. Agents should have mechanisms to observe and participate in the activities of the agents around them, producing both imitative and complementary actions.
7. Coordination through complementary action implies the recognition of recurrent predictable activity patterns. Building agents that can discover reliable emergent patterns of activity is possible using modern machine learning techniques. The recognition of activity is a necessary condition for disambiguating the meanings of actions in context, which provides grounding for the parsing of sequentially compositional forms. The recognition of familiar and well-structured activities also provides a robustness that allows the system to bear the transformations brought by decontextualization.

Embodiment is often thought of as a suite of phenomena that apply only to the interactions of an agent with its local environment. We have tried to show here that the current

trends in cognitive science toward embodied and situated theoretical frameworks also have implications for collective cognitive processes. In particular, incorporating embodiment into models of the emergence of symbolic language as a collective cognitive activity suggests a number of new ways to approach long-standing problems.

Acknowledgments

We are indebted to two anonymous reviewers and to Robert Goldstone and Georg Theiner for comments on an early draft of this paper authored by Edwin Hutchins only. Brian Hazlehurst, the architect of some of the best research in this area, provided key critical insights on this paper. Funding for this research was provided by NSF award no. 0729013, ''A multiscale framework for analyzing activity dynamics.''

References

Arbib, M. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, *28*, 105–124.

Batali, J. (1998). Computational simulations of the emergence of grammar. In J. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the evolution of language: Social and cognitive bases* (pp. 405–426). Cambridge, England: Cambridge University Press.

Becker, S., & Hinton, G. (1992). A self-organizing neural network that discovers surfaces in random-dot stereograms. *Nature*, *355*(6356), 161–163.

Cangelosi, A., & Harnad, S. (2000). The adaptive advantage of symbolic theft over sensorimotor toil: Grounding language in perceptual categories. *Evolution of Communication*, *4*(1), 117–142.

Cangelosi, A., & Parisi, D. (1998). The emergence of a ''language'' in an evolving population of neural networks. *Connection Science*, *10*(2), 83–97.

Cangelosi, A., & Parisi, D. (Eds.) (2002). *Simulating the evolution of language*. London: Springer-Verlag.

Corballis, M. C. (2002). *From hand to mouth: The origins of language*. Princeton, NJ: Princeton University Press.

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

Donald, M. (1991). *Origins of the modern mind: Three steps in the evolution of culture and cognition*. Cambridge, MA: Harvard University Press.

Gibbs, R. (2006). *Embodiment and cognitive science*. New York: Cambridge University Press.

Gibson, J. J. (1986). *The ecological approach to visual perception*. Hillsdale, NJ: Erlbaum.

Hazlehurst, B., & Hutchins, E. (1998). The emergence of propositions from the coordination of talk and action in a shared world. (K. Plunkett, Ed.) *Language and Cognitive Process*. (Special issue on Connectionist Approaches to Language Development), *13*(2/3), 373–424.

Hurford, J. (2002). Expression/induction models of language evolution: Dimensions and issues. In T. Briscoe (Ed.), *Linguistic evolution through language acquisition* (pp. 301–344). Cambridge, England: Cambridge University Press.

Hurford, J., Studdert-Kennedy, M., & Knight, C. (Eds.) (1998). *Approaches to the evolution of language: Social and cognitive bases*. New York: Cambridge University Press.

Hutchins, E., & Hazlehurst, B. (1995). How to invent a lexicon: The development of shared symbols in interaction. In N. Gilbert & R. Conte (Eds.), *Artificial societies: The computer simulation of social life* (pp. 157–189). London: UCL Press.

Hutchins, E., & Hazlehurst, B. (2002). Auto-organization and emergence of shared language structure. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 279–305). London: Springer-Verlag.

Kirby, S. (1999). Syntax out of learning: The cultural evolution of structured communication in a population of induction algorithms. In D. Floreano, J. D. Nicoud, & F. Mondada (Eds.), *Advances in artificial life, proceedings of the 5th European conference, ECAL'99* (pp. 694–703). Berlin: Springer.

Kirby, S. (2000). Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In C. Knight, M. Studdert-Kennedy, & J. R. Hurford (Eds.), *The evolutionary emergence of language: Social function and the origins of linguistic form* (pp. 303–323). Cambridge, England: Cambridge University Press.

Kirby, S. (2002). Natural language from artificial life. *Artificial Life*, *8*, 185–215.

Kirby, S., & Hurford, J. R. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 121–148). London: Springer.

Marocco, D., & Nolfi, S. (2007). Emergence of communication in embodied agents evolved for the ability to solve a collective navigation problem. *Connection Science*, *19*(1), 53–74.

Matsuzawa, T. (2003). The AI project: Historical and ecological contexts. *Animal Cognition*, *6*, 199–211.

Murphy, K. (2004). Imagination as joint activity: The case of architectural interaction. *Mind, Culture & Activity*, *11*, 270–281.

Noe, A. (2004). *Action in perception*. Cambridge, MA: MIT Press.

Nolfi, S. (2002). Evolving robots able to self-localize in the environment: The importance of viewing cognition as the result of processes occurring at different time scales. *Connection Science*, *14*, 231–244.

Nolfi, S. (2005). Emergence of communication in embodied agents: Co-adapting communicative and non-communicative behaviours. *Connection Science*, *17*(3/4), 231–248.

Oliphant, M. (1997). *Formal approaches to innate and learned communication: Laying the foundation of language*. Unpublished doctoral dissertation, University of California San Diego, Department of Cognitive Science.

Oliphant, M., & Batali, J. (1997). Learning and the emergence of coordinated communication. *Center for Research on Language Newsletter*, *11*(1), 1–46.

Peirce, C. S. (1958). On a new list of categories. In C. Hartshorne & P. Weiss (Eds.), *Collected papers of Charles Sanders Peirce* (Vol. 1, pp. 545–559). Cambridge, MA: Harvard University Press.

Resnick, M. (1994). Learning about life. *Artificial Life*, *1*(1/2), 229–241.

Savage-Rumbaugh, E. S. (1986). *Ape language: From conditioned response to symbol*. New York: Columbia University Press.

Savage-Rumbaugh, E. S., Romski, M. A., Hopkins, W. D., & Sevcik, R. A. (1989). Symbol acquisition and use by *Pan troglodytes*, *Pan paniscus* and *Homo sapiens*. In L. A. Marquardt & P. G. Heltne (Eds.), *Understanding chimpanzees* (pp. 266–295). Cambridge, MA: Harvard University Press.

Steels, L. (1996). Self-organizing vocabularies. In C. Langton & K. Shimohara (Ed.), *Proceedings of alife V* (pp. 177–184). Cambridge, MA: MIT Press.

Steels, L. (2001). Language games for autonomous robots. *IEEE Intelligent Systems*, *16*, 17–22.

Steels, L. (2003). Evolving grounded communication for robots. *Trends in Cognitive Science*, *7*(7), 308–312.

Steels, L., & Belpaeme, T. (2005). Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and Brain Sciences*, *28*, 469–529.

Steels, L., & Kaplan, F. (1999). Situated grounded word semantics. In T. Dean (Ed.), *Proceedings of the sixteenth international joint conference on artificial intelligence (IJCAI'99)* (pp. 862–867). San Francisco: Morgan Kauffman.

Steels, L., & Kaplan, F. (2002). Bootstrapping grounded word semantics. In T. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models* (pp. 53–73). Cambridge, England: Cambridge University Press.

Steels, L., Kaplan, F., McIntyre, A., & van Looveren, J. (2002). Crucial factors in the origins of word-meaning. In A. Wray (Ed.), *The transition to language* (pp. 252–271). Oxford, England: Oxford University Press.

Strum, S. C., Forster, D., & Hutchins, E. (1997). Why Machiavellian intelligence may not be Machiavellian. In A. Whiten & R. W. Byrne (Eds.), *Machiavellian intelligence II: Extensions and evaluations* (pp. 50–85). Cambridge, England: Cambridge University Press.

Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Cambridge, MA: Harvard University Press.

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

Turner, J. S. (2000). *The extended organism: The physiology of animal-built structures*. Cambridge, MA: Harvard University Press.