

HOW TO INVENT A LEXICON: THE DEVELOPMENT OF SHARED SYMBOLS IN INTERACTION

Edwin Hutchins

Brian Hazlehurst

Department of Cognitive Science
University of California, San Diego
La Jolla, California
USA

This paper appeared in N. Gilbert and R. Conte (Eds.), *Artificial Societies: The computer simulation of social life*. London: UCL Press.

INTRODUCTION

Human language provides, among other things, a mechanism for distinguishing between relevant objects in the natural environment. This mechanism is composed of two components – forms and meanings – which must be shared by the community of language users. The lexicon constitutes much of a language's form, but its description as a set of shared form-meaning resources for communication is seldom addressed by formal models of language. This paper presents a simulation model of how shared symbols (form-meaning pairs) can emerge from the interactions of simple cognitive agents in an artificial world. The problem of creating a lexicon from scratch is solved by having cognitive agents capable of organizing themselves internally – that is, agents which can learn to classify a world of visual phenomena – share their expressions of visual experience in interaction. The model is seen as an instantiation of a theory of cognition which takes symbols to be a product of inter- and intra-individual organizations of behavior, the result of cultural process.

Below, we first present a theoretical stance which has been derived (elsewhere) from empirical investigation of human cognitive phenomena. This stance leads to the articulation of some simple information-processing constraints which must hold for lexicons and their users. In the middle of the paper, we present two simulations employing communities of simple agents which allow us to model how a lexicon could come into existence or emerge from the interactions between agents in an extremely simple world of experience. The model allows us to explore issues which involve interaction between group- and individual-level properties of cognition, social organization, and communication. Near the end of the paper, we briefly review some findings from the literature on lexicons of natural language. We conclude the paper with an evaluation of the model as an instantiation of the theoretical stance.

A COMMUNITY OF MINDS AS COGNITIVE SYSTEM

Recently, we have been exploring a novel approach to cognitive anthropology. We've been trying to push the boundaries of a genuinely cognitive unit of analysis out beyond the skin of the individual. Ever since symbolic and cognitive anthropology embarked on their ideational odyssey in the 1950s they have proceeded away from the material and the social aspects of human life. Of course, many people are interested in *social cognition* in which the social world is the content of cognition. And in fact, there are good arguments for believing that human intelligence developed in the context of reasoning about social situations (Byrne & Whiten 1988, Levinson, in press). This kind of relationship between the social and the cognitive is important,

but it is still centered on the notion of the individual as the primary unit of cognitive analysis. The social world is taken to be a set of circumstances "outside" the individual, about which the individual reasons. What we intend, instead, is to put the social and the cognitive on equal theoretical footing by taking a *community of minds* as our unit of analysis.

This new perspective permits two things to happen that are not possible from the traditional perspective. First, it permits inquiry about the role of social organization in the cognitive architecture of the system and allows description of the cognitive consequences of social organization at the level of the community (Hazlehurst 1991, Hutchins, 1991). Second, it permits symbolic phenomena that are outside the individuals to be treated as real components of the cognitive unit of analysis (Hazlehurst 1994, Hutchins in press, Sperber 1985).

Making this move also presents an opportunity to view language in a new way. Cognitive science generally takes the existence of language as a given and concerns itself with the sorts of cognitive processes that must be involved when an individual processes language (in production and comprehension). From the perspective of the community of minds as cognitive system, a language's information bearing capacity, conventions of use, functionality, distribution, variability, etc. – all become determinants of the cognitive properties of the community because these things are instrumental in determining where, when, and what kinds of information move through the system (cf. Freyd 1983). This attention to the movement of information in the larger system necessarily brings the material world back into play since, having acknowledged symbols outside the head, we now must take seriously their material nature. Furthermore, we need not – indeed must not – ignore the means individuals have or develop for incorporating symbols into private as well as collective organizations of behavior. By redefining our unit of cognitive analysis, it seems to us that progress may be made towards reuniting the social and the material with the cognitive (cf. Goody 1977).

A MODEL OF THE EMERGENCE OF SHARED LANGUAGE

The existence of shared language is one of the central facts of human existence. Language appears to be closely tied to most high level cognitive activities. It mediates most of the interactions among members of the most social of all species. Once a language exists, it is not difficult to think of the means by which it could be maintained and propagated from generation to generation in a population. But without anyone to tell individuals which language to speak, how could a language ever arise? How could something structured come from that which is unstructured? It's a puzzle.

There is, of course, a vast speculative literature on the origins of language which we will not attempt to treat here. Rather, this study focuses more modestly on the development of sets of local lexical distinctions as may arise in small groups of cognitive agents engaged in shared tasks. In this paper we outline a scheme by which shared denotational resources, that will be called *symbols*, arise in the interactions among the members of such a community. This is certainly not a model of the development of a human language, but it does demonstrate how simple shared structures can arise where none existed before. These structures are products of a system whose organizing dynamics are an interplay between intra-individual and inter-individual coordinations.

In the presentation, these structures will be referred to as terms, descriptions, words, or patterns of acoustic features. There is, however, no strong a priori commitment to any particular level of linguistic representation here, and the structures described might just as well be thought of as patterns of denotational or even relational features. Each of the above terms takes its meaning from a claim about the function of these public representations in this artificial world. We take this stance to be an important methodological and theoretical component of this work. Representations don't get to be symbols by virtue of *us* creating them or calling them such, but rather, by our reading of what they do for the members of a community of artificial cognitive agents (Clancey 1989).

The model is based on six central theoretical assumptions, derived from a theory of *distributed cognition* (Hazlehurst 1991, 1994, Hutchins 1990, 1991, 1993, in press, Hutchins & Hazlehurst 1991, Hutchins & Klausen, in press).

1. No mind can influence another except via mediating structure. (The *no telepathy* assumption.)
2. No social mind can become appropriately organized except via interaction with the products of the organization of other minds, and the shared physical environment. (The *cultural grounding of intelligence* assumption.)
3. The nature of mental representations cannot simply be assumed, they must be explained. (The *shallow symbols* assumption - in contrast with the *deep symbols* assumption which brings symbols into the language of thought as an article of faith rather than as a consequence of cultural process.)
4. Symbols always have both a material and an ideal component. The material component is what makes form, structure, and differences possible. The ideal component is a function of the stance that organized individuals take toward these material forms. (The *material symbols* assumption.)

5. Cognition can be described as the propagation of representational state across representational media that may be internal to or external to individual minds. (The *distributed information-processing* assumption.)
6. The processes that account for the normal operation of the cognitive system should also account for its development through time. (The *no developmental magic* assumption.)

Below we present a computer simulation that is an implementation of these assumptions. It turns out to be a very robust procedure by which a community of individuals can develop a shared set of symbols. The simulation is *not* to be taken seriously as a model of any part of human history. It is the simplest possible scheme that captures the essential properties of the system being modeled. The simulations are not to be taken as representations of actual human cognition, language, culture, or experience, but as existence proofs that a particular kind of process is capable of producing a particular sort of outcome, in this case, a community with a shared lexicon. One is certainly free to question the extent to which it is reasonable to map this process onto plausible courses of human behavior and in the discussion section we consider ways in which the model is consistent with observations about human language.

THE CONSTRAINTS ON A SHARED LEXICON

The central problems of inventing a lexicon can be stated in terms of a description of the outcome. Consider two individuals, A and B, and a set of visual scenes or contexts in the world numbered $1, 2, 3, \dots, m$. Let the description that an individual uses for referring to a scene be denoted by the concatenation of the letter designating the individual and the number of the scene. For example, "B5" denotes the description that individual B uses for referring to the fifth scene.¹ Now, if the lexicon is to be *shared*, the word that A uses for any particular scene must be the same as that used by B. In notation: $A_1 = B_1, A_2 = B_2, \dots, A_m = B_m$. Simultaneously, if the lexicon is to be a lexicon at all, there must be differences between the material forms of the words used by each individual for different scenes. In notation: $A_1 \neq A_2 \neq \dots \neq A_m$ and $B_1 \neq B_2 \neq \dots \neq B_m$. (It wouldn't do to have a lexicon for m scenes

¹ In fact, a description is a vector of real values which we imagine to be articulatory features capable of generating a monolexemic word, a piece of agent-created structure, in this artificial world. The act of "referring" is a property of our construction of the world—we have built into the world the need for agents to internalize visual experiences and code that organization in externally realizable structures. This, of course, bypasses a long history of evolution which might select some of these properties as internal properties of agents. However, in granting these (on face value, plausible) assumptions we are better able to address cognitive and cultural issues with our simulations.

that were m homonyms.²) These two constraints must somehow be simultaneously satisfied in any process that is to develop a shared lexicon. For our purposes, *a shared lexicon is a consensus on a set of distinctions.*³

It is interesting that the same mathematical basis, although a different computational procedure, was developed by Hinton and Becker (1989) to show how modules in the brain could discover a shared communication protocol without a supervisor to specify how to communicate. The problem of discovering a lexicon may be quite general and seems to occur at a number of levels of organization in cognitive systems. Independent of its relation to problems solved by organic cognitive systems, the procedures described here might provide a general engineering solution to problems where modules must invent a means of communication but where the organization of the communications protocol cannot be specified in advance.

Cultural Process

Before turning to the simulation, we need to say a few more words about the theoretical stance. The six theoretical assumptions described previously can be assembled into a model of cultural process, the temporal characterization of distributed cognition shown in Figure 1. Our inventory of representational structure includes *natural structure* in the environment, *internal structure* in the individuals, and *artifactual structure* in the environment. Artifactual structure is a bridge between internal structures. Artifacts may provide the link between internal structures in one individual and those in another individual (as is the case in communication), or between one set of internal structures in an individual and another set of internal structures in that same individual (as is the case in using written records as a memory, for example). Internal structures provide bridges both between successive artifactual structures and between natural and artifactual structures. Following Sperber (1985, p. 76) we may say that "A representation [artifactual or internal structure] is of something [natural, artifactual, or internal structure] for some information processing device [internal or artifactual structure]."

² This model does not deal with homonyms or synonyms. However, note that it is a matter of current debate whether such categories actually exist in human language when the perspective of language use and learning is taken to examine instances claimed to be synonyms or homonyms (Clark, 1987; Slobin, 1985).

³See Appendix for a more formal treatment of this analysis.

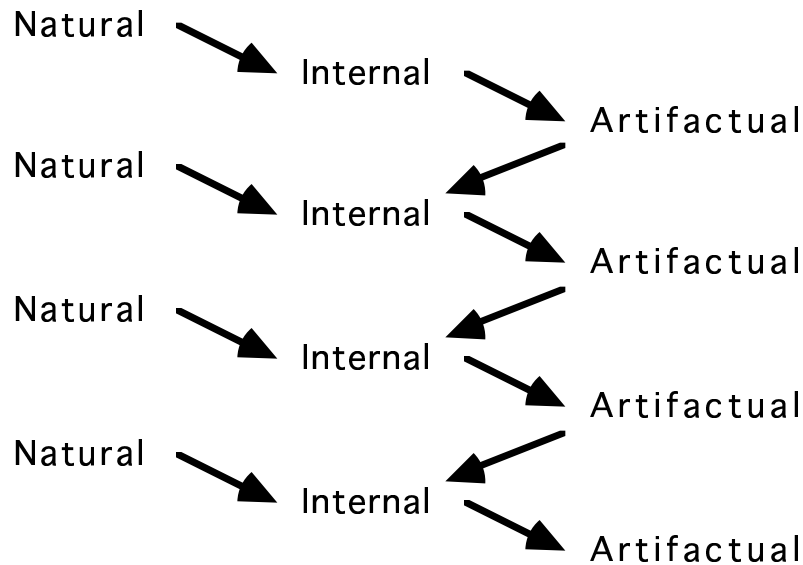


Figure 1. The relations of natural, internal, and artifactual structure which instantiate a cultural process. (The arrows represent the propagation of constraints. Constraints may be propagated by many means. We use the cover term coordination to refer to the satisfaction of constraints regardless of the mechanism by which constraint satisfaction is achieved.)

Connectionist Networks As Simple Agents

In the past eight years, developments in computational modeling employing *connectionist networks* have made it possible to think in new ways about the relations between structure inside a system and structure outside.⁴ A *network* is composed of two kinds of mathematical objects: *units* and *connections*. Units take on activation values, generally in the range of 0.0 to 1.0, and pass activation along one-way connections to other units. This passing of activation along a connection is modulated by a real value associated with that connection, the *connection strength* or *weight*. The passing of activation from input units to output units of the network can be treated as the implementation of a function, and viewed as the network's behavior in a given environment. Through modification of the network's weights, in time, the network adapts to the shape of that environment.

Connectionist networks of a class called *autoassociators* have particularly nice properties with respect to the problem of discovering and encoding structural regularities in their environment. Autoassociator networks learn

⁴The best background work on connectionism is the two volume set *Parallel Distributed Processing* by Rumelhart et al. (1986), and McClelland et al. (1986). The behavior of autoassociator networks is thoroughly analyzed in Chauvin (1988).

to duplicate on the output layer the identical pattern of activation presented to the input layer. Figure 2 shows a simple autoassociator network. It consists of three *layers* of units. Input units on the left, output units on the right, and hidden units in the middle. *Targets* are real valued vectors which are structurally similar to the output and input layers but, like inputs, are thought of as information external to the network. Targets are part of the environment which the network is made to learn (see below) – they provide the teaching signal. In the case of autoassociators, the targets are simply the input patterns themselves, allowing the network to learn about the environment with no additional teaching signal.

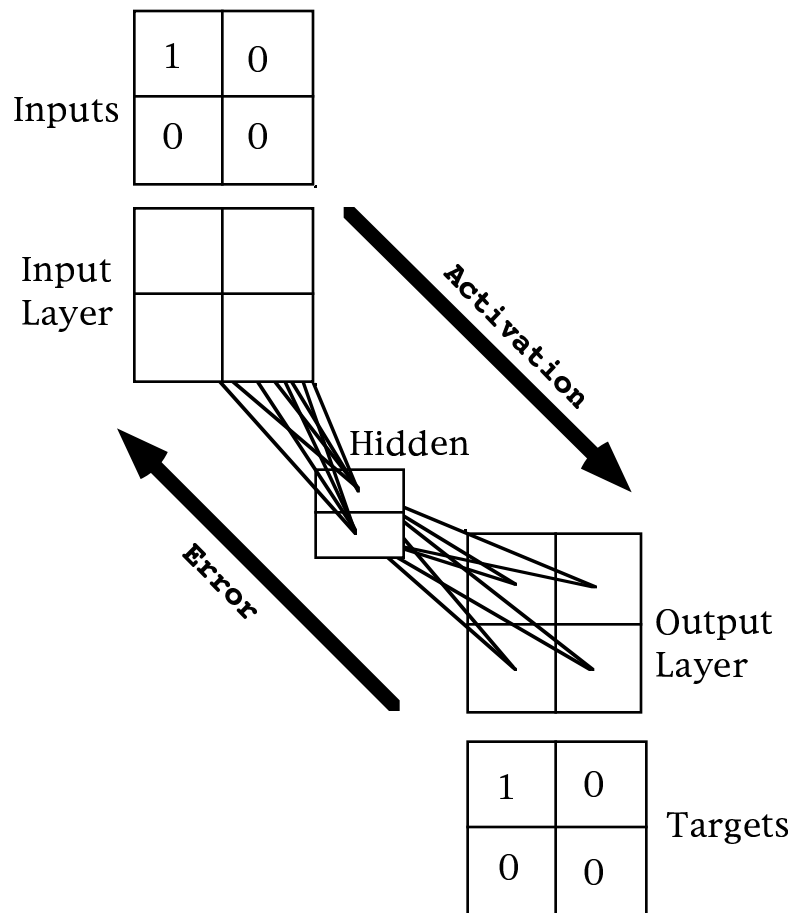


Figure 2. A typical autoassociation network and learning scheme. (See text for explanation of network components.)

Limitations on space make a full description of this kind of information processing system impossible. The following sentences will hopefully convey the style of computation entailed, if not the details.

Every unit in the input layer of a network has a unique connection to every unit in the hidden layer, and every unit in the hidden layer has a unique connection to every unit in the output layer (see Figure 2). The

strengths of these connections can be adjusted. The activations of the input layer are set by external phenomena. The activations of the other units are determined by the activations of the units from which they have connections and on the strengths of those connections. The task for the network is, starting from random connection strengths, to discover a pattern of connection strengths that will produce the desired output (i.e. the target) in response to a given set of inputs. Incremental improvement in accomplishing this task is referred to as *learning*. The networks modeled here use a procedure called the *back-propagation of error* to find an appropriate set of connection strengths. In this scheme, the output produced is compared to the target,⁵ and the difference between output and target is an error in the network's ability to perform this input-output mapping. The connections are then adjusted to reduce this error on future trials at this task. The problem for the network can be viewed as one of finding a set of weights which simultaneously meets the constraints imposed by all of the input-output mappings it is made to perform.

Meeting Constraint #1 Of The Lexicon: Words must discriminate between objects in the environment

Rumelhart, Hinton, and Williams (1986) have shown that under certain conditions, the activations of the hidden layer units of fully trained autoassociator networks converge on efficient encodings of the structural regularities of the input data set. That is, the connections between input and hidden units must produce activations at the hidden layer which can be used by the connections between hidden and output units to produce the target, under the constraints of the function which propagates activation. For example, given any four orthogonal input patterns and an autoassociator network with two hidden units, the hidden unit activations for the four input cases should converge on $\{(0, 0) (0, 1) (1, 0) (1, 1)\}$. This is because the network must use the activations of the hidden units to encode the four cases and the encoding scheme attempts to distinguish (optimally) among the cases.

Producing these efficient encodings is equivalent to feature extraction. That is, what the networks learn is how to classify the input data in terms of distinctive features or principal components. If we fold an autoassociator in half, and allow the hidden layer to produce representations which become a material part of the shared environment of interaction, then the (now "public") hidden layer encodings produce one of the properties we want in a

⁵For an autoassociator, the target is identical to the input, thus reducing the problem to an identity mapping on the input set.

lexicon.⁶ In Figure 3 we have relabeled these units *Verbal Input/Output* units because this is the location where agents produce words and receive words for comparison with the words of others.

If we take the remaining parts of the network to be a simple visual system – capable of classifying scenes in the environment – then the Verbal Input/Output layer is capable of generating patterns of activation in response to each visual scene encountered, and these patterns are (or become) maximally different from each other (within the boundaries set by learning and resource limitations of the network). Regarding network A's descriptions for the m scenes, this satisfies the constraints: $A1 \neq A2 \neq \dots A_m$.

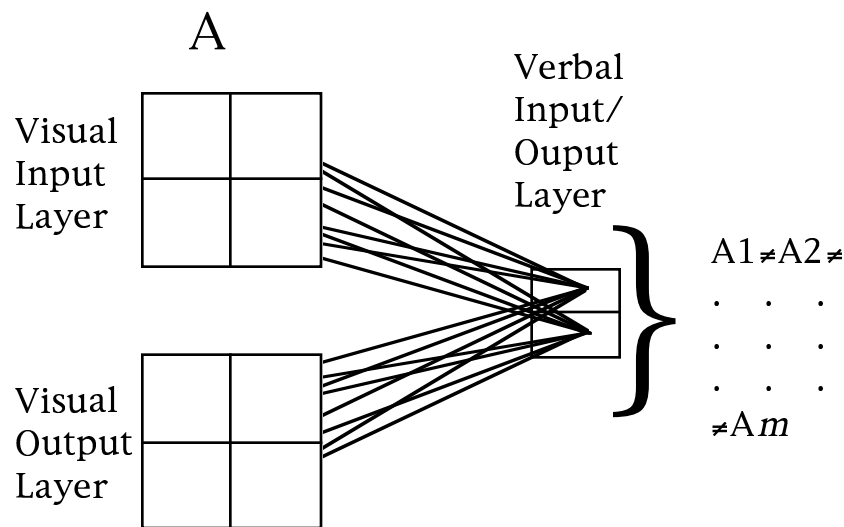


Figure 3. A modified autoassociator, with public hidden units. (By "folding" an autoassociator back on itself, we create a system capable of generating referentially meaningful, distinct representations of the m scenes.)

Meeting Constraint #2 Of The Lexicon: Word meanings must be shared

Virtually all work in connectionist modeling today is concerned with using connectionist networks to model aspects of the cognition of *individuals*. Our theoretical stance suggests that it might be useful to consider the properties of *communities* of networks. Of particular interest here is the fact that in traditional connectionist modeling, the programmer constructs the world of experience from which the networks learn. Modeling communities of networks suggests that the behavior of *other networks* might also be an important source of structure from which each network could learn. Connectionist programmers refer to the output patterns to be learned

⁶We thank Elizabeth Bates (personal communication, Feb., 1991) for coining the term *public hidden units* for this construction.

as the *teachers* for their networks. With a community of networks, we can let an important part of the teacher be embodied in the behavior of other networks. Thus, where traditional network modeling is concerned only with the relation of structure in the environment to internal structure, a model of interactions in a community of networks adds the universe of communicational artifacts to the picture.

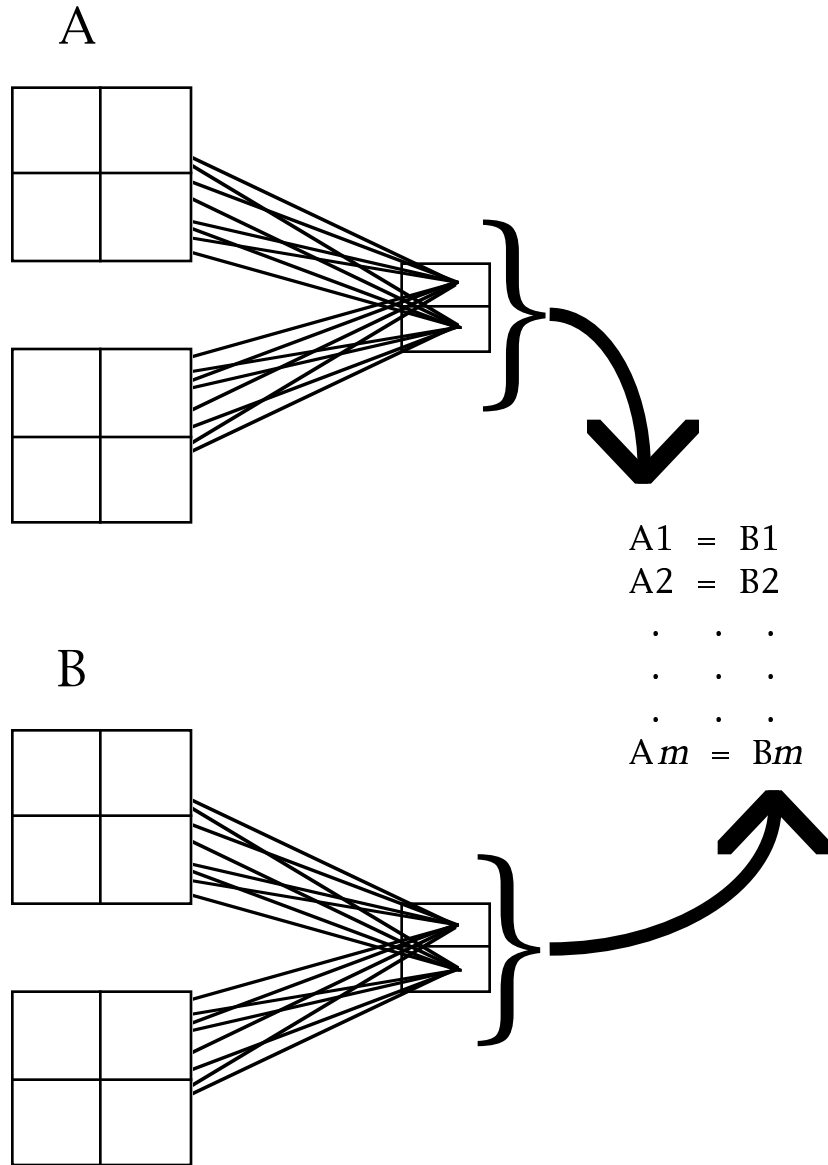


Figure 4. A scheme for evolving consensus on a set of distinctions. (By reciprocally constraining autoassociators at their hidden layers, consensus about the representations used to classify the m scenes can be achieved.)

It is easy to show that consensus among two networks (say A and B) can be achieved by making the output of one the teacher for the other. If each

takes the behavior of the other to be the target, then consensus will result. This satisfies the constraints that $A_1=B_1$, $A_2=B_2, \dots, A_m = B_m$.

IMPLEMENTATION

The simulation proceeds via interactions – one interaction is one time step in the simulation. An interaction consists of the presentation of a chosen scene (from the set of m scenes) to two chosen individuals, a *speaker* and a *listener* (from the set of n individuals). The functions which do this choosing determine what we call the *interaction protocol* of the simulation. The typical functions simply implement random selection from the domains of scenes and individuals, respectively. One of the individuals chosen (say A) responds to the scene by producing a pattern of activation on its verbal output layer (A speaks). The other individual (say B) also generates a representation of what it would say in this context but, as listener, uses what A said as a target to correct its own verbal representation. The listener, (B), is also engaged in a standard learning trial on the current scene, which means its own verbal representation – in addition to being a token for comparison with A's verbal representation – is *also* being used to produce a visual output by feeding activation forward to the visual output layer. The effects of this learning on B's future behavior can be stated as: (1) in this context produce a representation at verbal output more like what A said, and (2) produce a representation at visual output more like the scene itself.⁷

By randomly choosing interactants and scenes, over time every individual has the opportunity to interact with all the others in both speaking and listening roles in all visual contexts. The effect to be achieved is for the population to converge on a shared set of patterns of activation on the verbal output units that makes distinctions among the m scenes. That is, we hope to see the development of a consensus on a set of distinctions.

Below we discuss the results of two different sets of simulations. Simulation One was based upon a more complex network architecture and set of scenes. Simulation Two used the simpler network architecture already shown in Figures 3 and 4. The scenes of Simulation Two were simply the four orthogonal vectors $(1, 0, 0, 0)$, $(0, 1, 0, 0)$, $(0, 0, 1, 0)$, and $(0, 0, 0, 1)$. The purpose of discussing Simulation One is to demonstrate the qualitative effects of evolving a lexicon within a (relatively) complex system space. The purpose

⁷Implementing this combination of error signals is straightforward. One error signal is computed from the difference between produced word and target word, the other error signal is the usual error backpropagated from the visual output layer. These two signals are simply added together, and backpropagated to the visual input layer.

of discussing Simulation Two is to explore more analytically the nature of this kind of dynamical system in a more manageable system space.

RESULTS AND ANALYSIS

Simulation One

In this simulation, each individual is an autoassociator network consisting of 36 visual input units, 4 hidden units, 4 verbal input/output units and 36 visual output units, as shown in Figure 5. Notice that an additional layer of 4 hidden units appears in these networks. These additional resources were required by networks in this simulation in order for the community to converge on a shared lexicon.⁸ The scenes to be classified are 12 phases of a moon, represented as patterns in the 6 X 6 arrays shown in Figure 6.

⁸It is a well-known result from connectionist history that two layers of weights are required to perform mappings from input to output which are not linearly separable (Rumelhart, Hinton, & Williams, 1986). The range of verbal representations that individuals are attempting to map to in this simulation may constitute such a set, thus requiring the extra hidden layer to perform properly. We say "may," because the mapping itself is evolving as the lexicon is being constructed. Another reason for the required extra layer has to do with the large compression of data from input (36 units) to verbal input/output layer (4 units). This compression tends to swamp the verbal output layer with large activation values (even for an untrained network), reducing the network's flexibility to learn. Since the learning problem is composed from the behaviors of other (now inflexible) agents, the community is unable to converge upon a shared lexicon.

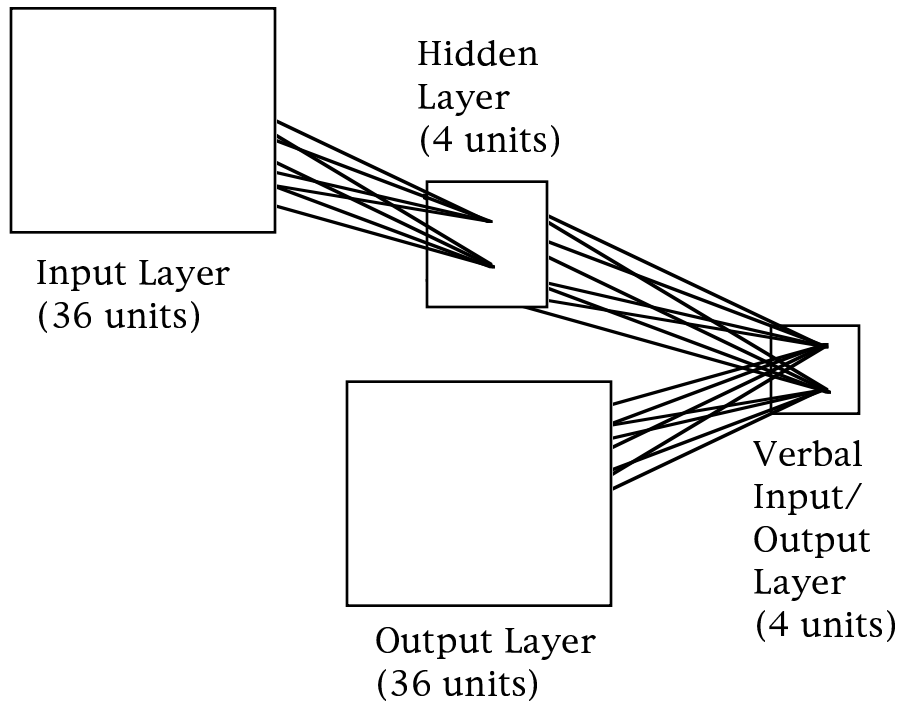


Figure 5. Network architecture for Simulation One. (Not all of the connections between layers are shown.)

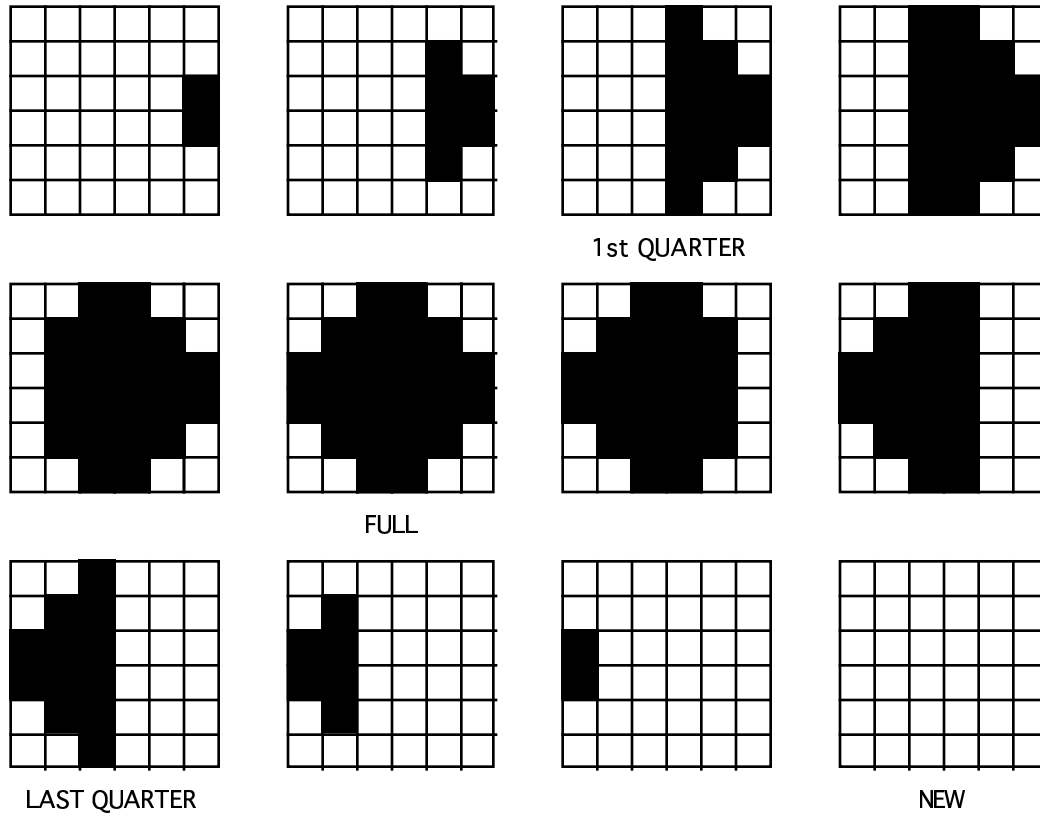


Figure 6. The scenes utilized in Simulation One. (These can be thought of as representations of the visual field associated with sight of the moon in 12 different phases.)

Developing consensus on a set of distinctions appears to be a highly likely final stable state of this dynamical system. Since the initial connection strengths of individuals are mid-range values and randomly assigned, early verbal representations do not differentiate between the scenes. Figure 7 shows the activation levels of the 4 verbal output units in response to the 12 scenes for some typical individuals, at the start of a simulation run. It is easy to see that there is little variation in the response of any individual to the different scenes. It is also easy to see that *consensus* (defined in terms of the degree of variance in terms used by individuals to represent the same scene) is quite high. That is, individuals' responses *do not* carry information which distinguishes the scenes, and these responses *are* highly similar across individuals in the community at the start of the simulation run.⁹

⁹The consensus in the starting state of the simulation is a product of the fact that random weights in the network tend to produce mid-range output values regardless of input to the network.

Figure 8 shows the same individuals after an average of 2000 interactions with each of the other individuals in the 5 member community. For the most part, individuals now respond *differently* to each of the 12 scenes, and all of the individuals *agree* with each other on how to respond. That is, we have a consensus on a set of distinctions. Due to the random starting weights of the networks, and the random interaction protocol functions which organize their learning experiences, there is no way to predict *which* lexicon will develop – but the procedure is robust in the sense that *some* well-formed lexicon or another develops nearly every time.¹⁰

¹⁰See *The formation of dialects*, section (below) for a discussion of some observed exceptions.

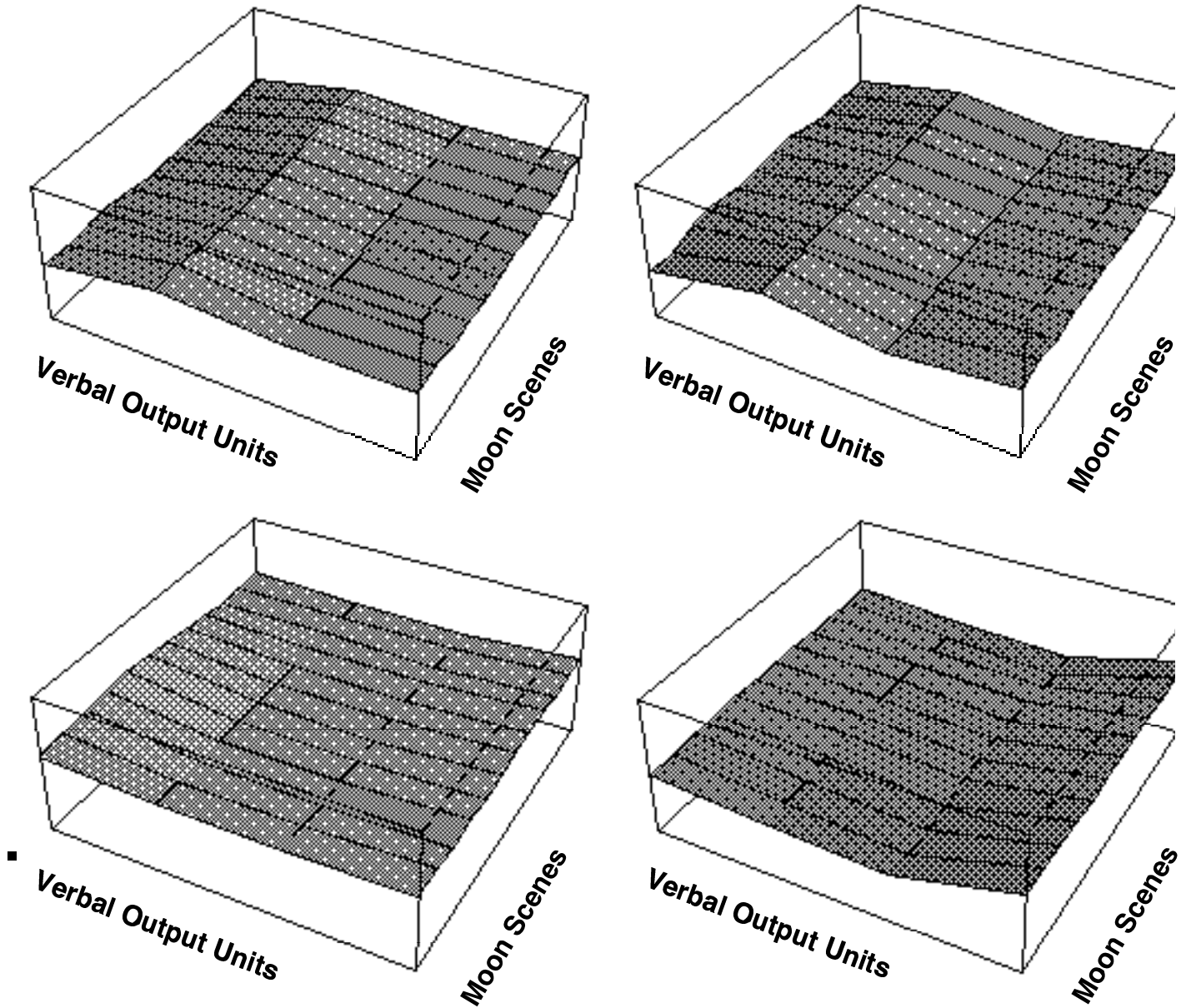


Figure 7. Four individuals of a five-member community at the start of a simulation run. (The surface represents the value of each verbal output unit, in the range 0.0 to 1.0, in response to each moon scene.)

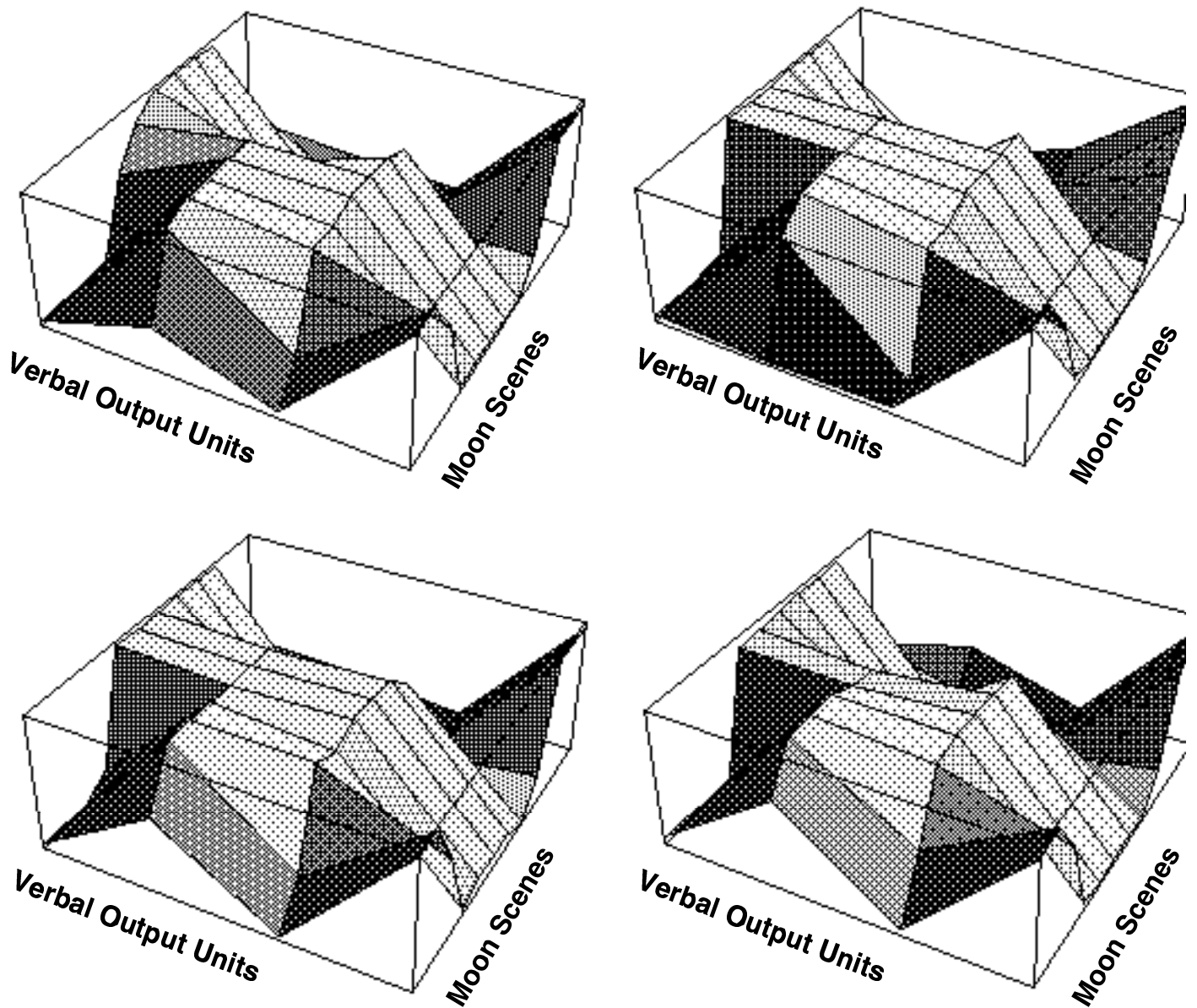


Figure 8. Four individuals of a five-member community after 50000 interactions. (Each individual has had, on average, 2000 interactions with each of the other four individuals. Half of these were in the role of listener and half in the role of speaker. The surface represents the value of each verbal output unit, in the range 0.0 to 1.0, in response to each moon scene.)

Simulation Two

In this set of simulation runs, we attempt to map out more analytically some of the properties of a simpler system. In particular, we view each simulation run as one of a large number of dynamical systems that are possible, given different initial conditions and parameter settings (Abraham & Shaw 1987, following Thom 1972). This infinite-dimensional space D of dynamical systems can be modeled by a function F which maps the following independent variables and functions into an instance of a dynamical system:

Scenes:

m = number of scenes

S = set of scenes $\{s_1, s_2, \dots, s_m\}$.

Individuals:

f_{-arch} = function which determines network architecture of an individual

n = number of individuals in community at start

W = set of starting weights of individuals

μ = learning rate

ψ = learning momentum.¹¹

Interaction protocol:

f_{-pop} = population control function for community

f_{-scene} = function which picks scene for an interaction

¹¹The *learning rate* is a parameter controlling the magnitude of effect one learning trial has on a network, i.e., the scale of magnitude by which changes are made to weights on each learning trial. The *learning momentum* is a parameter which influences the effects that variability in the learning set has upon network learning performance. That is, the learning momentum parameter determines the scale of magnitude by which recent learning trials continue to effect the current learning trial. (See McClelland, 1988, for implementation details regarding these learning parameters.) These two parameters could, conceivably, vary among individuals, perhaps also as functions of time. In the simulations of this paper we chose to fix these parameters, and not let them vary across individuals or time, in order to simplify the task of understanding the set of dynamical systems we are dealing with.

$f-ind$ = function which picks individuals for interaction.

Instantiation of these variables and functions (by the application of F) determines a unique dynamical system which evolves in time (t = cycles of simulation or interactions). In general, we expect different instantiations of these parameters to generate qualitatively different dynamical systems in D.

The parameter settings of simulation two

In order to analyze a small portion of the huge space D, we can make the following simplifications during the remainder of this simulation (i.e. apply F as follows):

Fix scenes to 4 orthogonal vectors of length 4:

$$m = 4$$

$$\mathbf{S} = \{(1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0), (0, 0, 0, 1)\}.$$

Fix individuals with identical network architecture, but initial weights are random values:

$f-arch$ = instantiate all individuals as shown in Figure 3

n = to vary across experiments

\mathbf{W} = a set of random numbers in the range -0.5 to +0.5

Fix learning parameters in time and across individuals:

$$\mu = .075$$

$$\psi = .9.$$

Grant individuals immortality, and no new individuals can be generated. Make interaction protocol functions random:

$f-pop$ = individuals live forever during a simulation run, and no new individuals are introduced.

$f-scene$ = random selection from the set of scenes \mathbf{S}

$f-ind$ = random selection from the community of individuals.

One benefit of establishing the three parameters \mathbf{W} , $f-scene$, and $f-ind$ as random variables is that (given statistically relevant samples of simulation

runs) these parameters can be approximated as fixed, thereby isolating the dynamical systems' dependence upon the one remaining variable, namely the number of individuals in the community (n).

Measures of the emerging lexicon: AVG1 and AVG2

Finally, having established a simulation run (or, preferably, a sample of simulation runs) as an instance of a dynamical system (by setting n), we can monitor the system's evolution with two measures of the community's language through time, $Avg1(t)$ and $Avg2(t)$. $Avg1$ is a measure of the average difference in *each individual's* verbal representations (i.e. it measures the average "term distinctiveness," across all scenes as denoted by each individual, averaged across all individuals). $Avg2$ is a measure of the variability in the *community of individuals'* verbal representations (i.e. it measures the average term variability for each scene across individuals, averaged across all scenes).¹² Our expectation is that $Avg1$ will tend toward 1.0 (the maximum difference in terms used across scenes) and $Avg2$ will tend toward 0.0 (the minimum difference in terms used by the community for each scene) as the system evolves. That is, we expect to see a consensus on a set of distinctions emerge.

Figures 9 and 10 show a simulation run with a community of 5 individuals ($n = 5$). The graphs of Figure 9 show the phase space of verbal representations (Unit 1 activation vs. Unit 2 activation) for each of the four scenes in the world. Each trajectory on a graph represents one citizen's term for that scene, parameterized by time. The trajectories all begin near the middle of each graph (verbal output activations near [.5, .5]) because, at the beginning of the simulation run, individuals are responding to the scenes with unorganized connection weights.¹³ As time proceeds, the trajectories of each graph head for one of the four corners (i.e. consensus regarding each scene increases). Furthermore, each graph's trajectories must reach a corner unoccupied by the trajectories of the other three graphs (i.e. term similarity decreases). Notice how two of the emerging lexicon's terms (Figures 9(a), 9(b)) compete with each other for a place in the (1, 0) corner before the term representing scene (0, 1, 0, 0) of Figure 9(b) finally wins this competition.

The two graphs of Figure 10 show how $Avg1$ and $Avg2$, plotted for the same simulation run as that shown in Figure 9, capture the two properties of the evolving system. Descriptions begin with a lot of consensus but lack

¹²See Appendix for a more formal definition of these measures.

¹³There has been no learning yet, and all individuals begin with random weight assignments to their connections, therefore, all units respond at mid-range levels of activation.

discrimination because, again, the networks are responding roughly the same (and un-informatively) at the beginning of the simulation. Between 1000 and 2000 interactions, as good representations for discriminating between scenes emerge (Avg1, the mean variability in each individual's descriptions for *different scenes*, goes up), the degree of consensus goes down (Avg2, the mean variability in descriptions representing the *same scene*, across individuals, goes up). As the rate of learning to discriminate between scenes slows down (Avg1 approaches the asymptote), the representations which serve as individuals verbal targets *change more slowly* – they become easier targets to follow, and consensus begins to emerge (Avg2 begins to descend again.)

<INSERT FIGURE 9 HERE>

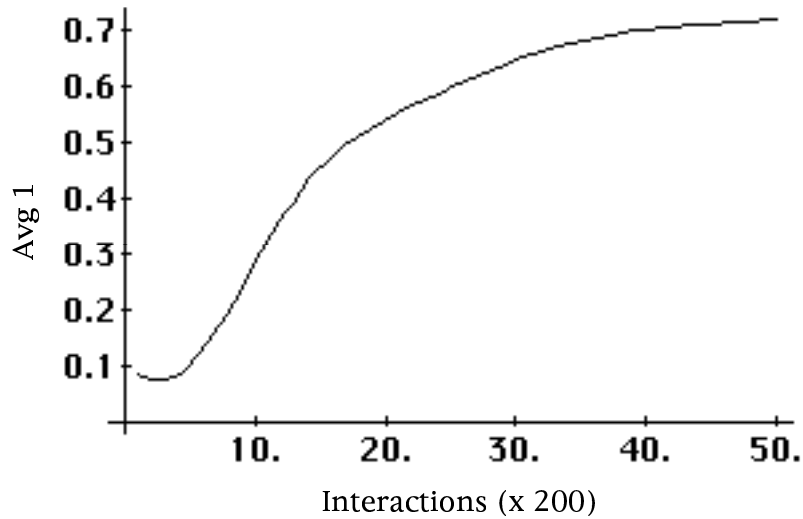


Figure 10(a). Emerging description discriminability. (Avg1 plotted as a function of simulation time steps, shows that the mean term variability across scenes increases as the system evolves. The data shown is from the same simulation run as that shown in Figure 9.)

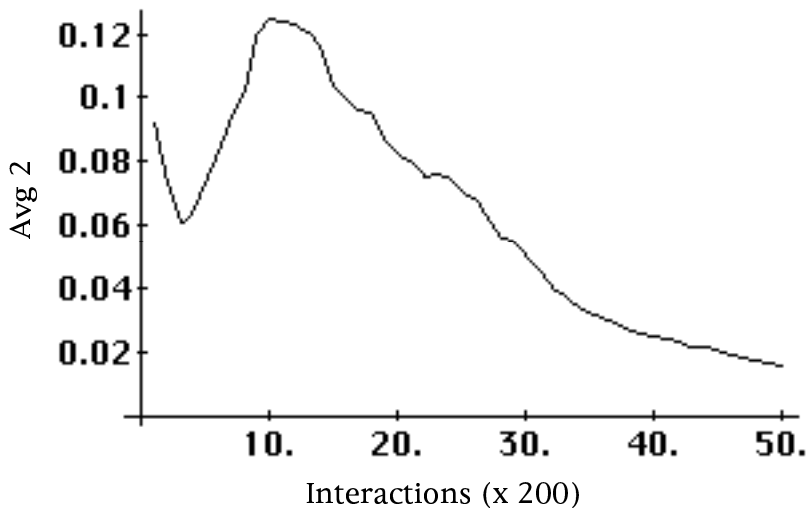


Figure 10(b). Emerging description consensus. (Avg2 plotted as a function of simulation time steps, shows that the mean term variability across individuals decreases as the system evolves. The data shown is from the same simulation run as that shown in Figure 9.)

The effect of varying community size

We are now in a position to ask how the dynamical system depends on population size n , given that we have fixed all of the other system parameters as discussed above. Figure 11 shows the results of 3 different experiments, each entailing a sample of 15 simulations. The simulations of each experiment were run with community member sizes n of 5, 10, and 15, respectively. The means and one standard deviation error bars for the 15 observations of each experiment (sampled at regular intervals within each simulation) are shown for the two measures of lexicon structure, Avg1 and Avg2. In all three experiments, individuals have participated (on the average) in the same number of interactions (namely, 2000) by the end of the time frame shown in Figure 11. The general pattern is the same as that seen in the single simulation run of Figure 10. Of particular interest, is the nature of the “decay” in the lexicon formation process shown by changes in the two measures of lexicon structure, Avg1 and Avg2, as community size gets larger. This decay is due to the greater difficulty of organizing large communities than small ones. Each experiment displayed in Figure 11 shows that Avg1 and Avg2 of the “average community” (represented by the plotted mean values of the 15 simulations in each graph) vary smoothly and asymptotically as a function of number of interactions. Therefore, the final steady-state of each experiment can be reasonably approximated by the final mean value of each graph in Figure 11. Taking the three points so collected for each measure (Avg1 and Avg2), it appears that the decay in the ability of a community to form a “good” lexicon increases exponentially by a factor of $1/n$. As community size n increases, the rate at which lexicons become less “good” slows down. This relationship, although limited to only three data points, is clearly represented in Figure 12. Of course, the meaning of what “good” is (and how good is “good enough”) can only be determined by the functional properties entailed in agents *using* the lexicon, something these simulations do not address. (But see Hutchins & Hazlehurst 1991, for how such a lexicon can be embedded in a larger context of community use.)

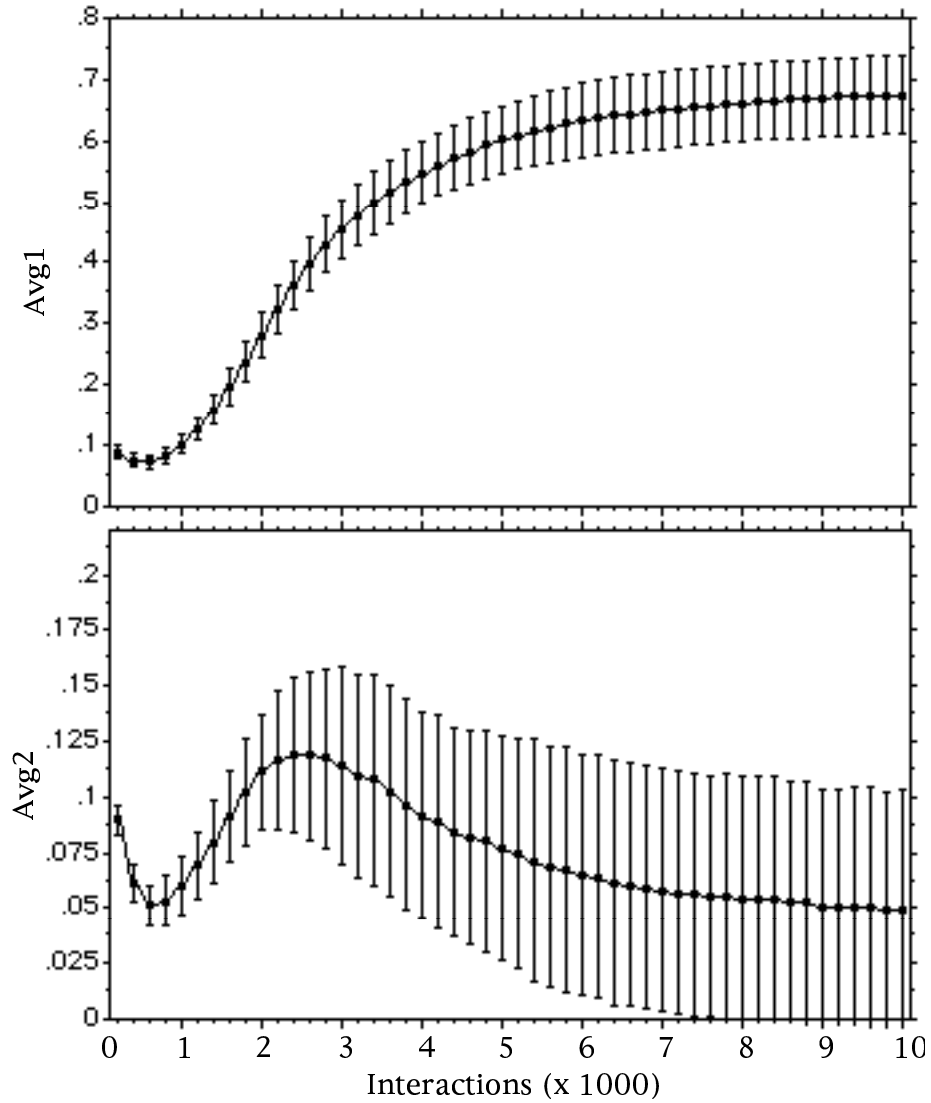


Figure 11(a). A sample of 15 simulations of community size 5 ($n = 5$). (Each individual in each simulation run participates, on average, in 2000 interactions.)

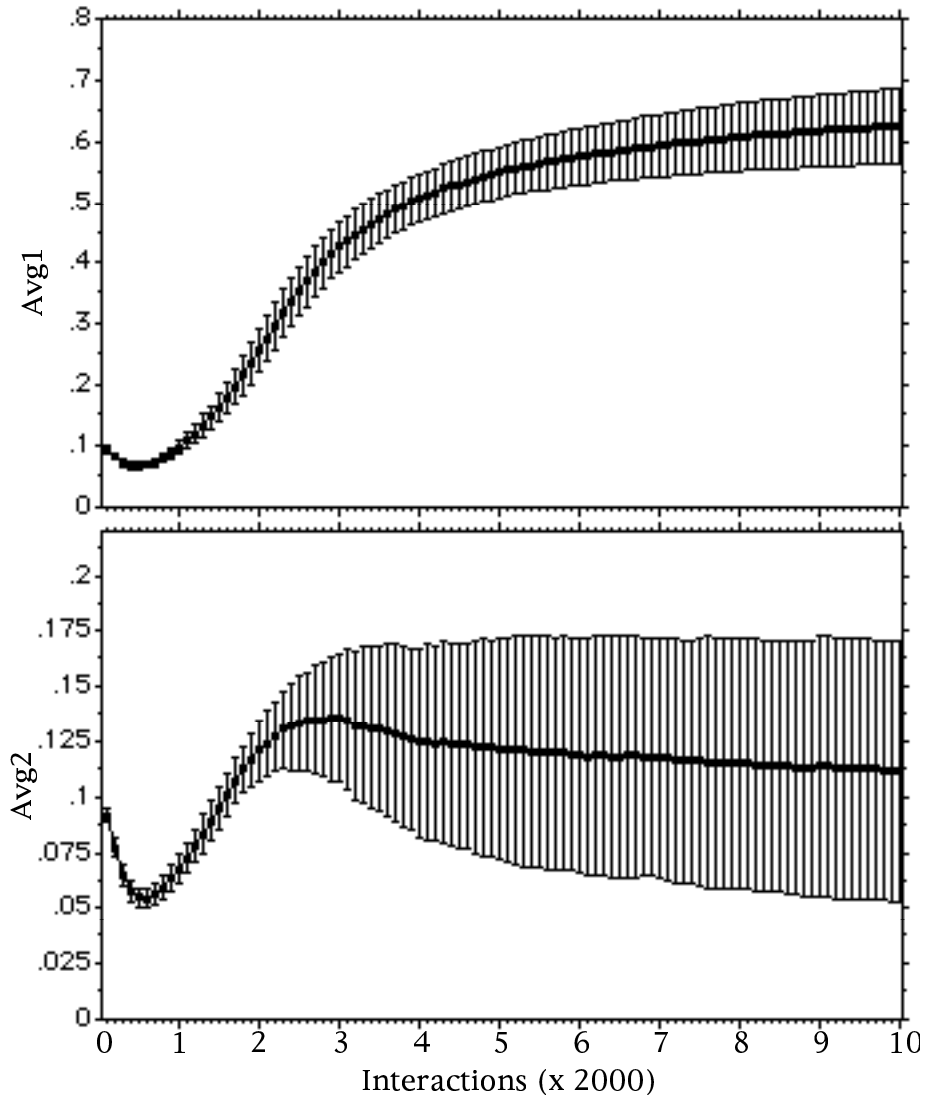


Figure 11(b). A sample of 15 simulations of community size 10 ($n = 10$). (Each individual in each simulation run participates, on average, in 2000 interactions.)

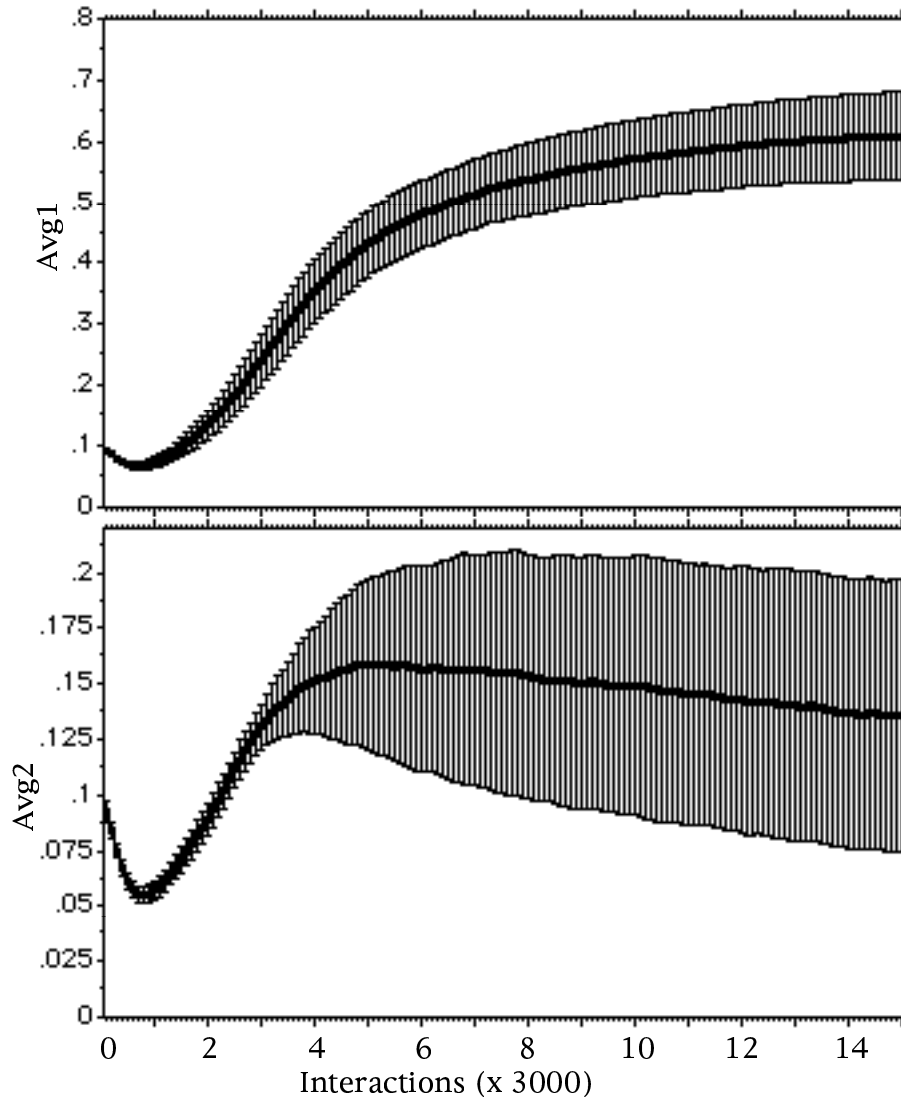


Figure 11(c). A sample of 15 simulations of community size 15 ($n = 15$). (Each individual in each simulation run participates, on average, in 2000 interactions.)

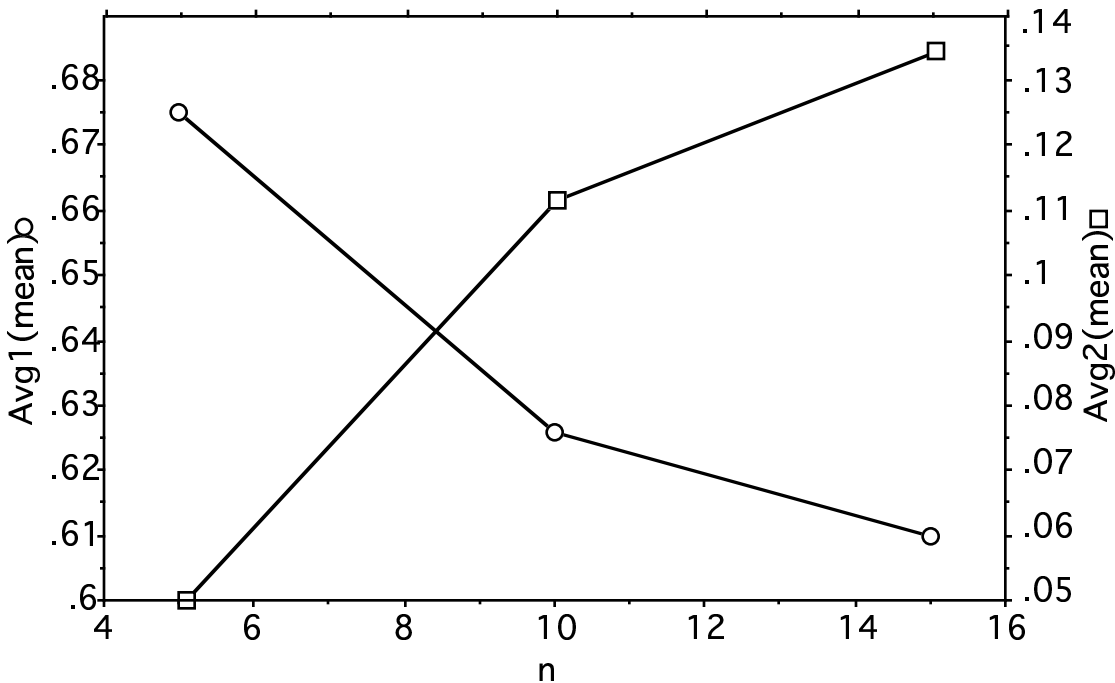


Figure 12. A representation of the effects of varying community size (n) on lexicon structure. (Each point shows the final mean values from the three experiments of Figure 11. These points represent estimates of the final steady-state of each of those dynamical systems.)

DISCUSSION

The Need For A Critical Period Of Language Learning

We have experimented with adding new individuals with random weight-structures to communities that have already developed a lexicon. Depending on the size of the community, the addition of the new individual may have quite different effects. The behavior of a new individual added to a large community with a highly shared lexicon will be entrained by the behaviors of the other members of the community and the newcomer will learn the shared lexicon. A new individual added to a small community may completely destroy the previously achieved solution. After such an event the community may or may not be able to recreate a well-formed lexicon with the new individual.

In running these simulations we found ourselves wishing that there was some principled way to reduce the learning rate once individuals had learned the language. In particular, one would like to reduce the learning rate at the point in the life cycle where individuals are likely to encounter many interactions with "disorganized" individuals. This would amount to implementing a critical period for language learning so that individuals learn

less from the linguistic behavior of others once they have reached sexual maturity.¹⁴ Perhaps evolution has engineered something like this into our species. We did not implement a critical period, however, because to do so seemed arbitrary and a violation of one of the core premises: The processes that account for normal operation of the system should also account for its development through time. In more complex situations, like that of biological evolution where adaptive searches are conducted in parallel at many levels of specification, it may be reasonable to expect violations of this premise.

The Formation Of Dialects

Occasionally, communities fail to create a well-formed lexicon. This happens because sometimes the random initial starting points of the networks in a community are incompatible with each other, and "unlucky" choices of the interaction protocol lead to divergence in the verbal representations of these individuals, which can not be overcome. In this case, the kind of language that one individual is predisposed to learn is not the sort that the other is predisposed to learn, and given their starting points and learning experiences, they never find a solution that can be shared.

The fact that some pairs of initial weight configurations are more compatible than others suggested that it might be advantageous to let the individuals discover and seek out those with whom they are compatible (as demonstrated in the similarity of their behaviors, i.e. their verbal representations). A sequence of experiments was run in which the choice of a listener for each speaker was biased in favor of those who had a history of speaking (i.e. using descriptions in former interactions between the two) which was similar to the speaker's own history of utterances. The result of implementing this interaction protocol was the formation of *dialects*, or clusters of individuals within which there is consensus on the descriptions and their referents. Since individuals interact more with "like minded" others, they are more susceptible to learning from the classification behaviors of those who act like they themselves do. We take this to be a very primitive implementation of what Sperber (1985) called *ecological patterns of psychological phenomena*.

¹⁴Of course, an alternative solution might be simply for individuals *not* to focus their learning lens too narrowly on any other single individual (or homogeneous group of individuals). The problem with this solution is that it works for organized (adult) individuals but doesn't work for disorganized (novice) individuals. See *The acquisition of a coherent lexicon* section.

The Acquisition Of A Coherent Lexicon

Consistent with the observations reported above about dialect formation and the education of new individuals, is an interesting interpretation of the model's performance regarding the ontogenetic problem of learning form-meaning pairs that are already established in the world. In particular, it seems that the complement to the need for a critical period of language learning (which ensures that the experience of older individuals is not lost by attending to the unorganized ramblings of novices) is the need for a set of consistent models during language acquisition (which ensures that novices are exposed to a system that is coherent, and therefore learnable).

Creating a lexicon from total lack of organization (demonstrated in Simulations One and Two), actually seems to be easier than creating a functional lexicon from a system that is organized in the wrong way. That is, novices introduced into the system who must accommodate a wide range of organized variance (as with dialects) have a hard time learning any part of the form-meaning system. On the other hand, new individuals exposed to well-formed lexicons (or one well-formed dialect) gain significant leverage on the problem of learning the form-meaning system via the mediation of consistent and well-formed terms. In fact, experiments were run which showed that the strictly visual problem of classifying scenes in the simulation world is simplified by the use of consistent and well-formed terms – *individuals abilities to learn the visual classification problem are enhanced by the existence of a coherent lexicon.*

This fact of the simulation seems to be due to the nature of the principal-component decomposition being conducted by the autoassociator in its learning (Chauvin 1988). By grounding this process in the constraints imposed at the hidden layer by coherently organized targets, the decomposition process is significantly accelerated. There is a sense, then, in which the structure of the lexicon is an important vehicle for the learning of the visual classification problem.

In the real world, we can cite two kinds of evidence which are consistent with this aspect of our model's behavior. The first has to do with the nature of language-acquisition environments. Children generally *do* learn language, over an extended period of time, in a relatively homogeneous population of language users. This is a fact which follows from the general nature of our species social organization. Second, there appears to be a good deal of empirical evidence for the fact that children *must* (or at least that they *act as though they must*) accommodate each and every form-meaning pair that they encounter (Clark 1983, 1987, Slobin 1985). That is, the language acquisition process apparently requires the learner to assume that words contrast in meaning, and children use this as a resource in classifying the objects or events which words denote (Clark, 1987).

Term Contrast And Consensus In Natural Language

Clark (1987) investigated the natural language version of our notion of a lexicon being a *shared set of distinctions*. Clark provides empirical evidence consistent with the claim that natural language lexicons exhibit two pervasive features:

1. Every two forms contrast in meaning (the principle of contrast, p. 2).
2. For certain meanings, there is a conventional form that speakers expect to be used in the language community (the principle of conventionality, p. 2).

The general claims made are that the systems which generate human lexicons are efficient (all words contrast in meaning) and conservative (established words have priority, and innovative words fill lexical gaps as opposed to replacing established words with the identical meanings) Evidence for these principles are cited from a wide range of natural language phenomena, including the observations that:

1. Lexical domains emphasize semantic contrasts.¹⁵
2. Syntactic constructions create semantic contrast. “[D]ifferences in form mark differences in meaning at both the lexical and the syntactic levels” (p. 6).
3. Well-established irregular forms in a language are maintained in the face of many resources (paradigms or patterns) for simplifying the language through regularization.
4. Innovative (new) words emerge as a consequence of a failure in an existing lexicon’s capacity for conveying the appropriate meaning (i.e. due to an inability of the existing words to establish the appropriate contrasts).

These principles of human language have parallel entailments for the acquisition of language.

1. Children rely on contrasting terms to tune their understanding of semantic fields to adult levels of specificity and generality.
2. Children assume (or at least act as though assuming) that new terms contrast with those that they already know.

¹⁵Clark claims that true synonyms do not exist in natural languages.

3. Children reject (both across and within languages) terms which they understand to be synonyms with terms that they already know.
4. Children productively generate novel terms to fill expressive needs, but these terms converge toward conventional usage as their language development proceeds.

We find Clark's analysis to be in general agreement with the phenomena modeled and suggested by our simulations, although her observations clearly exceed the range of phenomena covered by our simple simulations.

Grounding Meaning In Communicatory Practices Of The Community

We find Clark's analysis lacking in two respects: the first is made explicit by our model, the second is suggested by the general theoretical framework we have adopted although not specifically captured in the simulations we have presented here.

First, Clark's formulation tends to equate language use with form-meaning pairing, which lapses into a notion of meaning that is structurally derived rather than grounded in communicatory practice. This appears to stem from an analysis which takes language *form* as a proximal explanation for *function*.¹⁶ In particular, without an explicit means of grounding terms in the world of experience, meaning becomes too tightly attached to form, forcing Clark into use of the *conduit theory of meaning* (Reddy 1979, Lakoff 1987).

The conduit theory of meaning takes meaning to be something transferred between language users, as if meaning is *attached to* language forms, rather than something which expresses a relationship between perceiver/actor, context, and experience as a *consequence of* situated processing of language forms. For example, terms in the lexicon need only contrast to the extent that the communicatory functions required of the

¹⁶This fact seems to get Clark into trouble in her critique of Slobin's notion of *unifunctionality*—which denies the existence of multiple forms carrying the same meaning (something Clark agrees with), *and* denies the existence of multiple meanings being carried by the same form (something Clark disagrees with; 1987, p. 25). Clark claims, for instance, that the English inflection *-s* is a form used productively to map onto the concepts of plurality and possession. Clark's argument here is based *solely* on evidence from the structural regularities of parts of English morphology, effectively ignoring the possible communicative and learning functions which act to create contrasts even here. These are the properties of natural language which Clark relies upon to build her own case elsewhere in the paper. Furthermore, Clark (1987, p. 26) utilizes a logical analysis of semantic feature inheritance as an argument for rejecting Slobin's denial of true homonymy in natural language. This seems to be inconsistent with her use of language's communicative functions found elsewhere in her paper.

lexicon by the agents involved are served. Clearly, these functions will vary across language situations, participants, and traditions – because contexts, individual experiences, and the histories of use vary with each of these. This latter view of meaning becomes clear when one distinguishes between, and considers the interactions among, conventional forms of behavior (artifactual structure), individual experience (internal structure), and the physical world (natural structure). In our own simulations, this functional grounding of meaning is evident in the ways the lexicon (artificial structure) responds to tweaking different simulation parameters (natural structure). It is also evidenced in the variability of individual network weight configurations (internal structures) which can accommodate the same lexicon (artifactual structure) given stable environments (natural structure).

Second, Clark's analysis undervalues the explanatory power of one of her own principles – conventionality. The principle seems to be given the role of *describing* the structure of language, but no causal role in *creating* that state of affairs. For example, in explaining the fact that children's private and idiosyncratic novel words give way to conventional expressions, Clark cites children's *efforts to contrast terms* as the mechanism responsible for convergence toward use of the standard forms. "It is children's discovery that two forms do *not* contrast in meaning that leads to take-over by the established term," she says (ibid:18, emphasis in original). It seems to us that Clark is ignoring a large space of communicatory functions responsible for explaining why children adopt conventional expressions. Again, structural analysis provides only proximal explanations for the mechanisms involved. Our own modeling framework keeps in focus the higher order effects of language sharing – *consensus is a functionally important property in it's own right* – and seems to play a more active role in answering the question of why children conform to an established norm than Clark posits.

In brief, a community of language users (and generations of language users) constitutes a system which enables cognitive performance that can not be performed by individuals alone (Hutchins & Hazlehurst 1991). Children's convergence toward the use of established expressions would seem to be importantly related, not only to the individual-level problem of making meaningful distinctions in the here-and-now, but also to the community-level problem of constructing which distinctions are meaningful. Conventionality points to a complex cultural process – generating many language-shaping communicative properties – which only becomes clearer by taking a community of interacting language users as the unit of analysis.

THE MODEL AS THEORY INSTANTIATION

Our model explicitly represents the interactions of the three kinds of structure discussed in the beginning of the paper: natural, internal and artifactual. The patterns representing physical phenomena of the world

(scenes) are the natural structure. The patterns of activation on the verbal input/output units are the artifactual structure. The connection strengths in the networks are the internal structure that provide coordination between the two kinds of external structure, and are themselves a product of artificial structure mediating experience with the world. We see this as the smallest first step toward a system in which artifactual structures invoke the experience of that which is not present in the environment.¹⁷

Let us return to the central theoretical assumptions. As we have seen, no individual can influence the internal processing of another except by putting mediating artifactual structure in the environment of the other. However, by putting particular kinds of structure in each other's environments, they all achieve a useful internal organization.¹⁸ It is possible for each individual to achieve an internal classification scheme in isolation – this is what autoassociators are known to do by themselves. But such a classification would be useless in interaction with others. That is, idiosyncratic distinctions may be useful, but not as useful as shared ones. We have noted that learning to categorize the world is easier when mediated by coherently organized verbal representations. By forcing individuals to learn from the classification behavior of others we ensure that each individual can only become internally organized by interacting with the external products of the internal organization of others. The effects of this kind of system also enable individuals to tap the resources of an entire group (and ancestors of the group), enabling cognitive performance not achievable by individuals alone. This is the foundation upon which human intelligence is built (Hazlehurst 1994, Hutchins & Hazlehurst 1991, Hutchins in press).

Although this simulation is too simple to address the issue of symbolic representation directly, it suggests a way in which shared symbols that could subsequently come to serve internal functions could arise as a consequence of social interaction. Such symbols are outside the individual first as pieces of

¹⁷Work on the slightly longer step of developing propositional representations and letting symbols break free of the phenomena to which they refer is now in progress (cf. Hutchins & Hazlehurst 1991).

¹⁸Here we must acknowledge a hedge on our own claims. The model, as it stands, collapses semantic and phonological representations. There is no distinction between what an agent *conceives* of the scene and what it *says* about the scene. Likewise, what an agent says is unproblematically heard by the other agent participating in the interaction. This is, strictly speaking, a violation of the no-telepathy assumption, and the separation of internal and artifactual structure. Having acknowledged this discrepancy we can state that the model, as it stands, gains no explanatory power from this conflation and is therefore not a violation in principle. We have experimented with architectures which produce phonological representations from semantic representations, and vice-versa. Implementation of these constructions do not render invalid the general claims of the model presented here.

organized material structure – in the behavior of others – before they have explicit internal representations. Undoubtedly, such shared public forms can be given internal representations, as can any structural regularity in the environment whether natural or artifactual. This perspective, in which symbols are in the world first, and only represented internally as a consequence of interaction with their physical form and social consequences, is what we mean by the *shallow symbols* hypothesis. In this view, symbols and symbolic processing may be relatively shallow cognitive phenomena, residing near the surface of the functional organizations which are the result of interacting with material structures in a cultural process.

The computations performed by the networks are well characterized in terms of the propagation of representational state. The universe of inputs is propagated through the networks and re-represented at the output. These representations, or rather the functional capacities to produce them, then become distributed across the members of the community. This general notion of computation comes from Simon (1981, p. 153) who says, "Solving a problem simply means representing it so as to make the solution transparent." Simon may not have intended quite so broad a reading of his definition but it seems to capture well the behavior of this system. The structure of the natural world is fixed in this model, but the internal structures and the artifactual structures co-determine each other and co-evolve in the development of the lexicon. In the broadest sense, the solution arrived at was determined by the structure of the natural world as manifested in the phenomena encountered, in the random initial configurations of the internal states of the individuals, and in the instantiation of who actually learns from whom in the community. The process of developing a lexicon in this model is a process of propagating transformed representations of naturally occurring structure throughout a system that contains artificial structure as well.

Finally, even when two networks are in complete agreement with each other about the use of the lexicon, each has a unique internal structure. Individuals in our model are able to use the lexicon *without* needing to share the internal structures which enable that use. Through learning from each other in interaction, individuals become functional equivalents, not structural replicates of each other. There is no need to posit the existence of grandmother neurons which are responsible for the like behaviors of autonomous individuals. On the other hand, behaviors *are* shaped by the constraints of autonomous individuals interacting in (problematically) shared environments.

Within this theoretical framework, we claim that meaning can be retrieved from the unattractive positions of being equated with: (a) the results of a (usually innate) private language of mental symbols which stand for an uncontested, fixed, and nonsocial objective reality (Fodor 1976; cf. Lakoff

1987), or (b) an unproblematically shared, static (i.e. a nondevelopmental, nonhistorical, and often nonsocial) semantic knowledge base (ethnoscience and much of cognitive anthropology; cf. Hutchins 1980), or (c) a strictly public (i.e. superindividual, nonmental) symbol system (Geertz 1973; cf. Shore 1991). Meaning in our model is an evolving property of the interaction of internal, artificial, and natural structures. At any point in the evolution of such a system we take meanings to be characterizations of the functional properties of individuals' viz.-a-viz. the environment (including each other). Meaning, for each individual, describes a range of possibilities for action. In a more complicated world, these possibilities would be constrained by representation of the consequences of those actions, something we understand to be true of human meaning systems. In the limited world we have created, the range of possibilities for action is restricted to producing something like monolexic utterances in the contexts of shared binary scenes. In the real world the possibilities are much greater, some of which we hope to address in future work.

APPENDIX

In a community of agents who live in a world with scenes, the community language at time t (the recording of all agents descriptions of the scenes) can be written in matrix form as

$$L_t = \begin{pmatrix} S_{1,1}(t) & S_{1,1}(t) & S_{1,1}(t) & \dots & S_{1,1}(t) \\ \vdots & \ddots & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ S_{1,1}(t) & S_{1,1}(t) & S_{1,1}(t) & \dots & S_{1,1}(t) \end{pmatrix}$$

where $S_{i,j}(t)$ is the j^{th} agent's description of the i^{th} scene at time t . Then the sequence of matrices $\{L_0, L_1, \dots, L_\varphi\}$ represents the evolution of the community language from time $t = 0$ to $t = \varphi$. At any time, the language L_t can be analyzed for at least two properties

- (1) Avg1, the average difference between all pairs of descriptions of scene for each individual
- (2) Avg2, the average difference between pairs of individuals descriptions of the same scene for all individuals and all scenes

Avg1 gives a measure of individual abilities to use terms to distinguish between scenes, while Avg2 gives a measure of the community ability to agree on the terms used to identify scenes

More formally assume $S_{i,j}(t)$ and $S_{l,p}(t)$ are real valued vectors of length γ , then define a distance metric

$$d(S_{i,j}(t), S_{l,p}(t)) = \sqrt{\frac{\sum_{k=1}^{\gamma} (r_{i,j}^k - r_{l,p}^k)^2}{\gamma}},$$

where $r_{i,j}^k$ is the k^{th} real value in vector $S_{i,j}(t)$.

Then,

$$Avg\ 1(t) = \frac{\sum_{j=1}^n \left(\frac{\sum_{(i_1, i_2) \in P_2(m)} d(s_{i_1, j}(t), s_{i_2, j}(t))}{\frac{m^2 - m}{2}} \right)}{n},$$

where $P_2(m)$ is the set of pairs of integers from 1 to m , and $\frac{m^2 - m}{2}$ is the size of this set

Similarly,

$$Avg\ 2(t) = \frac{\sum_{i=1}^m \left(\frac{\sum_{(j_1, j_2) \in P_2(n)} d(s_{i, j_1}(t), s_{i, j_2}(t))}{\frac{n^2 - n}{2}} \right)}{m}.$$

ACKNOWLEDGMENTS

Research support was provided by grant NCC 2-591 to Donald Norman and Edwin Hutchins from the Ames Research Center of the National Aeronautics & Space Administration in the Aviation Safety/Automation Program. Everett Palmer served as technical monitor. Additional support was provided by a fellowship from the John D. and Catherine T. MacArthur Foundation.

Correspondence and requests for reprints should be sent to Edwin Hutchins, Department of Cognitive Science, University of California at San Diego, La Jolla, CA 92093-0515 USA.

BIBLIOGRAPHY

- Abraham, R. & C. Shaw 1987. Dynamics: A visual introduction. In *Self organizing systems*, F. E. Yates (ed.), 543--97. New York: Plenum.
- Bryne, R. & A. Whiten (eds) 1988. *Machiavellian intelligence*. New York: Oxford University Press.
- Chauvin, Y. 1988. *Symbol acquisition in humans and neural (PDP) networks*. Unpublished doctoral dissertation, University of California, San Diego.
- Clancey, W. 1989. Ontological commitments and cognitive models. *Proceedings of The Eleventh Annual Conference of the Cognitive Science Society*. Ann Arbor, MI.
- Clark, E. 1983. Convention and contrast in acquiring the lexicon. In *Cognitive development and the development of word meaning*, T. B. Seiler & W. Wannenmacher (eds), 67--89. Berlin: Springer-Verlag.
- Clark, E. 1987. The principle of contrast: A constraint on language acquisition. In *Mechanisms of language acquisition*, B. MacWhinney (ed.), 1--33. Hillsdale, NJ: Lawrence Erlbaum Assoc.
- Fodor, J. 1976. *The language of thought*. Sussex: Harvester Press.
- Freyd, J. 1983. Shareability: The social psychology of epistemology. *Cognitive Science*, 7, 191--210.
- Geertz, C. 1973. Thick description: Toward an interpretive theory of culture. In *The interpretation of cultures*, 3--30. New York: Basic Books.
- Goody, J. 1977. *The domestication of the savage mind*. Cambridge: Cambridge University Press.
- Hazlehurst, B. 1991. The cockpit as multiple activity system. Unpublished manuscript, University of California, San Diego.
- Hazlehurst, B. 1994. *Fishing For Cognition: An ethnography of fishing practice in a community on the west coast of Sweden*. Unpublished doctoral dissertation, University of California, San Diego.
- Hinton, G., & S. Becker 1989. An unsupervised learning procedure that discovers surfaces in random-dot stereograms. Unpublished manuscript, University of Toronto, Canada.
- Hutchins, E. 1980. *Culture and inference*. Cambridge, MA: Harvard University Press.

- Hutchins, E. 1990. The technology of team navigation. In *Intellectual teamwork: Social and technical bases of cooperative work*, J. Galegher, R. Kraut, & C. Egido (eds), 191--219. Hillsdale, NJ: Lawrence Erlbaum Assoc.
- Hutchins, E. 1991. Social organization of distributed cognition. In *Perspectives on socially shared cognition*, L. Resnick, J. Levine, & S. Teasley (eds), 283--387. Washington DC: The American Psychological Association.
- Hutchins, E. 1993. How a cockpit remembers its speed. Unpublished manuscript, University of California, San Diego.
- Hutchins, E. (in press). *Cognition in the wild*. Cambridge, MA: MIT Press.
- Hutchins, E. & B. Hazlehurst 1991. Learning in the cultural process. In *Artificial life II*, C. Langton, C. Taylor, D. Farmer, & S. Rasmussen (eds), 689--706. Redwood City, CA: Addison-Wesley.
- Hutchins, E. & T. Klausen (in press). Distributed cognition in an airline cockpit. In *Cognition and communication at work*, Y. Engeström & D. Middleton (eds). New York: Cambridge University Press.
- Lakoff, G. 1987. *Women, fire, and dangerous things*. Chicago: University Of Chicago Press.
- Levinson, S. (in press). Interactional biases in human thinking. In *The social origins of intelligence*, E. Goody (ed). New York: Cambridge University Press.
- McClelland, J. 1988. *Explorations in Parallel Distributed Processing*. Cambridge, MA: MIT Press.
- McClelland, J., Rumelhart, D., & The PDP Group 1986. *Parallel distributed processing: Volume 2, explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.
- Reddy, M. 1979. The conduit metaphor: A case of frame conflict in our language about language. In *Metaphor and thought*, A. Ortony (ed.), 284--324. New York: Cambridge University Press.
- Rumelhart, D., McClelland, J., & The PDP Group 1986. *Parallel distributed processing: Volume 1, foundations*. Cambridge, MA: MIT Press.
- Rumelhart, D., Hinton, G., & Williams, R. 1986. Learning internal representations by error propagation. In *Parallel distributed processing: Volume 1, foundations*, D. Rumelhart, J. McClelland, & The PDP Group (eds), 318--62. Cambridge, MA: MIT Press.
- Shore, B. 1991. Twice-born, once conceived: Meaning construction and cultural cognition. *American Anthropologist*, 93, 9--27.

- Simon, H. (1981). *The sciences of the artificial*. Cambridge, MA: MIT Press.
- Slobin, D. 1985. Crosslinguistic evidence for the language-making capacity. In *The crosslinguistic study of language acquisition*, D. Slobin (ed.), 1157--260. Hillsdale, NJ: Lawrence Erlbaum Assoc.
- Sperber, D. 1985. Anthropology and psychology: towards an epidemiology of representations. *Man*, 20(1).
- Thom, R. 1972. *Stabilite structurelle et morphogenese: Essai d'une theorie generale des modeles*. New York: Benjamin.