Information and Meaning: Use-Based Models in Arrays of Neural Nets

PATRICK GRIM, PAUL ST. DENIS and TRINA KOKALIS
Group for Logic and Formal Semantics, Department of Philosophy, SUNY at Stony Brook, Stony Brook, NY 11794, USA; E-mail: pgrim@notes.cc.sunysb.edu

Abstract. The goal of philosophy of information is to understand what information is, how it operates, and how to put it to work. But unlike 'information' in the technical sense of information theory, what we are interested in is *meaningful* information. To understand the nature and dynamics of information in this sense we have to understand meaning. What we offer here are simple computational models that show emergence of meaning and information transfer in randomized arrays of neural nets. These we take to be formal instantiations of a tradition of theories of meaning as use. What they offer, we propose, is a glimpse into the origin and dynamics of at least simple forms of meaning and information transfer as properties inherent in behavioral coordination across a community.

1. Introduction

There is much that we want to know about information. For starters, we want to know just what information *is*. We want to understand the dynamics of information systems, we want to be able to judge strategies for information acquisition, and we want to engineer efficient techniques for information filtering, management, and application. We want to know what information is, how it operates, and how to put it to work.

The term 'information', however, carries a range of senses. It is perhaps significant for the approach we take here that the earliest occurrences for 'information' listed in the OED are forms of a verb: as 'instigation' is to 'instigate', so 'information' is to 'inform'. Only later does the term appear as a designator for something conveyed in the course of information: a piece of knowledge or an item of intelligence. For our purposes it is unfortunate that the term also carries a defined technical sense within information theory. The questions we pursue here have far more to do with the natural sense of 'information' than the technical sense; in order to clear the ground for what we regard as the interesting philosophical work it is important to clearly identify the technical sense and to put it aside.

In ordinary language, 'information' carries clear ties to meaning: something must be meaningful in order to qualify as information. 'Meaningful information' is quite close to being redundant, and it is not clear that any information could be literally meaningless. We do complain that 'Mrs. Bumpus rambled on and on, belaboring us with meaningless information,' but of course in such a case the information is really only *relatively* meaningless; of little or no significance for the projects we have in hand. Had Mrs. Bumpus literally been gabbling nonsense

Minds and Machines **14:** 43–66, 2004.

syllables, we certainly would not have referred to what she was giving us as meaningless *information*. If information is tied to meaning, it is unlikely that we will be able to fully understand information in the normal sense without understanding meaning. The modeling work we outline below attempts to track the two together.

In the sense familiar within information theory, on the other hand, 'information' is defined in terms of syntax and need have no clear ties to meaning. There the term is defined purely as a measure of probabilities over syntactic strings within some larger pool. What are commonly spoken of as 'messages' in information theory are strings defined purely syntactically, as concatenations of syntactic units. The 'information' carried by a syntactic string is simply a measure of the improbability of randomly drawing that particular string from a larger set; the lower the probability of a syntactic string, the more 'information' it is said to carry. Despite the considerable interest and applicational importance of information theory, it must be said that the theory is couched in misleadingly semantic terminology. 'Information' in the ordinary sense is a richly semantic category: it is because of its meaning that something counts as information at all. 'Information' in the sense of information theory is a probabilistic measure on syntax alone, and the true topic of information theory is efficient syntactic coding.

Claude Shannon, the founder of information theory, was perfectly aware of this. For him (and for Bell Labs) the essential problem was an engineering problem of efficient transfer. Issues of meaning were irrelevant:

> Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. (Shannon, 1949, p. 31)

For our purposes, on the other hand, and for at least much of the philosophy of information, meaning will be not irrelevant but central. In asking what information is, what we want to know is the nature of information in a sense that demands meaning. We want to understand the dynamics of systems of meaningful information transfer, and want to be able to judge strategies for acquisition of meaningful information. It is information in this full-blooded semantic sense that we want to be able to filter, manage, and apply.

2. Modeling Meaningful Information

How is one to model meaningful information? How is one to model meaning? The models we want to introduce here fall within the broad tradition of theories of meaning as use.

A family of older approaches take meaning to be a relation. A sound or gesture is meaningful because it stands in a particular relation to something, on these approaches, and the thing to which it stands in the proper relation is taken to be its meaning. The crucial question for any relational theory of meaning, of course, is precisely what the meaning relation *is* and precisely what it is a relation *to*.

One classical theory casts the central relation as one of reference; the meaning relation is taken to be a relation of reference from words to things in the world. Words have meanings because they have referents, and the meaning of a word is the thing to which it refers. In various forms such a theory of meaning can be found in Augustine (c 400), Mill (1994) and Russell (1921, 1940). Although Shannon is not concerned with meaning, he echoes a referential theory in the quote above when he speaks of messages having meaning when they "refer to ... certain physical ... entities."

A second relational theory portrays meaning as a relation between a sound or gesture and the images, ideas, or internal representations it stimulates or is used to express. On such a view the meaning of the word is the thing in the head it is used to convey. Communication becomes an attempt to transfer the contents of my head into yours, or to make the contents of your head match mine. An ideational theory of this sort can be found in Aristotle (c 330 BC), Hobbes (1651) and Locke (1689), with a significantly more sophisticated contemporary echo in Fodor (1975). Shannon is appealing to an ideational theory of this sort when he speaks of messages having meaning when they "are correlated ... with certain ... conceptual entities."

A third approach is to consider meaning as a relation neither to things in the world nor to the contents of heads but to some abstract object, removed from the world and yet non-psychological. Here the classical representative is Frege (1879).

There does not appear to be any contemporary philosophical consensus on theories of meaning. Certainly there is no consensus in support of relational theories (see, for example, Ludlow, 1997). In some theories the entities to which the relation is supposed to stand themselves remain obscure–the representational entities of the ideational theory, for example, or the abstract entities of Frege's. In all relational theories the relation itself remains unclear, and none go so far as to indicate how the meaning relation is established. In ideational theories in particular, a gesture has meaning because it expresses or stimulates some event in a sender's or receiver's head. But what gives *that* event a meaning to begin with?

Although relational theories may be held in suspicion by many philosophers, it is our impression that they constitute the dominant picture of meaning across the various disciplines involved in contemporary modeling regarding communication and language. A referential theory, in which the meaning of a term is taken to be the object or situation it applies to, is more or less explicit in Batali (1995), Oliphant and Batali (1996), and MacLennan and Burghardt (1994). We take it to be a clear indication that the theory in play is a relational theory of the ideational variety when the measure of 'identity of meaning' or 'successful communication' is correspondence between individuals' representation maps or signal matrices. Ideational theories are clear in Levin (1995), Parisi (1997), and in the work of Martin Nowak (Nowak et al., 1999, 2000).

Relational theories are not the only option, however. Much current philosophical work, to which ours is not allied, relies on a Tarskiian theory of truth to do much of the work traditionally expected of a theory of meaning (Quine, 1960; Davidson, 1967, Larson and Segal, 1995). Of interest since the later Wittgenstein (1953) are also theories which emphasize not meaning as something that a word somehow *has* but communication as something that members of a community *do*. Wittgenstein is notoriously hard to interpret, but one clear theme is an insistence that meaning is to be understood not by looking for 'meanings' either in the world or in the head but by understanding the role of words and gestures in collaborative action within a community. Unfortunately, theories of meaning as use have often tended to go no farther than hints. It will not be enough to gesture vaguely at either meaning or information as socially constituted in interaction.

We think of the models we offer here as computer instantiations that carry a theory of meaning as use significantly farther. We think of them as use-based models for information as well: in particular, models of how it can be that something comes to carry communicative information, and how a system of information transfer can arise. The surprising results that appear in these models, we think, offer important support for an approach to both meaning and information in terms of use. We offer them as a step in the direction of greater explicitness and precision.

On the view we try to model here, a grasp of what meaning and information are will come not by looking for the right relations involving the right kind of object — meanings, for example — but by attention to the coordinated interaction of agents in a community. In practical terms, the measure of meaningful communication and information transfer will be functional coordination alone. The understanding we seek may thus come with an understanding of the development of patterns of functional communication within a community, but without our ever being able to identify a particular relation as the 'meaning' or 'information' relation, and without being able to identify a particular object — concrete, ideational, or abstract — as the thing that is the 'meaning' of a particular term or the item that is the 'information' conveyed.

An analogy may be helpful. We now have a wonderful biological grasp of the phenomenon of life, elegantly summarized for example in the 'replicators' of Dawkins (1976). Our contemporary understanding of life is very different, however, from an earlier picture in which life was thought to be some kind of component, quality, or even fluid that live bodies have and that dead bodies lack. This earlier picture appears in the Biblical tradition of a 'breath of life', for example; as recently as Mary Shelley's *Frankenstein* (1831), life is portrayed as something like a spark that must be added in order to animate dead tissue.

As long as we looked for life as a mysterious spark within an individual at a time, we were bound to misunderstand it. Only when we came to see life as a property of a class of organisms developing over time did we start to grasp its principles. We think that the same may be true of information and meaning. As long as we think of these as entities, they will remain mysterious. In order to understand meaning and information these too must be seen as properties of coordination within a community of individuals that develops over time. That is, at any rate, the conviction that drives the models we offer here.

In applying tools of formal modeling within a theory of meaning as use our most immediate philosophical precursors are Lewis (1969) and Skyrms (1996). Although the modeling literature for communication in other disciplines may be dominated by relational views of meaning, this dynamical approach also has its representatives: we note with satisfaction some comments in that direction in Hutchins and Hazelhurst (1995) and fairly explicit statements in Steels (1996, 1998).

Here we also want to add a proviso. What we outline below are models for meaning and information transfer that show surprising patterns of emergence and growth for meaning and simple information networks.[1] We don't want to suggest that these are the only possible models; that the patterns of emergence they exhibit are the only ways that meaning and information can appear, or that the forms of meaning and information they track are the only forms that meaning or information can take. In the end meaning and information may each be many things rather than one, and may require a range of overlapping theories. What we have to offer is thus not a total theory but an intriguing set of models, suggestive of dynamics possible for some of the central phenomena at issue.

3. A Use-Based Model of Meaning and Information

Given the motivation outlined in the preceding section, it should be clear that there will be no component in our model that is labelled 'the meaning' of a gesture or signal, nor any object transmitted that is conceived of as 'the information' passed. Despite the absence of identifiable 'meanings' or 'pieces of information', we think of ours as a model in which clear patterns of meaningful information transfer do emerge. Those patterns appears as a theory of meaning or information as use would lead one to expect: in the dynamics of behavioral coordination across a community.

The environment of our model is a cellular automata array: a 64×64 square grid in which each cell represents an individual. Each cell in the array touches eight others on its sides and corners; these are the individual's immediate neighbors. Technically, the array also forms a torus; ours is a 'wrap-around' array in which individuals on the bottom edge have neighbors at the top edge and those at the left edge have neighbors on the right (Figure 1).

The individual cells in the model have a range of simple behaviors — they can open their mouths, hide, or coast in neutral — but they do not move. What does move are food sources and predators, which follow a random walk across the array. When a food source lands on an individual with an open mouth, that individual 'feeds' and gains points. When a predator lands on an individual that is not hiding, that individual is 'hurt' and loses points. Both mouth opening and hiding, however, exact an energy cost. Energy costs can be avoided by an individual that coasts

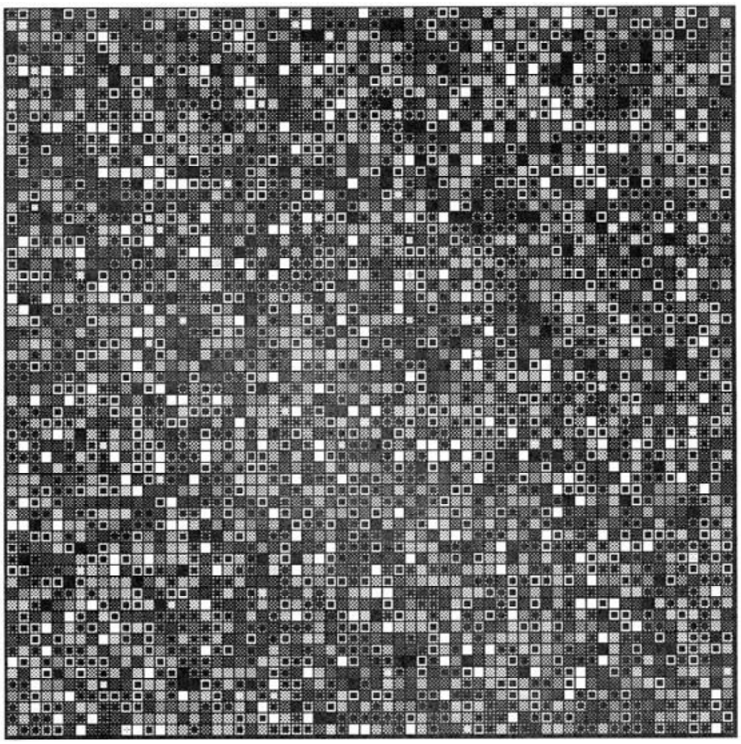

Figure 1. A randomized cellular automata array of strategies. Strategies are shown as different colors, open mouths as central black squares.

in neutral, but an individual in neutral can gain no points from feeding and is still vulnerable to predation because it is not hiding. It should also be noted that ours are food sources which are not consumed; like a school of fish, perhaps, they continue their random walk and may feed the next individual down the line. Predators are never sated, continuing their hunt in a random walk.

As part of each individual's behavioral repertoire, it can also make one of two arbitrary sounds, heard only by itself and its eight immediate neighbors. Each cell instantiates a behavioral strategy: in response to hearing such a sound, an individual might open its mouth, hide, do both or neither. In response to successful feeding, it might make a particular sound. But sound-making, like mouth-opening and hiding, exacts an energy cost.

Within these parameters, one can envisage particular strategies that cells might instantiate, and can imagine clusters or communities of individuals with corresponding strategies. We are particularly interested in communities of 'communicators', envisageable either as sharing a pattern of meaningful signaling or as instantiating a network of information transfer. A community of 'communicators' might make sound 1 when fed, for example, and open their mouths when sound 1 is heard from an immediate neighbor. Since food sources migrate from cell to cell,

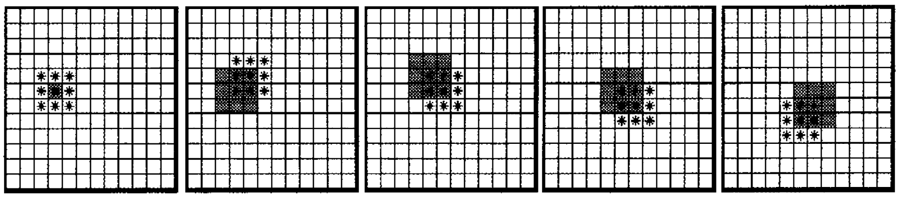

Figure 2. Migration of a single food source across a hypothetical array of communicators. In the left frame, a food source dot lands on an open mouth, indicated by gray shading. That central individual makes a sound * heard by its immediate neighbours, which in the second frame open their mouths in response. One of them feeds successfully, making a sound heard by its immediate neighbours, which are shown opening their mouths in the third frame. The result in a community of communicators is a chain reaction of efficient feeding.

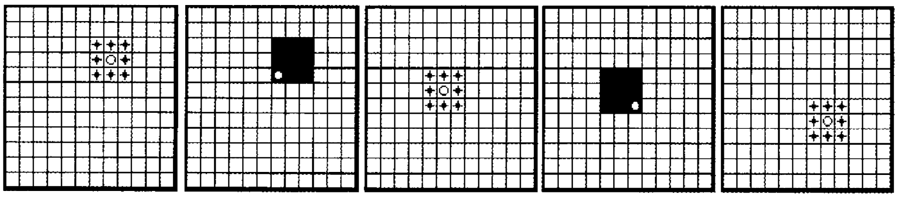

Figure 3. Migration of a single predator across a hypothetical array of communicators. In the left frame, a predator is shown as a hollow dot landing on a cell that is not hiding. The victim makes a sound shown as a cross, heard by itself and its immediate neighbours. The predator moves on, but the cell on which it lands in the second frame has avoided victimization by hiding, and so does not make a sound. The predator thus lands on an unwarned victim again in the third frame, and that victim sends out a warning cell. The result in a community of communicators is avoidance of predation only on every other round.

such a pattern of behavior instantiated across a community would increase chances of feeding (Figure 2) A community of 'communicators' might also make sound 2 when hurt, and hide when sound 2 is heard from an immediate neighbor (Figure 3).

The individuals that we embed in our array are simple neural networks, and their strategies are the behavior patterns dictated by those networks. The inputs to each net include whether the individual is fed on the current round, whether it is hurt, and any sounds heard from itself or immediate neighbors. The net's outputs dictate whether the individual opens its mouth on the next round, whether it hides, and whether it makes either of two sounds heard by immediate neighbors. 'Communicators' represent just one of a very large range of behavioral strategies.

Suppose we start with an array of neural nets with entirely randomized weights. Over a course of 100 rounds (a 'generation'), our individuals gain points from feeding and lose points from predation and energy expenditure. At that point we have each of our individuals see if any immediate neighbor — any of the eight touching cells — has gained a higher score. If so, we have them do a partial training on a sampling of the behavior of that neighbor that has amassed the most points. Changes in weights within individual nets follows standard learning algorithms, but learning is unsupervised in any global sense. There is no central supervisor or universal set of targets: learning proceeds purely locally, as individual nets do a partial training on the responses of those immediate neighbors that have proven most successful at a given time.

In such a context, might communities of ‘communicators’ emerge? It should be clear from the outline above that there is no reward built in for ‘communication’ or ‘information exchange’: individuals amass gain points only by successfully feeding and avoiding predation. With just individual gains at issue, could a network of neural nets in this simple environment nonetheless learn to communicate? Could an information network arise regarding food sources and predators?

The answer is ‘yes’. Within such a model, starting from an array of purely randomized neural nets, flourishing communities of communicators arise and spread. Members of such a community uniformly generate a particular sound at feeding and react to hearing that sound by opening their mouths. They generate a different sound when hurt and react to hearing that sound by hiding. In actually working with such a model it is almost impossible not to think of what emerges as a simple signaling system: a simple pattern of communication in which particular sounds carry a particular *meaning*. In actually working with such a model it is almost impossible not to think of sound-making and reaction to sound as a simple case of information transfer.

In our more sober moments, we remind ourselves that ours are merely computer models: that the blips on the screen operate in terms of an update mechanism across an array of randomized neural nets, and may not include any entity that literally means anything by anything. A fruitful mathematical model for gravity, however, need not itself function in terms of gravitational attraction, and a fruitful model for meaning and information need not literally have meaning or information inside. The models do seem to capture something important about the dynamics of meaning and information systems. From that perspective they may help us understand how real meaning and real information systems might arise and function.

There are individual features that our models share with some of their predecessors in the modeling literature. But there are also sharp contrasts with earlier models, and no previous model contains all the characteristics we take to be important.

In line with the relational models that tend to motivate them, most neural net models involving language to date have been models of idealized individuals. In the spirit of a theory of meaning as use, we on the contrary take communication to involve dynamic behavioral coordination across a community. We thus follow the general strategy of MacLennan and Burghardt (1994) and Hutchins and Hazelhurst (1995) in emphasizing the community rather than the individual in building a model of language development. Luc Steels’ (1998) outline of this shared perspective is particularly eloquent: Language may be

> a mass phenomenon actualised by the different agents interacting with each other. No single individual has a complete view of the language nor does anyone control the language. In this sense, language is like a cloud of birds which

attains and keeps its coherence based on individual rules enacted by each bird. (p. 2)

Another essential aspect of the model offered here is spatialization, carried over from our own previous work in both cooperation and simpler models for communication (Grim, 1995, 1996; Grim et al., 1998, 2000, 2001). Our community is modeled as a two-dimensional cellular automata array. Each individual interacts with its immediate neighbors, but no individual interacts with all members of the community as a whole. Our individuals are capable of making arbitrary sounds, but these are heard only by immediate neighbors. Communication and information transfer thus proceed purely locally, regarding food sources and predators that migrate through the local area. Fitness is measured purely locally, and learning proceeds purely locally as well: individuals do a partial training, using standard algorithms, on that immediate neighbor that has proven most successful. Spatialization of this thorough-going sort has not generally been exploited in earlier modeling of communication.[2]

It has been noted that the reward structure of the model is entirely individualistic. In many previous models both 'senders' and 'receivers' are simultaneously rewarded in each case of 'successful communication', rather than rewards tracking natural benefits that can be expected to accrue to the receiver alone.[3] The need for a model of how communication or information transfer might originate without symmetrical reward or shared tasks is noted explicitly by Ackley and Littman (1994), Noble and Cliff (1996), Parisi (1997), Cangelosi and Parisi (1998), Dyer (1995), and Batali (1995). Batali writes:

> While it is of clear benefit for the members of a population to be able to make use of information made available to others, it is not as obvious that any benefit accrues to the sender of informative signals. A good strategy, in fact, might be for an individual to exploit signals sent by others, but to send no informative signals itself. Thus there is a puzzle as to how coordinated systems of signal production and response could have evolved. (p. 2)

In an overview of various approaches, Parisi (1997) is still more explicit:

> In the food and danger simulations the organism acts only as a receiver of signals and it evolves an ability to respond appropriately to these signals. It is interesting to ask, however, where these signals come from. . . Why should the second individual bother to generate signals in the presence of the first individual? The evolutionary 'goal' of the first individual is quite clear. Individuals who respond to the signal 'food' ('danger') by approaching (avoiding) the object they currently perceive are more likely to reproduce than individuals who do not do so. Hence, the evolutionary emergence of an ability to understand these signals. . . But why should individuals who perceive food or danger objects in the presence of another individual develop a tendency to respond by emitting the signal 'food' or 'danger'? (p. 129)

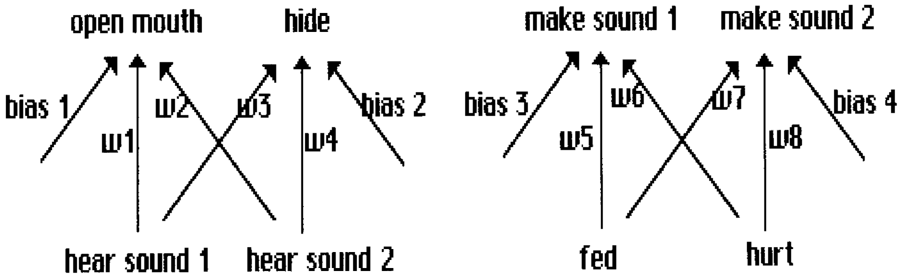

Figure 4. Perceptron architecture.

The models we offer here show the emergence of a system of communication or information transfer within a large community of neural nets using a structure of rewards that fits precisely the outline called for.

4. Learning to Communicate in an Array of Perceptrons

Here we begin with the simpler of two species of neural nets we have employed; their more complicated relatives are left to the following section. In the models offered here, our individuals employ neural nets with no hidden layers: the behavior of each individual is generated by a two-layer perceptron of the form shown in Figure 4.[4]

From Figure 4 it is clear that our neural architecture divides into two distinct halves: a right half that reacts to being fed or hurt by making sounds, and a left half that reacts to sounds by opening its mouth or hiding. No feed-forward connection goes from hearing sounds, for example, directly to making sounds. This 'two lobe' configuration for communication or information management seems to have been re-invented or re-discovered repeatedly in the history of the literature. Many note an intrinsic distinction between the kinds of action represented here by (1) making sounds and (2) mouth-opening or hiding in response to sounds heard.[5] It is also clear that such a structure builds in no presumption that a signal will be read in the same way that is sent: what de Saussure (1916) calls the 'bi-directionality' of signals. If bi-directionality nonetheless emerges, as indeed it does in our communities of 'communicators', it will be as a consequence of learning in the environment rather than any inherent structural constraint.

As indicated in Figure 4, our nets carry two weighted inputs to each output node, together with one weighted bias. In order to keep things as simple as possible in this first species of neural nets, we 'chunked' our weights: each of the twelve weights (corresponding to the twelve arrows) are 'chunked' at unit intervals between -3.5 and $+3.5$. For the network as a whole this gives us some 68 billion numerical combinations. Not every numerical difference in networks will make a behavioral difference. We are however working with 38,416 possible behavioral strategies: 38,416 different patterns of mouth opening, hiding, and sounding on different configurations of inputs.

The nets operate as follows. We use a bipolar coding for inputs, so that 'hear sound 1', for example, takes a value of +1 if the individual hears sound 1 from itself or an immediate neighbor on the previous round, and takes a value of −1 if it does not. Each input is multiplied by the weight shown on arrows from it, and the weighted inputs are then summed at the output node. To that is added the value (positive or negative) of the bias, which might alternatively be thought of as a third weight with a constant input of 1. If the total at the output node is greater than 0, we take our output to be +1, and the individual opens its mouth, for example; if the weighted total is less than 0, we take our output to be −1, and the individual keeps its mouth closed. Our own earlier work (Grim, et al., 2000), in line with results from Nowak and Sigmund (1990, 1992), showed the importance of a measure of stochastic imperfection in simulated environments: here in a random 5% of cases an individual will open its mouth, and in a random 5% it will hide, regardless of weights and inputs.

We code our behavioral strategies in terms of the outputs they give for possible pairs of inputs. The possible inputs at 'hear sound 1' and 'hear sound 2' for the left lobe of the net can be thought of as 00 (neither sound heard), 01 (just sound 2 heard), 10 (just sound 1 heard), or 11 (both sounds heard). Outputs for a given strategy will be pairs representing the output values for 'open mouth' and 'hide' for each of these pairs of inputs. We code the left-lobe behavior of a given strategy as a series of 8 binary digits. The string 00 00 00 11, for example, represents a behavior that outputs an open mouth or a hide only if both sounds are heard, and then outputs both. We can use a similar pattern of behavioral coding for the right lobe, and thus encode the entire behavior of a net in terms of 16 binary digits.

Of the 38,416 strategies in this sample space, there are precisely two that qualify as what we term 'perfect communicators'. Pattern 00011011 00011011 represents an individual that makes sound 2 whenever it is fed, and opens its mouth when it hears sound 2. It makes sound 1 whenever it is hurt, and hides whenever it hears sound 1. Pattern 00100111 00100111 represents an individual with a symmetrical behavior in which only the sound-correlations are changed: it reacts to sound 2 by eating and respond to being fed by making sound 2, reacts to sound 1 by hiding and responds to being hurt by making sound 1.

We initially populate our array with neural nets carrying 12 random weights, randomizing over our 68 billion numerical strategies. Individuals gain 1 point for successful feeding and lose 1 point for predation. We use 100 food sources and 200 predators, and set an energy cost for opening one's mouth or hiding at 0.05 points, with energy expenditure for making any sound at 0.05 points as well.[6] Over the course of 100 rounds, our individuals total their points and scan their immediate neighbors to see if any has garnered a higher score. If so, they do a partial training on the behavior of their highest-scoring neighbor.

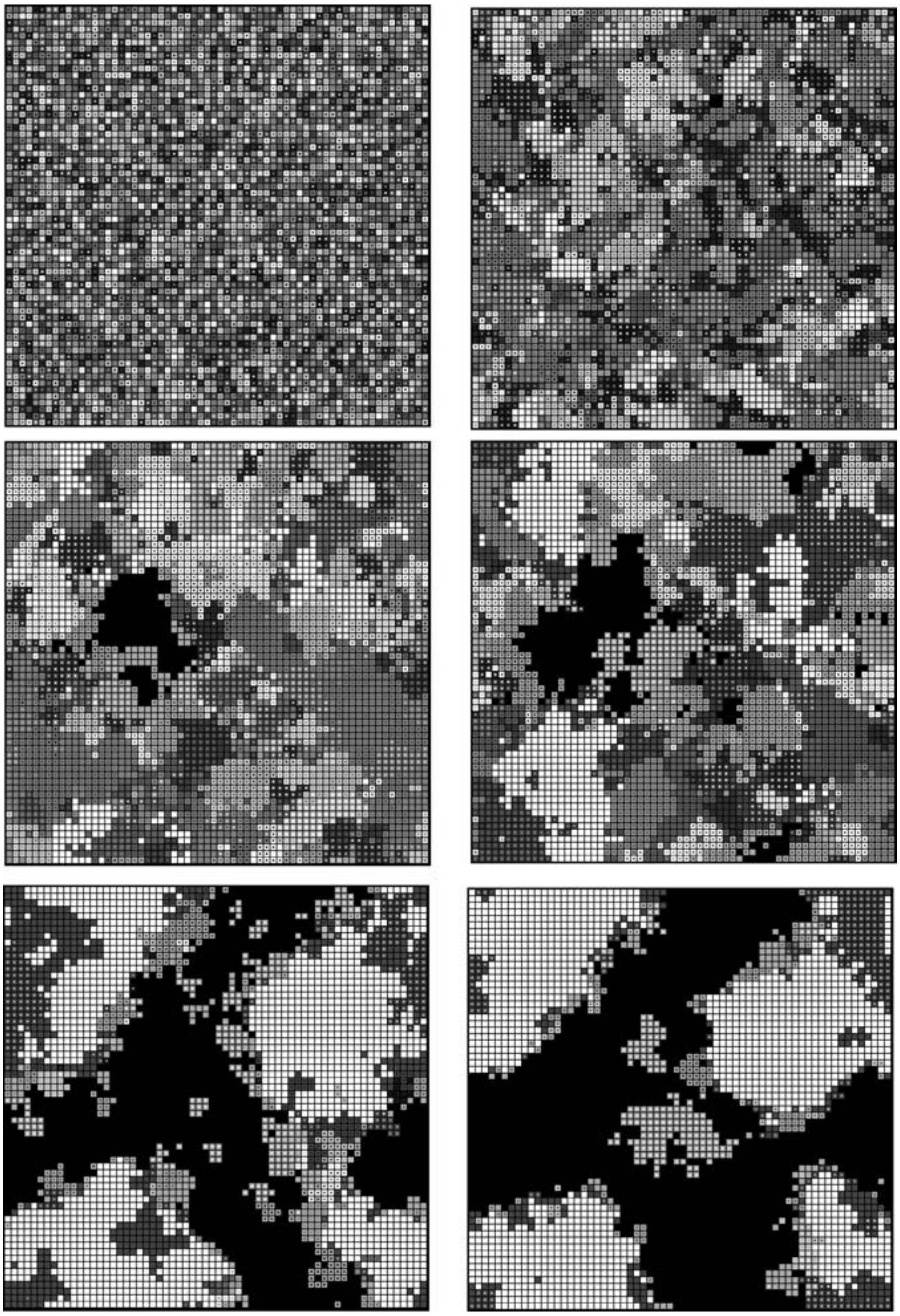

Figure 5. Emergence of two dialects of perfect communicators, shown in solid black and white, in a randomized array of perceptrons with partial training on successful neightbours. Generations 1, 10, 50, 100, 200 and 300 shown. The full development can be seen at http://129.49.17.140/im/im5.htm.

Our perceptrons are trained using the standard delta rule. For a set of four random inputs, the cell compares its outputs with those of its target neighbor. At any point at which the behavior of the training cell differs form that of its target, we nudge each of the responsible weights and biases one unit positively or negatively. Within the limits of our value scale, use of bipolar values for target and input allow us to calculate this simply, using $w_{new} = w_{old} + (\text{target} \times \text{input})$ and $\text{bias}_{new} = \text{bias}_{old} + \text{target}$.

Our training run consists of only four random sets of inputs, with no provision against duplication. If a cell has a neighbor with a higher score, in other words, it compares its behavior with that of its neighbor over four random sets of inputs, changing weights where there is a difference. Training will obviously be partial: only four sets of inputs are sampled, rather than the full 16 possible, and indeed the same set may be sampled repeatedly. The learning algorithm is applied using each set of inputs only once, moreover, leaving no guarantee that each weight will be shifted enough to make the behavioral difference of a complete training. The idea of partial training was quite deliberately built into our models in order to allow numerical combinations and behavioral strategies to emerge from training which might not previously have existed in either teacher or learner, thereby allowing a wider exploration of the sample space of possible strategies. In all of the runs illustrated below, for example, there are no ‘perfect communicator’ cells in our initial randomizations; those strategies are ‘discovered’ by the dynamics of partial training.

Suppose we start, then, with an array of 4096 perceptrons with randomized weights and biases. With a sample space of 38,416 strategies, will communities of our two perfect communicators emerge?

The answer is ‘yes’. Figure 5 shows the evolution of such a randomized array over the course of 300 generations. Here our ‘perfect communicators’ are coded in pure black and pure white: other strategies are coded using combinations of background colors and central dots. The full development of the array can be seen in action at http://129.49.17.140/im/im5.htm.

We can also graph the emergence of communication in terms of proportions of the population. Figure 6 shows the emergence, from an initial randomized array of neural nets, of precisely our two ‘perfect communicators’.

From arrays of neural nets randomized across possible weights, with partial training on successful neighbors in a spatialized array, communities of ‘perfect communicators’ routinely and robustly emerge and flourish.

Within an established community of communicators, the dynamics of sound and behavior clearly resemble patterns of meaningful communication or transfer of information; indeed it is difficult *not* to see an operating model as functioning in terms of meaning and information. In our more sober moments, as we’ve noted, we remind ourselves that these we are working with blips on a screen controlled by a computational algorithm. What these clearly seem to suggest, however, is a simple dynamic in which patterns of information transfer can arise and flourish.

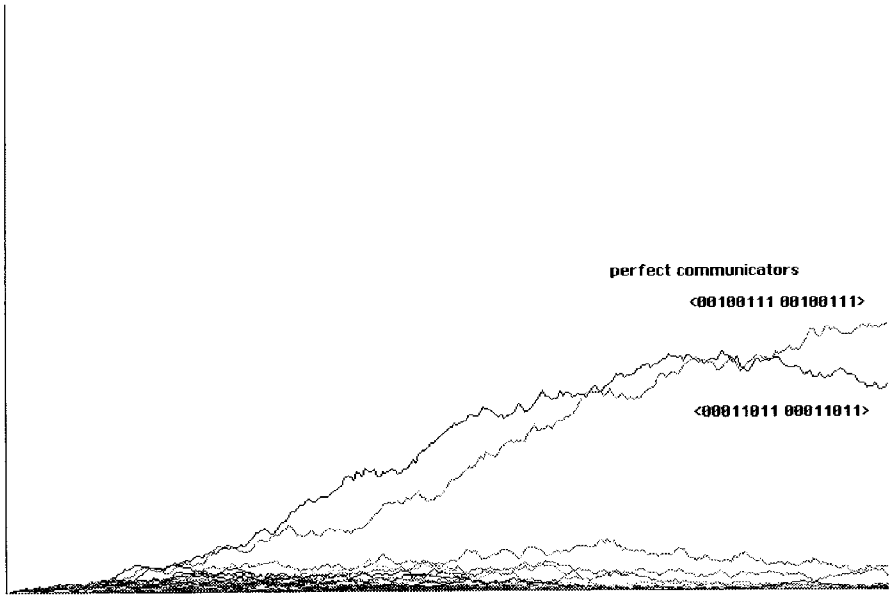

Figure 6. Emergence of communication in a randomized array of perceptrons with partial training on successful neighbors. Percentages of population graphed over 300 generations.

5. Emergent Information Systems in Backpropagation Nets

Here we repeat the model construction using more complex neural nets.

It has long been known that neural nets of just two layers, such as those employed above, are incapable of representing all of the Boolean functions: exclusive 'or' and the biconditional are exceptions. One of the behaviors that our nets above cannot instantiate, for example, is mouth opening just in case sound 1 is heard, or sound 2 is heard, but not both. This structural limitation dulls the impact of the otherwise remarkable perceptron learning convergence theorem: that the simple delta rule is adequate to train any perceptron, in a finite number of steps, to any function it can represent (Rosenblatt, 1959, 1962; Minsky and Papert, 1969, 1990; Fausett, 1994). Historically, this limitation posed a significant stumbling block to the further development of neural nets in the 1970s. It was known even then that the addition of intermediate layers to perceptrons would result in multi-layer neural nets which could model the full spectrum of Boolean functions, but the simple delta rule was known to be inadequate for training multi-layer nets.

With the use of continuous and differentiable activation functions, however, multi-layer neural nets can be trained by backpropagation of errors using a generalized delta function. This discovery signaled the re-emergence of active research on neural nets in the 1980s (McClelland and Rumelhart, 1988). Here again there is a convergence theorem: it can be shown that any continuous mapping can be

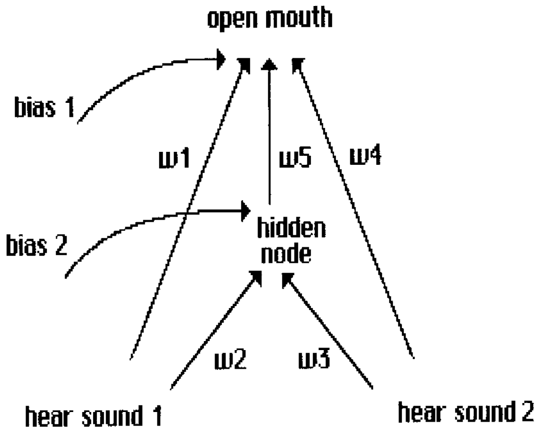

Figure 7. The quadrant structure of our backpropagation nets.

approximated to any arbitrary accuracy by using backpropagation on a net with some number of neurons in a single hidden layer (White, 1990; Fausett, 1994).

The most complicated neural nets we have to offer here exploit backpropagation techniques in order to train to the full range of Boolean functions of their inputs. Each of our nets is again divided into two 'lobes,' with inputs of two different sounds on the left side and outputs of mouth opening or hiding, inputs of 'fed' and 'hurt' on the right side with outputs of two different sounds made. Each of these lobes can be thought of as divided into two quadrants, as can our earlier nets, but here our quadrants are structured with so as to include a single hidden node (Figure 7).

The feedforward neural nets most commonly illustrated in the literature have hierarchically uniform levels — all inputs feed to a hidden layer, for example, and only the hidden layer feeds to output. For reasons of economy in the number of nodes and weights to be carried in memory over a large array of neural nets, the design of our nets is not hierarchically uniform. As is clear from Figure 7, inputs feed through weights w1 and w4 directly to the output node as well as through weights w2 and w3 to a hidden node. The output node receives signals both from inputs directly and through weight w5 from the hidden node.

At both the hidden node and the output node we use a sigmoid activation function $f(\chi) = \frac{2}{1+\exp(-\chi)} - 1$ equivalent to $\frac{1-\exp(-\chi)}{1-\exp(-\chi)}$, graphed in Figure 8.

In our sample quadrant, bipolar inputs -1 or +1 from 'hear sound 1' and 'hear sound 2' are first multiplied by weights w2 and w3, initially set between -3.5 and +3.5. At the hidden node, those products are added to a constant bias b2 set initially in the same range. The total is then treated as input to the activation function, generating an output somewhere between -1 and +1 that is sent down the line to the output node.

The signal from the hidden node is multiplied by weight w5, which is added at the output node to the product of the initial inputs times weights w1 and w4.

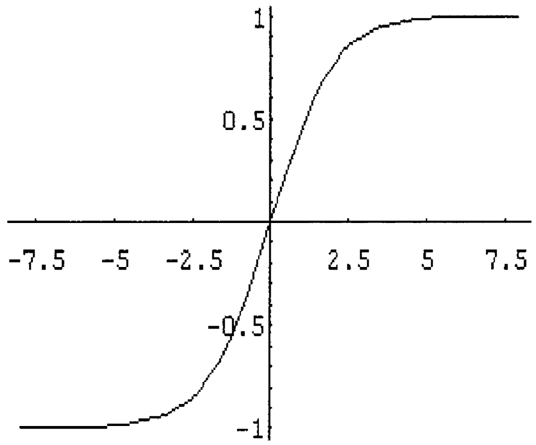

Figure 8. Activation function.

Bias b1 is also added to the sum. Here again all initial weights and biases are set between -3.5 and +3.5. This output is again passed through our activation function, with final output >0 treated as a signal to open the mouth, for example, and an output ≤ 0 as not opening the mouth. With different weight settings, this simple multi-layered structure is adequate to represent all 16 Boolean functions.

We employ a form of backpropagation appropriate to nets with this structure,[7] using the derivative of our activation function

$$f'(\chi) = \frac{[1 + f(\chi)][1 - f(\chi)]}{2}. \tag{1}$$

Training can be illustrated in terms of a single quadrant.

For a particular pair of inputs, we will at training have a particular target *t*: the output (-1 or +1) toward which our net is to be trained for those inputs. We operate our net feedforward, as outlined above, to obtain a final output *o* of -1 or +1. We calculate an output error information term δ_o = (*t*-*o*).

δ_o is applied directly to calculate changes in weights w1 and w4 on lines feeding straight from inputs. In each case the weight change Δ is a learning rate *lr* times δ_o times the input signal that was fed down that line. Our learning rate is set at a constant 0.02 throughout.

$$\Delta_w 1 = lr \times \delta_o \times \text{input(sound1)}$$
$$\Delta_w 4 = lr \times \delta_o \times \text{input(sound2)}$$

A bias can be thought of as a weight with a constant input of 1, and the change for bias b1 is calculated in the same way:

$$\Delta b1 = lr \times \delta_o. \tag{2}$$

The weight change for w5, from hidden node to output, follows a similar pattern, though Δw5 is calculated in terms of the signal which was sent down the line from hidden to output node in the feedforward operation of the net:

$$\Delta w5 = lr \times \delta_o \times \text{output(h)}. \tag{3}$$

Weight changes for w2 and w3 are calculated by backpropagation. Here we first calculate a new error information term δ_h = w5 $\times$ δ_o $\times$ $f'(\text{inp}_h)$, where $f'(\text{inp}_h)$ is the derivative of our activation function applied to the sum of weighted inputs at our hidden node. Changes in weights w2 and w3 are then calculated in terms of δ_h and our initial inputs:

$$\Delta_w 2 = lr \times \delta_h \times \text{input(sound1)}$$
$$\Delta_w 3 = lr \times \delta_h \times \text{input(sound2)}$$

The change in bias b2 will be simply $lr \times \delta_h$.

Once all weight and bias changes are calculated, they are simultaneously put into play:

$$\begin{aligned} w1 &= w1 + \Delta w1 \\ w2 &= w2 + \Delta w2 \\ w3 &= w3 + \Delta w3 \\ w4 &= w4 + \Delta w4 \\ w5 &= w5 + \Delta w5 \\ b1 &= b1 + \Delta b1 \\ b2 &= b2 + \Delta b2. \end{aligned}$$

For the sake of simplicity we have outlined the basic structure of our nets and our training algorithm above in terms of an isolated quadrant. Our nets as a whole are four times as complicated, of course, with two lobes of two quadrants each (Figure 9).

Each of our more complex neural nets employs a total of 20 weights, plus eight biases, requiring a total of 28 variable specifications for each net at a given time. In the networks of the previous section, we used discrete values for our weights: weights could take values only at unit intervals between -3.5 and +3.5. For the simple delta learning rule used there, this was a useful simplification. Backpropagation, however, demands a continuous and differentiable activation function, and will not work properly with these 'chunked' approximations. Here, therefore, our individual nets are specified at any time in terms of 28 real values in the range between -3.5 and +3.5.

Had we used 'chunked' weights at 0.5 intervals, we would have been dealing with 7^{28} different numerical strategies. With real values for weights, the ceiling

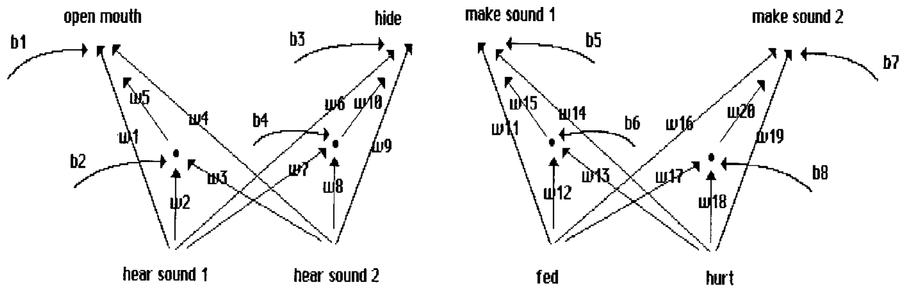

Figure 9. The full architecture of our backpropagation nets.

is lifted and we are dealing at least theoretically with a nondenumerably infinite number of numerical strategies.[88] But of course not every difference in weights makes a difference in output behavior. Each quadrant is capable of 16 different output patterns for a complete set of possible inputs. Our sample space is therefore one of 65 536 distinct behavioral strategies.

Here as before we can code our behavioral strategies in terms of binary strings. Pairs of digits such as 01 represent a lobe's output for a single pair of inputs. A coding 00 01 01 11 can thus be used to represent output over all possible pairs of inputs to a lobe: (-1,-1), (-1, +1), (+1, -1), and (+1,+1). A double set 01111000 00100011 serves to represent the behavior of both lobes in a network as a whole.

Of the 65,536 distinct behavioral strategies that can thus be encoded, there are still only two that qualify as 'perfect communicators'. The pattern 00011011 00011011 represents an individual that makes sound 1 whenever it is fed and reacts to sound 1 by opening its mouth, that makes sound 2 whenever it is hurt and reacts to sound 2 by hiding. It will both hide and open its mouth when it hears both sounds and will make both sounds when both hurt and fed. Pattern 00100111 00100111 represents an individual with a symmetrical behavior in which only the sound correlations are changed. This second individual makes sound 2 when it is fed and reacts to sound 2 by opening its mouth, makes sound 1 when hurt and reacts to sound 1 by hiding.

There are also variants on the pattern of perfect communicators that differ by a single digit in their encoding. Those that play the most significant role in our runs are 'right-hand variants', which differ from one or the other of our perfect communicators in just one of the last two digits, applicable only on those rare occasions when an individual is both fed and hurt at the same time. Patterns 00011011 00011010 and 00011011 00011001 differ from a perfect communicator in that they each make just one sound rather than two in the case that they are simultaneously fed and hurt. Patterns 00100111 00100110 and 00100111 00100101 vary from our other perfect communicator in the same way. For our two 'perfect communicators' there are thus also four minimally distinct 'right-hand variants' out of our 65,000 behavioral strategies.

We initially randomize all 28 weights as real values between -3.5 and $+3.5$ for each of the 4096 neural nets embedded in our 64×64 array. 100 food sources and 200 predators wander across the array as before. When a cell has a predator on it and is not hiding, it is ‘hurt’ and loses 1 point; when it has food on it and has its mouth open, it is ‘fed’ and gains 1 point. Each of these activities carry an energy expenditure of 0.05 points. In ‘neutral’ an individual is neither hiding nor has its mouth open, which saves it from energy expenditure but makes it incapable of capturing food and leaves it vulnerable to victimization by a predator. When hurt or fed, an individual makes one sound, both, or neither, where making a sound also carries an energy expenditure of 0.05 points. Here as before, it should be noted, ours is an ‘imperfect’ world: individuals open their mouths and hide in a random 5% of all cases regardless of inputs and internal structure.

Over the course of 100 rounds (a ‘generation’), our individuals collect points from successful feeding and lose points as victims of predation. At the end of each generation they scan their eight immediate neighbors to see if any has garnered more points. If so, they do a partial training on the behavior of the highest-scoring neighbor. All this is as it was before; what differs is the structure of the nets themselves, the full sample space of behavioral strategies, and training by the backpropagation algorithm outlined above.

Full training by backpropagation standardly requires a large number of epochs, each consisting of the complete training set in a randomized order. Here, however, we use only a single training epoch. At the end of each 100 rounds, should any neighbor have a higher score, individuals do a partial training on the behavior of their highest-scoring neighbor. Training uses a complete but randomized set of possible inputs for each quadrant and takes the more successful neighbor’s behavioral output for each pair of inputs as target. This cannot, of course, be expected to be a full training in the sense that would make behaviors match; training using a single epoch will typically shift weights only to some degree in a direction that accords with the successful neighbor’s behavior. Often the resulting behavior will match neither the initial behavior of the ‘trainee’ nor the full behavior of its more successful neighbor.

Figure 10 shows a typical result with 1 epoch of training over the course of 300 generations. Rather than plotting all 65,000 behavioral strategies, we have simplified the graph by showing only those strategies which at one point or another appeared among the top 20 in the array. Here the two strategies that emerge from a sample space of 65,000 are our two ‘perfect communicators.’ Starting from a randomized configuration it is also possible, however, for one or another ‘right-hand variant’ to play a significant role as well.

Despite the greater sophistication of the backpropagation model, and despite the wider sample space of possible behaviors, the central result is the same. Starting from an array randomized across possible behaviors, and with a partial training on the behavior of successful neighbors, communities of communicators instantiating networks of information transfer seem spontaneously to emerge and flourish.

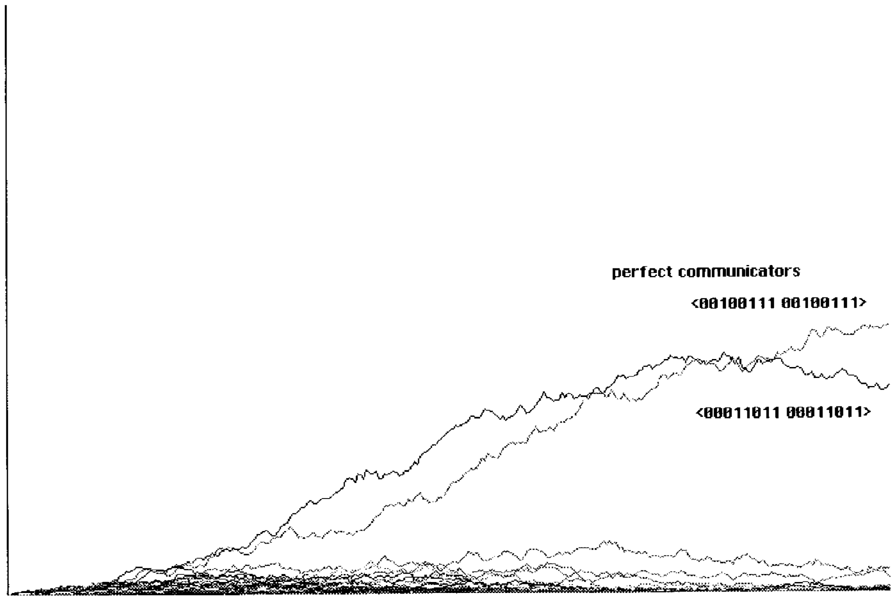

Figure 10. Emergence of perfect communication using backpropagation in an array of randomized neural nets. One training epoch used, 300 centuries shown.

6. Conclusion

In the sense that is of interest for philosophy of information, information is intimately linked to meaning. To understand what information is, how it operates, and how to put it to work, we will have to understand meaning as well.

Here we have tried to introduce some simple use-based models for meaning and information transfer, expanding the tradition of theories of meaning as use with the techniques of computer modeling. The magic in these models is that something very much like a pattern of meaning in a community, and something very much like a network of information transfer, emerges systematically from localized training on successful neighbors in a spatialized network of initially randomized neural nets.

In the alternative philosophical tradition of ideational relational theories of meaning (Aristotle c. 330 BC; Hobbes, 1651; Locke, 1689; Fodor, 1975), and in much current modeling work in other disciplines (Levin, 1995; Hutchins and Hazlehurst, 1995; Parisi, 1997; Nowak et al., 1999, 2000), the 'meaning' of a sound or gesture is taken to be a correspondence between sound and some internal representation. For the patterns of meaning and information transfer that seem to appear in the models offered here, however, ideational theories are significantly less plausible than a theory written directly in terms of patterns of use. In the models sketched here, learning proceeds throughout in terms of weight-shifting toward the behavior of a successful neighbor. When a community of communicators emerges from an array of randomized neural nets, it is not internal matches but convergence to a behavioral strategy that is crucial.

In the models offered above there is in fact no guarantee that the internal workings of behaviorally identical strategies — primary candidates for internal 'representations' — are themselves identical. There are in principle non-denumerably many neural configurations which may show the same behavioral strategy. In training to match a neighboring 'perfect communicator', a neural net may not only fail to match the absolute values of its neighbor's weights, but may not even match its over-all structure of relative weight balances. What arises in a community is a pattern of coordinated behavior, but in evolving from an initially randomized array of neural nets that coordinated behavior need not be built on any uniform understructure in the nets themselves. There is thus no guarantee of matching internal representations in any clear sense, no guarantee of matching internal 'meanings' or internal 'pieces of information' that are transferred, and no need for internal matches in the origin and maintenance of patterns of communication or information across a community.

The basic philosophical lesson is a Wittgensteinian one, here given a more formal modeling instantiation. To understand information we will have to understand meaning; indeed it appears that an understanding of information transfer and of meaningful communication can only be achieved together. For at least some kinds of information and at least some kinds of meaning, we want to suggest, a primary route to that understanding will be attention to the behavioral dynamics of patterns of use in a community over time.

Notes

[1] 'Emergence' has acquired a fairly standard and fruitful use despite lacking any adequate technical definition. What we have in mind is what Cariani calls 'computational emergence': "Complex global structures or behaviors arise from local computational interactions" (Cariani, 1992). In Chalmers' terms, "complex, interesting high-level function is produced as a result of combining simple low-level mechanisms in simple ways," or more colloquially, "something stupid buys you something smart" (Chalmers, 1990).

[2] See also Grim et al. (2003). Aside from our own previous work, Ackley and Littman (1994) is perhaps the most consistently spatialized model to date, with local communication and reproduction limited at least to breeding those individuals in a 'quad' with the highest fitness rating. Theirs is also a model complicated with a blizzard of further interacting factors, however, including reproductive 'festivals' and a peculiar wind-driven strategy diffusion.

[3] An assumption of mutal benefit from communicative exchanges is explicitly made in the early theoretical work of Lewis (1969). MacLennan (1991) offers a model in which both 'senders' and 'receivers' are both rewarded, with communicative strategies then perfected through the application of a genetic algorithm. As Ackley and Littman (1994) note, the result is an artificial environment "where 'truthful speech' by a speaker and 'right action' by a listener cause food to rain down on both " (p. 40). In some studies the target has been explicitly limited to 'communication for cooperation' (MacLennan and Burghardt, 1994; Levin, 1995; Wagner, 2000). That restriction, of course, also limits the generalizability of such models to communication or information transfer in general. Within

neural net models in particular, symmetrical rewards for 'successful communication' characterize Werner and Dyer (1991), Saunders and Pollack (1996), and Hutchins and Hazelhurst (1995).

[4]In earlier studies we have considered similar environments in which strategy change was by pure imitation of successful neighbours (Grim et al., 2000) or by genetic algorithm crossover with successful neighbours (Grim et al., 2001). In these studies we found similar results, indicating that the forces driving emergence of information transfer are inherent in the environmental structure rather than in the specific mode of strategy change. Here we focus on learning in arrays of neural nets as the mode of change; greater detail can be found in Grim et al. (2002).

[5]MacLennan (1991) similarly distinguishes 'emissions' from 'actions', for example, and Oliphant and Batali (1997) distinguish 'transmission behavior' from 'reception behavior.' It also seems natural to embody that distinction in the neural architecture of the individuals modeled: Werner and Dyer (1991) separate precisely these two functions between two different sexes, Cangelosi and Parisi (1998) note that the architecture of their neural nets uses two separate sets of connection weights for the two kinds of action, and Martin Nowak notes that his active matrix for signal-sending and his passive matrix for signal-sending can be treated as completely independant (Nowak et al., 1999, 2000).

[6]The reason for using twice as many predators lies in the dynamics shown in Figures 2 and 3 above. In an array composed entirely of 'communicators', a chain reaction can be expected in terms of food.

[7]We are deeply indebted to Laurene Fausett for helpful correspondence regarding training algorithms for nets of the structure used here. Our simple net combines perceptron-like connections (along weights w1 and w4) with crucial use of a single hidden node; it will be noted that the training algorithm also combines a perceptron-like training for w1, w4, and w5 with full backpropagation to update w2 and w3.

[8]'At least theoretically' because of course computer instantiation does not deal with true reals.

References

Ackley, D. and Littman, M. (1994), 'Altruism in the Evolution of Communication', in R.A. Brooks and P. Maes, eds., *Artificial Life IV: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems*, Cambridge, MA: MIT Press, pp. 40–48.

Aristotle (c 330 BC), *De Interpretatione*, in R. McKeon, ed., *Basic Works of Aristotle*, New York: Random House, 1971.

Augustine (c 400), *Confessions*, in J.J. O'Donnell, ed., Oxford: Oxford University Press, 1992.

Batali, J. (1995), 'Small Signaling Systems Can Evolve in the Absence of Benefit to the Information Sender', unpublished.

Cangelosi, A., and Parisi, D. (1998), 'The Emergence of a 'Language' in an Evolving Population of Neural Networks', *Connection Science* 10, pp. 83–97.

Cariani, P. (1991), 'Emergence and Artificial Life', in C. Langton, C. Taylor, J. D. Farmer and S. Rasmussen, eds., *Artificial Life II, SFI Studies in the Sciences of Complexity, vol. X*, Reading, MA: Addison-Wesley, pp. 775–797.

Chalmers, D., 1990, 'Thoughts on Emergence', http://www.u.arizona.edu/∼chalmers/notes/emergence.html.

Davidson, D. (1967), 'Truth and Meaning', *Synthese* 17, pp. 304–323.

Dawkins, R. (1976), *The Selfish Gene*, Oxford: Oxford University Press.

de Saussure, F. (1916), *Cours de Linguistique Generale*, Trans. as *Course in General Linguistics*, R. Harris, trans., Duckworth, 1983.

Fausett, L. (1994), *Fundamentals of Neural Networks*, Prentice-Hall.

Fodor, J. (1975), *The Language of Thought*, New York: Crowell.

Frege, G. (1879), *Begriffsschrift*, in Jean van Heijenoort, ed., *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*, Harvard University Press, 1967.

Grim, P. (1995), ‘Greater Generosity in the Spatialized Prisoner’s Dilemma’, *Journal of Theoretical Biology* 173, pp. 353–359.
Grim, P. (1996), ‘Spatialization and Greater Generosity in the Stochastic Prisoner’s Dilemma’, *Biosystems* 37, pp. 3–17.
Grim, P., Mar, G. and St. Denis, P. (1998), *The Philosophical Computer: Exploratory Essays in Philosophical Computer Modeling*, MIT Press/Bradford Books.
Grim, P., Kokalis, T., Tafti, A., and Kilb, N. (2000), ‘Evolution of Communication in Perfect and Imperfect Worlds’, *World Futures: The Journal of General Evolution* 56, pp. 179–197.
Grim, P., Kokalis, T., and Kilb, N. (2001), ‘Evolution of Communication with a Spatialized Genetic Algorithm’, *Evolution of Communication* 3, pp. 105–134.
Grim, P., St. Denis, P. and Kokalis, T. (2002), ‘Learning to Communicate: The Emergence of Signaling in Spatialized Arrays of Neural Nets’, *Adaptive Behavior* 10, pp. 45–70.
Grim, P., Wardach, S. and Beltrani, V. (2003), ‘Location, Location, Location: The Importance of Spatialization in Modeling Cooperation and Communication’, Research Report #03-01, Group for Logic & Formal Semantics, Dept. of Philosophy, SUNY at Stony Brook.
Hobbes, T. (1651), *Leviathan*, in R.E. Flathman and D. Johnston, eds., Norton Critical Edition, 1997.
Hutchins, E. and Hazlehurst, B. (1991), ‘Learning in the Cultural Process’, in C. G. Langton, C. Taylor, J. D. Farmer and S. Rasmussen, eds., *Artificial Life II, SFI Studies in the Sciences of Complexity, vol. X*, Reading, MA: Addison-Wesley, pp. 689–708.
Hutchins, E. and Hazlehurst, B. (1995), ‘How To Invent a Lexicon: The Development of Shared Symbols in Interaction’, in N. Gilbert and R. Conte, eds., *Artificial Socities: The Computer Simulation of Social Life*, UCL Press, pp. 157–189.
Larson, R. and Segal, G. (1995), *Knowledge of Meaning: An Introduction to Semantic Theory*, Cambridge, MA: MIT Press/Bradford Books.
Levin, M. (1995), ‘The Evolution of Understanding: A Genetic Algorithm Model of the Evolution of Communication’, *BioSystems* 36, pp. 167–178.
Lewis, D. (1969), *Convention: A Philosophical Study*, Harvard University Press.
Locke, J. (1689), *Essay Concerning Human Understanding*, in Peter H. Nidditch, ed., Oxford University Press, 1979.
Ludlow, P., ed. (1997), *Readings in the Philosophy of Language*, MIT Press/Bradford Books.
MacLennan, B.J. (1991), ‘Synthetic Ethology: An Approach to the Study of Communication’, in C. G. Langton, C. Taylor, J. D. Farmer and S. Rasmussen, eds., *Artificial Life III, SFI Studies in the Sciences of Complexity, vol. X*, Reading, MA: Addison-Wesley, pp. 631–655.
MacLennan, B.J. and Burghardt, G.M. (1994), ‘Synthetic Ethology and the Evolution of Cooperative Communication’, *Adaptive Behavior* 2, pp. 161–188.
McClelland, J. L. and Rumelhart, D.E. (1988), *Explorations in Parallel Distributed Processing*, Cambridge, MA: MIT Press.
Mill, J.S. (1884), *A System of Logic, Ratiocinative and Inductive, Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*, London: Longmans, Green, and Co.
Minsky, M. and Papert, S. (1969, 1990), *Perceptrons: An Introduction to Computational Geometry*, Cambridge, MA: MIT Press.
Noble, J. and Cliff, D. (1996), ‘On Simulating the Evolution of Communication’, in P. Maes, M. Mataric, J. Meyer, J. Pollack and S. Wilson, eds., *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*, MIT Press, pp. 608–617.
Nowak, M.A., Plotkin, J.B. and Krakauer, D.C. (1999), ‘The Evolutionary Language Game’, *Journal of Theoretical Biology* 200, pp. 147–162.
Nowak, M.A., Plotkin, J.B. and Jansen, V.A. (2000), ‘The Evolution of Syntactic Communication’, *Nature* 404, pp. 495–498.
Nowak, M.A. and Sigmund, K. (1990), ‘The Evolution of Stochastic Strategies in the Prisoner’s Dilemma’, *Acta Applicandae Mathematicae* 20, pp. 247–265.

Nowak, M.A. and Sigmund, K. (1992) 'Tit for Tat in Heterogeneous Populations', *Nature* 355, pp. 250–252.

Oliphant, M. and Batali, J. (1997) 'Learning and the Emergence of Coordinated Communication', *Center for Research on Language Newsletter* 11(1).

Parisi, D. (1997), 'An Artificial Life Approach to Language', *Brain and Language* 59, pp. 121–146.

Quine, W.V.O. (1960), *Word and Object*, Cambridge, MA: MIT Press

Rosenblatt, F. (1962), *Principles of Neurodynamics*, New York: Spartan Press.

Russell, B. (1921), *The Analysis of Mind*, London: George Allen & Unwin.

Russell, B. (1940), *An Inquiry into Meaning and Truth*, London: George Allen & Unwin.

Saunders, G.M. and Pollack, J.B. (1996), 'The Evolution of Communication Schemes Over Continuous Channels', in P. Maes, M. Mataric, J. Meyer, J. Pollack and S. Wilson, eds., *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*, MIT Press, pp. 580–589.

Shannon, C. (1949), *The Mathematical Theory of Communication*, University of Illinois Press.

Shelley, M. (1831), *Frankenstein*, Everyman's Library, 1922.

Skyrms, B. (1996), *Evolution of the Social Contract*, Cambridge University Press.

Steels, Luc (1996), 'Emergent Adaptive Lexicons', in P. Maes, M. Mataric, J. Meyer, J. Pollack, and S. Wilson, eds., *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*, MIT Press, pp. 562–567.

Steels, L. (1998), 'Synthesizing the Origins of Language and Meaning Using Co-Evolution, Self-Organization and Level Formation', in J. R. Hurford, M. Studdert-Kennedy and C. Knight, eds., *Approaches to the Evolution of Language: Social and Cognitive Bases*, Cambridge University Press, pp. 384–404.

Wagner, Kyle (2000), 'Cooperative Strategies and the Evolution of Communication', *Artificial Life* 6, pp. 149–179.

Werner, G. and Dyer, M. (1991), 'Evolution of Communication in Artificial Organisms', in C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen, eds., *Artificial Life II, SFI Studies in the Sciences of Complexity, vol. X*, Addison-Wesley, pp. 659–687.

White, H. (1990), 'Connectionist Nonparametric Regression: Multilayer Feedforward Networks Can Learn Arbitrary Mappings', *Neural Networks* 5, pp. 535–549.

Wittgenstein, L. (1953), *Philosophical Investigations*, Macmillan.