Making Meaning in Computers:
Synthetic Ethology Revisited

Technical Report UT-CS-05-549

Bruce J. MacLennan[*]

Department of Computer Science
University of Tennessee, Knoxville
www.cs.utk.edu/~mclennan

May 5, 2005
Revised November 1, 2006

Abstract

This report describes *synthetic ethology*, a scientific methodology in which we construct synthetic worlds in which synthetic agents evolve and become coupled to their environment. First we review the motivations for synthetic ethology as an experimental methodology and explain how it can be used to investigate intentionality and meaning, and the mechanisms from which they emerge, with an especial emphasis on communication and language. Second, we present several examples of such experiments, in which genuine (i.e., not simulated) meaningful communication evolved in a population of simple agents. Finally we discuss the extension of the synthetic ethology paradigm to the problems of structured communications and mental states, complex environments, and embodied intelligence, and suggest one way in which this extension could be accomplished. Indeed, synthetic ethology offers a new tool in a comprehensive research program investigating the neuro-evolutionary basis of cognitive processes.

Keywords: artificial language, artificial life, autonomous robot, categorization, cellular automata, communication, cooperation, ecological validity, embodiment, ethology, evolution, intentionality, learning, meaning, micro-world, neural networks, pragmatics, semiotics, simulation, situatedness, symbol, symbol grounding, synthetic ethology

[*] This report is an unedited preprint of a chapter appearing in *Artificial Cognition Systems*, ed. by Angelo Loula, Ricardo Gudwin and Joâo Queiroz, Idea Group, 2006, ch. 9 (pp. 252–83).

1 Introduction

Synthetic ethology was developed as a methodology for constructing experiments in which artificial agents could exhibit real (i.e., not simulated) intentionality and other mental phenomena. Our first experiments using this methodology demonstrated the evolution of communication in a population of simple machines and illustrated ways of relating the emergence of meaning to underlying mechanisms (MacLennan, 1990, 1992, 2001, 2002; MacLennan and Burghardt, 1993). In these experiments, as I will explain, the communications were meaningful to the artificial agents themselves, but they were only secondarily and partly meaningful to the experimenters.

This chapter has two purposes. The first is to review the motivations for synthetic ethology as an experimental methodology and to explain how it can be used to investigate intentionality and meaning, and the mechanisms from which they emerge, with an especial emphasis on communication and language. The second purpose is to reconsider these issues with the hindsight of fifteen years, and discuss new approaches to the use of synthetic worlds in the scientific investigation of problems in epistemology and cognitive science.

2 Background

2.1 Definition of Synthetic Ethology

Synthetic ethology can be defined as an experimental methodology in which the mechanisms underlying cognitive and intentional phenomena are investigated by constructing synthetic agents and observing them in their *environment of evolutionary adaptedness* (EEA, the environment in which they have evolved), which is also synthetic. These synthetic worlds are commonly constructed inside a computer. I will briefly summarize the most important considerations motivating the synthetic ethology paradigm (a fuller discussion can be found in MacLennan, 1992).

In discussing his research program in *synthetic psychology*, which was a direct inspiration for synthetic ethology, Braitenberg (1984, p. 20) distinguished "uphill analysis and downhill invention." By this he meant to distinguish the enormous difficulty of analyzing natural systems, as opposed to the comparative simplicity of synthesizing systems exhibiting a behavior of interest. His intention was to advocate the synthesis of neural networks and robots exhibiting intelligent behavior as an important adjunct to the analysis of intelligent agents in nature. Synthetic ethology extends this approach to phenomena for which populations and their evolution are relevant, such as communication.

The synthetic approach is especially valuable for investigating phenomena that depend essentially on the evolutionary history of the agents. Our ability to test evolutionary hypotheses about natural species is limited; we cannot go back into the past and restart the evolution of a species with different initial or boundary conditions, but we can do this with synthetic populations. That is, in synthetic ethology we can make systematic investigations of the effects of various parameters on the evolutionary outcome. Synthetic-ethology experiments are also facilitated by the rapid pace of evolution in synthetic worlds.

The ability to rerun evolution is just one example of the greater experimental control afforded by synthetic ethology over ordinary ethology. Because synthetic ethology constructs the world in which the experiments take place, every variable is under control, and we can intervene in the experiment whenever it is advantageous to do so. Some examples of useful control include the abilities to determine the genotypes in the population, to allow genetically identical initial populations to evolve under different conditions, and to inspect, control, or alter the behavioral mechanism (e.g. neural network) of an agent. Furthermore, since the entire synthetic world is contained in the computer, any mechanism underlying intentional or meaningful behavior is potentially open for inspection. This characteristic is critical, because it allows connecting the behavioral mechanisms (corresponding in natural organisms to neuron-level structures and processes) to the social-evolutionary level (that is, the evolution of a population over many generations). When meaningful behavior is observed in the population, there need be no "ghost in the machine"; the underlying mechanism is completely accessible.

2.2 Intrinsic Intentionality and Meaning

Even in a philosophical context, *intentionality* has several (interrelated) meanings. For instance, Searle (1983, p. 1) says, "Intentionality is that property of many mental states and events by which they are directed at or about or of objects and states of affairs in the world." For my purposes intentionality may be defined informally as the property of a physical state or process when it is *about* something else (e.g., Blackburn, 1994, p. 196; Gregory, 1987, p. 383; Gutenplan, 1994, p. 379). For example, states in our brains may instantiate propositional attitudes about real or imaginary objects, such as beliefs, doubts, desires, hopes, fears, memories, anticipations, and so forth. However, our linguistic expressions are also generally about something (their semantics), and therefore potentially meaningful to us (if we understand the expression). Thus we are led to another ambiguous term: "meaning." One of its senses, and the one that I will use in this paper, is to refer to the intentional aspect of signals, linguistic expressions, and similar meaning-bearing physical states and processes (e.g., Gutenplan, 1994, p. 386; Searle, 1983, pp. 26–9). Thus we may refer to the *meanings* of a vervet's alarm call, a peacock's mating display, or a human utterance or gesture.

Many physical states and processes are meaningful, and therefore intentional, in this broad sense. For example, the states of a computer memory are almost always *about* something — for example, a student's academic record is *about* that student — but no one claims that the information is meaningful to the computer in the same sense that it is meaningful to us. Therefore philosophers (e.g., Dennett, 1987, p. 288–9; Haugeland, 1997, pp. 7–8) distinguish *derived* intentionality from *intrinsic* (or original, authentic, etc.) intentionality. *Our* intentional states and processes (e.g., brain states, communicative activities) normally have *intrinsic intentionality*, because they are meaningful to us (the bearers, producers, or consumers of the states and processes). In contrast, information in a computer memory or database has *derived intentionality*, because it is not intrinsically meaningful to the computer, and derives its intentionality only from its meaningfulness to us, the users of the computer.

Intrinsic intentionality is a fundamental (even defining) property of mental states, cognition, communication, and many related phenomena and processes. We can judge the intrinsic intentionality of our own internal states by introspection (the basis of the Chinese Room Argument), but this approach cannot be applied to artificial agents or even to most animals; this complicates the scientific investigation of intrinsic intentionality's physical basis. One of the principal motivations of synthetic ethology is the creation of systems that exhibit intrinsic intentionality, but are simple enough to permit complete explication of the underlying mechanisms. To accomplish this, we must identify non-introspective criteria of intrinsic intentionality.

How can we determine if physical states are intrinsically meaningful to agents, natural or artificial? For the purposes of this article, the argument must be abbreviated and restricted to a particular manifestation of intentionality, namely communication (see MacLennan, 1992, for a fuller discussion). We may begin with Grice's (1957) analysis of the *meaning* of a communication act as the speaker's intention to affect the audience by means of the audience's understanding of that intention. More generally we may say that in a *communication act* one agent behaves with the *intention* (purpose) of eliciting a response (external or internal) in other agents that perceive the behavior or its result. Here, of course, "intention" refers to a certain goal-directed or purposeful behavioral disposition or state of mind (e.g., Blackburn, 1994, p. 196; Gutenplan, 1994, pp. 375–9). (Therefore, an intention is an instance of intentionality by virtue of its being *about* its goal or purpose. Thus an internal state may be "intentional" both in the narrower sense of being an intention and in the wider sense of having intentionality.)

Determination of intention or purpose is problematic, of course, especially in the context of non-human agents, such as the artificial agents used in synthetic ethology, but we can learn from the ways that ethologists make these determinations about non-human animals. In general, ethologists explain the purpose of an innate behavior in terms of its selective advantage in the species' EEA. In particular, questions of whether genuine communication is taking place are answered by looking at its effect on the inclusive fitness of a group of animals in their EEA (Burghardt, 1970). In broad terms, we may say that an animal's behavior is purposeful if it has, or has had, the probability of being relevant to the survival of that animal or its group. We can apply a similar criterion in synthetic ethology, indeed more rigorously than it is applied in natural ethology; for we can test directly whether particular behaviors or internal states of the agents contribute to their survival in the environments in which they have evolved. Such experiments can also reveal the meaning of these states or behaviors to the agents, that is, their specific relevance to the agents' inclusive fitness. Certainly, this is not the only approach to determining purpose, but it has the advantage of being applicable to very simple artificial agents.

2.3 Ecological Validity and Pragmatic Context

Ecological validity refers to the fact that many behaviors are adaptively meaningful only in a species' EEA, that is, only in the environment that has conditioned the species' adaptations. When agents are placed in conditions that are too different from their EEA, they behave in abnormal ways, from which it may be difficult to draw valid conclusions about normal behavior (Neisser, 1976, pp. 2, 7–8). Indeed, this is the motivation for ordinary (natural) ethological methods, which relate behavior to its EEA, as opposed to behaviorist methods, which typically study behavior in unnatural laboratory settings. Internal ("mental") states and external signals acquire meaning through their functional role in the life of the agents, and so they can be understood best in relation to their EEA. Therefore synthetic ethology strives for ecological validity by studying behaviors and cognitive phenomena in their synthetic EEA.

A related issue is the *pragmatic context* of a behavior. In particular, when we are dealing with communication, and especially when we are concerned with non-human communication, we must recognize that communication is rarely purely semantic, that is, serving the purpose of transmitting a proposition (truth-bearing signal). Indeed, communication may be deceptive, among humans as well as other animals, and often serves non-propositional purposes. This is well known from studies of animal communication as well as from philosophical investigations of ordinary language use (e.g., Austin 1975; Wittgenstein, 1958, sec. 19). Now, *pragmatics* refers to the purpose served by a communication or other behavior, and, as for intentionality, this purpose can be derived or intrinsic, that is, derived from the designer of an artificial agent, or intrinsic to the agent itself. Therefore, in order to investigate behaviors with an intrinsic pragmatic context, those behaviors must be fulfilling some purpose intrinsic to the agents, considered either individually or as a population. Synthetic ethology addresses these issues by investigating agents in an environment in which their behaviors matter *to them*. This is so because an agent's behaviors affect its inclusive fitness, that is, the reproductive fitness of itself or related agents. (It is not necessary, of course, that the agents be *aware* that the behavior matters to them.)

In summary, we may say that a communication act is *meaningful* to an agent if within a pragmatic context it serves a purpose, which is understood as the likelihood of increasing the selective advantage of the agent or its group in its EEA (thus maintaining ecological validity). In broad terms, the *meaning* of the communication act may be identified with this purpose. (This is the sense with which "meaning" and "meaningfulness" are used in this chapter.)

2.4 Synthetic Worlds are Physically Real

Since most synthetic ethology experiments take place on general-purpose digital computers, there is a danger of confusing them with simulations. The distinction is important because, as has been remarked often, no one gets wet when a meteorologist simulates a hurricane on a computer. If we are using simulated agents to investigate intentional phenomena, then the objection may be made that although the agents may *simulate* understanding (for example), they do not *really* understand anything, since nothing real is taking place in the computer; it is all simulated. For example, it may be claimed, there is no true meaningful communication, only simulated communication; any apparent intentionality is either derived or simulated (i.e., illusory).

Putting aside, for a moment, the topic of simulation, it is important to stress that synthetic ethology depends crucially on the fact that a computer is a physical device (made of silicon etc.), and therefore that the states and program-controlled state-changes within the computer are (real) physical states and state-changes (involving the movement of electrons etc.)[1] In effect, the program determines the way in which physical law operates within the confines of the computer. Of course, in contrast to *simulated* physical processes, the *real* physical processes in a computer cannot disobey physical law, but these (real) physical processes can be controlled by the program (itself a physical configuration in the computer). In particular, a computer is a non-equilibrium thermodynamic system, and so a program controlling the computer can deploy physical law in such a way that synthetic agents are also (literal, physical) non-equilibrium structures, which must behave (physically) in a specified manner in order to maintain their structure. Thus the behavioral and cognitive processes in the synthetic agents have real relevance to their continued existence ("survival") as real, physical non-equilibrium systems.

It is on this basis that we can claim that agents in synthetic ethology experiments exhibit intrinsic and not just derived intentionality. That is, within the synthetic world constructed in the computer (which is physically real, despite being synthetic), the internal states and behaviors of the agents will have a real influence on their persistence as definite physical structures (particular arrangements of matter and energy in the computer). Therefore these states and behaviors are meaningful *to them*. By observing the relevance of these states and behaviors to the agents, we, as outside observers, may *infer* (more or less correctly and more or less precisely) the meaning of the states and behaviors for the agents, much as if we were observing another animal species. The meaning *we* attribute to the states and behaviors will be *derived* from their meaning to the agents. For us the states and behaviors have only derived intentionality, but for the agents they have intrinsic intentionality.

2.5 Synthetic Ethology Experiments vs. Ethological Simulations

I must say a little more about the difference between synthetic ethology and ordinary ethological simulations, since the two are easy to confuse. First, recall that a scientific *model* is a representation of some system of interest, such a representation being chosen because it is easier to investigate in some way than the system being modeled (e.g. Bynum, Brown, & Porter, 1981, pp. 272–4). Thus a model always has reference to some other *subject system* (which is the primary system of interest and justifies the existence of the model), and so we are concerned with the relation of the model to this subject system, such as its accuracy, range of applicability, etc. There are of course many kinds of models: mathematical models, mechanical models, computer models, and so on. A (computer) *simulation* is a particular kind of model in which a computer is programmed to be a model of some other subject system of interest. It is worth noting that in a running simulation we have one physical system (the programmed computer) being used as a model of another (the simulated subject system). In this way a computer model is like a mechanical model, in which one physical system serves as a model for another.

To understand the difference between experiments in synthetic ethology and ordinary ethological simulations, we can look at an analogous distinction in the physical sciences. First consider a scientific investigation using an ordinary computer simulation of a hurricane. Here one physical system (the programmed computer) is used as a model of anoth-

[1] I hope I may be excused for belaboring the obvious fact that a computer is a physical object and that its computation is a physical process. It is a common source of confusion due, no doubt, to our habit of treating programs and computers as abstractions (e.g., Turing machines).

er (the earth's atmosphere), and the usefulness of the model (in a particular context of questions) will depend on the relation between the two systems. Next consider a different scientific investigation, in which we attempt to discover and validate gas laws (e.g., Boyle's, Charles') by placing various gases in a cylinder in which their pressure, volume, and temperature may be controlled and measured. In this case we have created a specialized physical system designed in such a way that we can control the relevant variables and investigate their interaction. However, this physical system is not a model of anything else; its purpose is not to facilitate the investigation of some *other* system, but to facilitate the investigation of *its own* properties in order to discover or validate general laws. Certainly an improved understanding of the gas laws will help us to understand atmospheric hydrodynamics, but our gas cylinder is not intended as a model of anything in the atmosphere.

Similarly, although synthetic ethology makes no attempt to simulate specific natural systems, it may produce scientific results that are relevant to a wide variety of natural systems. This is because it is directed toward basic science, for synthetic ethology is based on the observation that fundamental, general scientific laws have usually been discovered and confirmed by means of experiments in which there are relatively few variables, which can be controlled precisely. Behaviorist experiments have the control but lack ecological validity. Ethological field studies have ecological validity, but the number of variables is enormous and difficult to control. Synthetic ethology maintains both control and ecological validity by having the agents evolve in a complete but simple synthetic world. Furthermore, because the mechanisms of behavior are transparent, synthetic ethology may facilitate the discovery of causal laws, whereas the nervous-system complexity of animals defeats a detailed causal account of behavior (at least with current technology). The goal of synthetic ethology, therefore, is to discover fundamental scientific laws of great generality underlying intentional phenomena in natural and synthetic systems. Once discovered, their applicability to particular natural systems could be confirmed through ordinary ethological investigations and simulations.

It remains to consider one more issue concerning the difference between synthetic ethology and ordinary ethological simulations, for it is relevant in assessing the relevance to synthetic ethology of simulation studies of language evolution, animal behavior, and related topics. A system that is designed as a model of some other subject system may be found to exemplify properties of much wider interest. For example, Lorenz's computer simulation of weather patterns turned out to be more interesting as an example of chaotic dynamics (Gleick, 1987, ch. 1). That is, we may shift attention from the simulating system as a model of something else to a system interesting in its own right; this is a change in the *goal* of the scientist using the system, not in the nature of the system. Therefore it is quite possible that an ethological simulation of some interesting natural process (e.g., the evolution of vervet alarm calls) might turn out to be interesting in its own right as a physical system exemplifying general scientific principles, and thus be usable in synthetic ethology. As in the case of Lorenz's simulations, we might use the system to investigate principles that are much more widely applicable than just to vervet alarm calls. Thus we may find particular ethological simulations that can be reinterpreted and used as synthetic ethology experiments.

3 Review of Early Results

In this section I will review, very briefly, our experiments demonstrating the evolution of intrinsically meaningful communication in a population of simple machines (MacLennan, 1990, 1992, 2001, 2002; MacLennan and Burghardt, 1993). I will also mention some related experiments by other investigators, but I will not attempt a comprehensive literature review.

3.1 One-symbol Communication

3.1.1 Method

In our first experiments we wanted to determine if it was even possible for genuine communication to evolve in an artificial system (for additional detail, see MacLennan, 1990, 1992, 2001, 2002; MacLennan and Burghardt, 1993). Therefore, in order to put selective pressure on the evolution of communication we decided to select for behavior that would be aided by communication, but could be accomplished less effectively without it. That is, there should be something relevant to the agents for them to communicate about. Therefore we decided to select for a simple kind of cooperation that would be more likely if one agent had information about another agent that was not directly accessible to the first agent.

This requirement was satisfied by placing each agent in a *local environment*, the state of which was directly accessible only to that agent. In accord with our goal of keeping the experiments as simple as possible, each local environment was defined to be in one of a small number L of discrete states; in this first series of experiments, $L = 8$. Each agent could sense the state of its local environment, but not alter it. The states of the local environments were randomized periodically so that there would be no way to predict them.

In addition to the local environments associated with each agent, there was a single *global environment* to which all the agents had access. It could be sensed by all the agents but also modified by them. Thus the global environment provided a potential medium for communication, but of course there was no built in requirement that the agents use it for this or any other purpose. For simplicity, the global environment was restricted to a small number G of discrete states; in these experiments, $G = 8$. See Fig. 1 for a schematic diagram of the relation of the agents and environments.

Our agents were capable of only two kinds of behavior: they could *emit* (that is, change the global environment) or they could *act* (attempt cooperation). In these experiments cooperation took a very simple form, namely, an agent A attempted to cooperate by trying to match the local-environment state of a specific other agent B. Since there were L possible local-state values, in any given situation there were L different possible actions A might take, denoted **act** (m), for each local-state value m; such an action was an attempt to match m against B's local environment. If B's local environment was indeed in state m, then a match occurred, and the agents were deemed to have cooperated; if they did not match, the attempted cooperation failed.

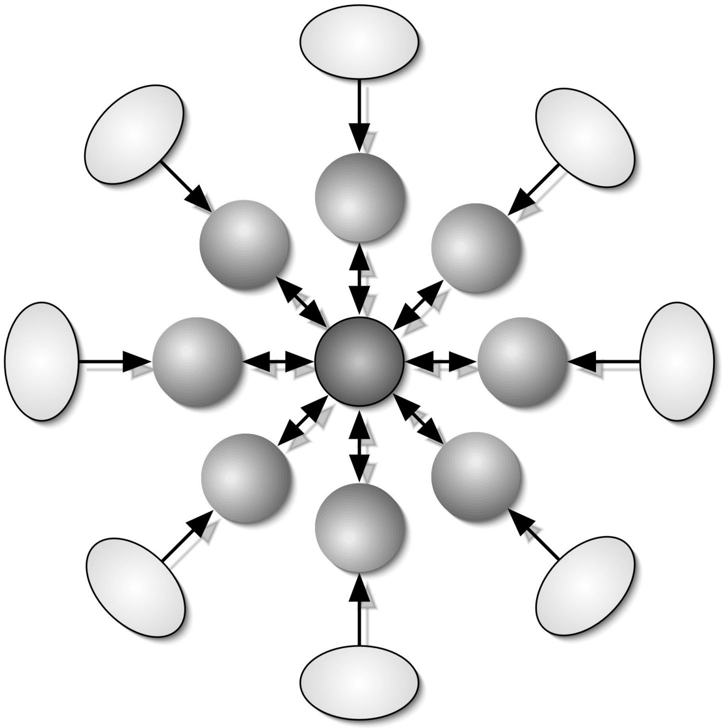

Figure 1: Schematic of environment structure: outer ovals are local environments, central circle is global environment, and spheres are agents. This diagram shows only eight agents and their local environments.

Since *B*'s local-environment state was unpredictable, in the absence of some form of communication the chances of a successful match were $1/L$. The target of the attempted match *B* was required to be the last emitter, that is, the last agent to have changed the global environment. Restricting the attempted match to pertain to a specific agent made random actions less likely to succeed than if a match to any other agent were allowed.

Although this matching process was simple, it was genuine cooperation, for it contributed to the continued existence of the cooperating agents. Specifically, when an agent succeeded in matching the local environment of the last emitter, both agents received a point of credit. If it failed, then neither agent was credited. (We also investigated variants in which there was penalty for failed matching attempts, differential rewards, etc.) As will be explained in more detail later, these credits were used to determine which agents were more likely to "survive" (persist as organized structures in the population) and to reproduce. Therefore, successful cooperations (by chance or any other means) enhanced the fitness (probability of survival and reproduction) of the cooperating agents.

These rules for cooperation may seem somewhat abstract and arbitrary. They were chosen for their simplicity, since synthetic ethology does not require us to simulate any particular natural system. Nevertheless, there is a sort of story we can tell that may make them more comprehensible. When an agent *A* emits, it can be thought of as a call that it needs help in dealing with something in its local environment. Another agent *B* may respond by acting, but it will succeed in cooperating only if it acts in a way appropriate to the agent A's local environment (i.e., if its action matches A's local environment). Be that as it may, in synthetic ethology we are free to define the "laws of nature" in our synthetic world in whatever way required for our experiments.

Our agents require some behavioral control mechanism, which allows them to sense the state of the global environment and of their own local environment, and then to behave in either of two ways: change the global environment to a particular value or act in a particular way. We use the notation **emit** (g) to denote the action of changing the global environment to state g, and **act** (l) for the action of attempting match state l in the last emitter's local environment. In addition, it is useful if an agent's actions depend on its own internal state, which can be thought of as the contents of its short-term memory. (In this first series of experiments the agents did not have any memory, however.)

We have used two behavioral control mechanisms in our experiments, finite-state machines (FSMs) and artificial neural networks (ANNs). In this first series of experiments we used FSMs. They get their name from the fact that they have a finite number I of internal or memory states. Therefore, a FSM can be defined by a table, called a *state transition table*, that describes the machine's behavior for each possible combination of inputs and internal state. For our machines, the inputs are the shared global state and the machine's own local-environment state. Thus there are $I \times G \times L$ possible conditions to which the machine must respond, and therefore $I \times G \times L$ entries in its state transition table. Such a FSM's transition stable can be visualized as follows:

old int. state	global state	loc. env. state	new int. state	response
⋮	⋮	⋮	⋮	⋮
3	7	2	5	**emit** (1)
⋮	⋮	⋮	⋮	⋮
4	0	1	4	**act** (4)
⋮	⋮	⋮	⋮	⋮

To the right of the double line, the table lists a new internal state and a response for each possible combination of old internal state, global-environment state, and local-environment state, as enumerated to the left of the double line. For example, this table says that if the machine is in internal (memory) state 3, and the shared global state is 7, and the machine's own local-environment state is 2, then the machine will change its internal state to 5 and respond with **emit** (1), which changes the global-environment state to 1. Similarly, if its internal state is 4, the global state is 0, and its local-environment state is 1, then it will stay in internal state 4 and respond with **act** (4), which attempts to match 4 against the local-environment state of the last emitter.

In each of these conditions the table must determine a new internal state (I possibilities) and either emit (G possibilities) or act (L possibilities). So there are $I(G + L)$ different ways the machines can respond to each condition. Since in this first series of experiments $I = 1$ (no memory), there are GL table entries and $G + L$ possible responses. In particular, since $G = 8 = L$, there are 64 entries, each defining one of 16 responses. (Henceforth we will assume $I = 1$ unless otherwise stated.)

It is worth observing that although the FSM is a simple behavioral control mechanism, it has some subtlety in this application. This is because the behavior of an agent is always determined by the combination of global, local, and internal state. Therefore, for example, the signal emitted for a particular local-environment state depends also on the global state and the internal state; that is, the signal emission is context-dependent. So also, the way an agent responds to a signal in the global environment depends on its own internal state (as we would expect), but also on its own local environment, another kind of context-dependence. An agent's response to each particular combination of circumstances is potentially unique; there is no built in ability to generalize over similar situations. Therefore if agents have evolved that are able to signal their local-environment state in a context-free way, that is, independently of the global environment and their internal state, it is because they have adapted to do this in every combination of global and internal state with this local state. Likewise, to respond to a signal independent of context, the agents must have this response for every combination of local and internal state. So the evolutionary problem that the population has to solve is actually quite difficult. (As will be discussed later, we have also investigated behavioral control mechanisms, such as ANNs, which do have some ability to generalize.)

We also experimented with a simple form of single-case learning, which could be enabled or disabled as an experimental control. When learning was enabled, it operated in the following way. Suppose in global-environment state g and local-environment state l the table for an agent B defined the response **act** (m), that is, attempt to cooperate with action m. Further suppose that this was the incorrect action, because local-environment state of the last emitter (A) was n. In this case the table for agent B will be changed to **act** (n) under conditions (g, l). In other words, it is changed to *what would have been* the correct action in the current conditions. This is actually a very weak form of learning, since there is no guarantee that action n will be correct the next time the global environment is g and this agent's local-environment state is l.

Our goal was to investigate communication in its environment of evolutionary adaptedness, and therefore it was necessary for our population of agents to evolve. The behavioral table of a FSM is represented simply by a "genetic string" of GL genes (representing the possible conditions), each of which has $G + L$ alleles (representing possible responses to those conditions). This was coded simply as a string of GL numbers in the range 0 to $G + L - 1$. This string defines the *genotype* of an agent, which is used to initialize its behavioral table when it is "born" (see below). The behavioral table, representing the agent's *phenotype* is constant throughout the agent's "life" if learning is disabled, but if learning is enabled, the behavioral table (phenotype) may change according to the learning rule.

Our goal was to select for cooperative activity, therefore we counted the number of successful cooperations for each agent over a specified interval of time. Specifically, each agent was given several opportunities (5, in these experiments) to respond to a given configuration of (randomly determined) local-environment states. Then the local environments were re-randomized and the agents were tested again; this was repeated several times (10, in these experiments). The *fitness* of an agent was defined to be the number of successful cooperations in this interval of simulated time, and thus reflects the rate of cooperation.

The preceding events are called a *breeding cycle* because at the end of them two agents are chosen to “breed,” producing a single “offspring,” and one agent is chosen to “die” and be replaced by that offspring. Thus the population size is constant (100 in these experiments). The probability of choosing an agent to breed was made proportional to its fitness (rate of cooperation), and the probability of dying was inversely related to its fitness. (Preliminary experiments showed that always choosing the most fit to breed and the least fit to die led to premature convergence in the population.) After replacement of the “dead” agent by the parents’ offspring, a new breeding cycle began, and the run continued for a specified number of breeding cycles (5000, in most experiments). The initial populations had randomized genomes.

Genetic algorithms (GAs) typically replace the entire population after breeding, thus defining non-overlapping generations (e.g., Goldberg, 1989). We decided to use incremental replacement — one individual at a time — to allow a simple form of “cultural transmission” of learning, which we thought might be important when learning was enabled.

It remains to say how the genotype of the offspring was determined. The genetic operators were similar to those used in GAs, but with minor differences. First, a genetic string for the offspring was determined by *two-point crossover* of the parents’ genotypes. That is, two uniformly random numbers k, l were chosen in the range 1 to GL. The offspring’s genes in the range k to l were taken from one parent, and those in the range 1 to $k - 1$ and $l + 1$ to GL from the other. Finally, with low probability (typically 0.01) a random gene was mutated by replacing it with a randomly chosen allele in the range 0 to $G + L - 1$. The resulting genetic string became the genome of the offspring and determined its phenotype (initial behavioral table).

During these experiments we gathered several kinds of data in order to assess whether communication had evolved in the population. According to Burghardt’s (1970) definition of communication, if genuine communication is taking place then it ought to have a demonstrable positive effect on the inclusive fitness of the population. Therefore the most fundamental information we gathered was *degree of coordination*, defined as the average number of cooperations per breeding cycle. We computed both the maximum and average for the population. The time series of these quantities (smoothed by a moving average) allowed us to track any progressive changes in the rate of cooperation in the population and its best representative.

In order to be able to investigate any communication that might evolve, we also compiled a *co-occurrence table* during each experiment. This was a $G \times L$ matrix in which the (g, l) entry reflected the frequency with which global-environment state g and local-environment state l co-occurred in a successful cooperation (that is, the correct action l was performed). If no communication were taking place, then one would expect all (g, l) pairs to be equally likely. On the other hand, if systematic communication were taking place, in which certain global states g (“symbols”) are used to denote certain local-environment states l (“situations”), then one would expect a non-uniform distribution.

Furthermore, if communication were evolving in the population, then one would expect to see a change in the co-occurrence matrix over time, from a uniform distribution at the beginning of the experiment and becoming progressively more structured as communication emerged. Therefore, we computed the co-occurrence table over the recent history of the simulation (50 breeding cycles), so that it would reflect the behavior of the population at a particular time. To quantify the degree of structure in the matrix we used several measures, including entropy, coefficient of variation, and chi-squared. By plotting these quantities as a function of time we were able to see changes in degree of structure as evolution progressed.

3.1.2 Symbols and Communication

In anticipation of discussing the results of these experiments, it is worth observing that although the preceding use of "symbols" to refer to global-environment states might be considered metaphorical, it is in fact consistent with the technical use of this term in semiotics (e.g., Colapietro, 1993, p. 190), deriving from Peirce (1955, 102–3 = *CP*[2] 2.243–52). To explain this, it is necessary for me to reiterate several characteristics of these synthetic ethology experiments. The global and local states are both physical states in the synthetic world external to the agents; they correspond to physical conditions in the vicinity of an animal. (Global-environment states are analogous to physical sound waves in the air; local-environment states are analogous to conditions nearby an animal, such as the physical presence of a food source, but these are just analogies.) Global-state values and local-state values come from conceptually distinct alphabets, although for programming convenience in these experiments they were both represented by nonnegative integers (i.e., the global-state values are in {0, …, *G*–1} and the local-states are in {0, …, *L*–1}). As will be seen when we discuss the results of these experiments, statistical regularities in the co-occurrence tables (e.g., Tables 2 and 3) show that global-state values are used systematically by the agents to denote local-state values. Therefore, certain physical conditions (global states) are used to *denote* other physical conditions (local states). Thus the global states are *referential*, which qualifies them as *signs* in the usage of semiotics (Colapietro, 1993, pp. 179–80).

Furthermore, the relationship between the global-state values and local-states values (between the *signs* and their *referents*) is arbitrary and conventional. In different experiments (i.e., with different random initial populations) we get different correspondences between the signs and their referents. Indeed, the correspondence must be arbitrary, for the global- and local-state values are drawn from different alphabets and the machines have no mechanisms for comparing the elements of these different spaces. (They can, in effect, compare local states to local states and global states to global states, but not local states to global states; this is implicit in the structure of the FSMs. Also, the agents have no access to the fact that in the computer both global-state values and local-state values are represented by nonnegative integers, that is, by bit strings.) Therefore, the correlation between the sign vehicle (global-state value) and its object (local-state value) is not based on any similarity or physical connection, and so, in Peirce's terms, these signs are neither *icons* nor *indices*, but *symbols* (Colapietro, 1993, p. 190; Peirce, 1955, pp. 102–4 = *CP* 2.243–52, 2.304). The initially random correspondences are amplified by evolution and (in some experiments) learning; that is, they are conventional.

[2] "*CP*" refers to Peirce's *Collected Papers* (1931–5).

While on the topic of semiotic issues, it will be worthwhile to mention several other characteristics of these symbols and the agents' use of them. First, these symbols are context sensitive, for an agent's behavior is always determined by a combination of the global-environment state and its local-environment state. As it turns out, this context dependence is of no use to the agents in these experiments (and indeed interferes with the straight-forward use of the global environment for communication), and so the agents have to adapt (through evolution and perhaps learning) to ignore their local-environment state when attempting to respond to signals in the global environment. (Conversely, they need to adapt to ignore the global environment when attempting to signal the state of their local environment.)

Second, we normally expect an external sign to correspond to some internal representation, its *sense* (*Sinn*), which is distinct from its *referent* (*Bedeutung*). As I have remarked elsewhere, the agents in these experiments are so simple that they can hardly be said to have psychological states. Nevertheless, if we look closely we can see a rudimentary internal representation of external conditions. For when an agent responds to its environment, the computer implementing the FSM must copy the global- and local-environment states from their physical locations in the computer's memory, and use them to compute an index into the FSM's transition table. Thus these two physical conditions external to the agent are copied to different physical locations and integrated to determine the agent's behavioral response. (Had we not used nonnegative integers to represent the global-state values, a less trivial transformation to the internal representation would have been required.) We find similarly rudimentary internal representations in simple organisms such as bacteria.

Third, we can understand a symbol as a triadic relation between a sign, its referent, and the *receiver* or *interpreter*, and we can observe this triadic relation in our synthetic ethology experiments. For example, the co-occurrence matrices (e.g., Tables 2 and 3) show how the symbols (global-environment states) are interpreted by the population at a given point in time. A substantially different population (either in a different experiment, or in the same experiment at a substantially different point in time) may interpret the symbols differently. Furthermore, as will be discussed later, we can sometimes discern two or more subpopulations that interpret one or more symbols differently. Finally, we can look at individual agents and determine what a symbol means to *it* (i.e., how it responds to the symbol) in any particular context.

Indeed, one of our FSM's behavioral rules (i.e., rows in its transition table) can be understood directly in terms of the five-place relation that, according to Morris (1964, p. 2), characterizes a *sign process* or *semiosis*. For suppose that under conditions (g, m) the FSM R responds **act** (n) and that the last emitter's local environment is in state n. Then in this semiosis, (1) global state g is the *sign*, (2) R is the *interpreter*, (3) **act** (n) is the *interpretant* (the effect of the sign qua sign), (4) local-state n is the *signification* (or *referent*), and (5) m is the *context* of the sign.

Finally, before discussing the experimental results, it is worth observing that synthetic ethology affords much easier experimental control than natural ethology, and we made use of this control in these experiments. The fundamental control addressed Burghardt's (1970) definition of communication: for genuine communication to be taking place we would have to show that it was contributing to the inclusive fitness of the population. Therefore we wanted to be able to compare the evolution of the population under conditions in which it was possible for communication to evolve with those under which it was impossible. Thus we designed the experiments so they could be run with communication suppressed or not. To suppress communication, we randomized the global-environment state at every opportunity, in effect raising the noise level so high that the only potential medium of communication was unusable. This allowed us to run parallel experiments with genetically identical (random) initial populations differing only in whether communication was suppressed or not. The second major control that we used in this series of experiments was to enable or disable the simple learning rule described above. This allowed us to do some preliminary investigations of whether learning facilitated the evolution of communication or not.

3.1.3 Results

We ran over one hundred experiments with the parameters as described above. In most cases we made three parallel runs with the same random initial populations: (1) communication suppressed, (2) communication not suppressed and learning disabled, and (3) communication not suppressed and learning enabled. This allowed us to investigate the effects of communication and learning independently of the initial population. Although there was considerable quantitative difference from experiment to experiment, due to the random initial populations and the many other random factors, nevertheless the qualitative results were quite predictable. Therefore, in the following I will discuss a typical experiment.

Analysis (MacLennan, 1990) shows that in the absence of communication agents can be expected to exhibit a degree of coordination of 6.25 cooperations per unit time (1 breeding cycle, in this case). Indeed, when communication is suppressed we find that the average degree of coordination begins at this value and stays very close to it. Nevertheless, linear regression shows a slight upward trend, 3.67×10^{-5} coop. / unit time / unit time, a somewhat surprising result discussed later. The degree of coordination was up to 6.6 coop. / unit time after 5000 breeding cycles, and had been as high as 6.95 (see Fig. 2).

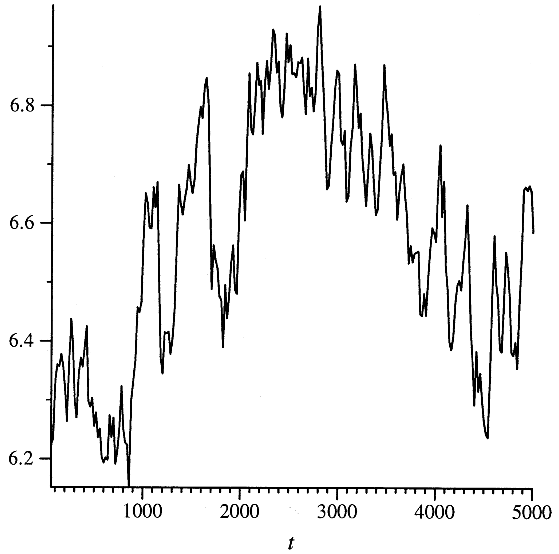

Figure 2: Degree of Coordination: Communication Suppressed.

When communication was not suppressed (but learning was still disabled) the degree of coordination began at the chance level, but increased at a rate of 9.72×10^{-4} coop. / unit time / unit time, a rate 26 times as great as when communication was suppressed. After 5000 breeding cycles the degree of coordination was up to 10.28 coop. / unit time, which is 60% higher than when communication was suppressed, and had been as high as 10.6 (see Fig. 3). This significant difference shows that the population is using the global environment for genuine communication, since it has increased the agents' inclusive fitness.

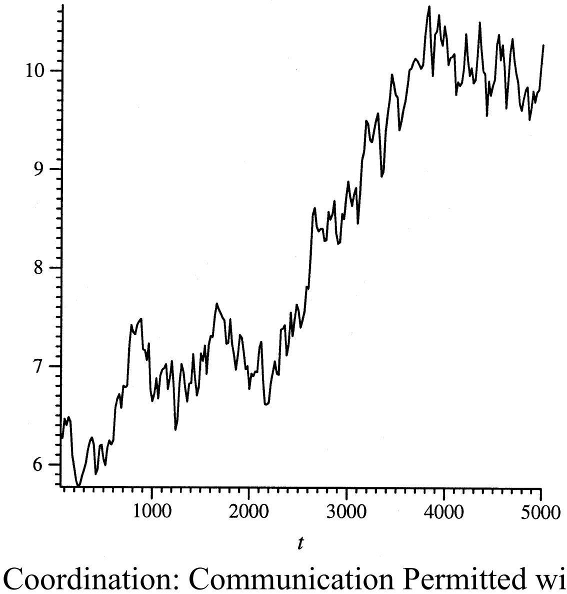

Figure 3: Degree of Coordination: Communication Permitted with Learning Disabled.

When communication was not suppressed and learning was enabled, the degree of coordination began at about 45 coop. / unit time, which is much higher than the 6.25 expected with neither communication nor learning. This high level of coordination, achieved before communication has evolved, is a consequence of the fact that an agent has several opportunities to respond to a local-environment configuration before the local environments are re-randomized. So there is a baseline advantage to learning even in the absence of communication. Over the 5000 breeding cycles, the degree of coordination increased at a rate of 3.71 × 10^{-3} coop. / unit time / unit time, which is 3.82 times the rate when learning was disabled and approximately 100 times the rate when communication was suppressed. By the end of the experiment the degree of coordination had reached 59.84 coop. / unit time, which is 857% above that achieved when communication was suppressed (see Fig. 4). Therefore learning reinforces the selective benefits of communication.

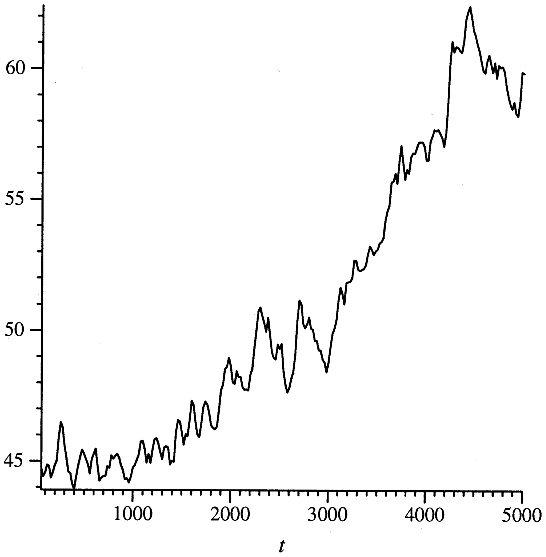

Figure 4: Degree of Coordination: Communication Permitted with Learning Enabled.

Finally, we must consider why there is a very slow increase in the degree of coordination even when communication is suppressed. This results from the population evolving to exploit a loophole in the fitness determination rule by means of *partial cooperation*. Recall that a cooperation is judged to have taken place only if the action of an agent matches the local environment of the *last emitter*. (To allow it to match any local environment would increase the probability of chance cooperation too much, decreasing the selective pressure.) Therefore the population can increase the probability of cooperation by co-evolving so that agents emit only in a small subset of local-environment states and only act in ways appropriate to this same subset. As a further consequence we observed that in long experiments with communication suppressed, the uniform co-occurrence matrix would become slightly structured due to attempted cooperations being attempted in only a subset of the local-environment states (see below). This explanation was confirmed by Noble and Cliff (1996). It is worth emphasizing that these simple agents found a way to improve their performance that was not anticipated when we designed the experiment, and that required some investigation in order to explain.

We can get more information about the agents' evolved ability to communicate by inspecting the co-occurrence tables. As previously mentioned, we quantified the structure of the tables by several measures, including entropy. For $G = 8 = L$ the maximum entropy, which occurs with a uniform distribution, is $H_{\max} = 6$ bits. For comparison, we can compute the entropy for an ideal code, in which there is a one-to-one correspondence between global- and local-environment states; it is $H_{\text{ideal}} = 3$ bits.

Table 1: Co-occurrence Matrix: Communication Suppressed

loc.→ **glob.↓**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
0	94	130	133	34	166	0	150	682
1	16	105	279	228	261	307	0	118
2	0	199	229	12	0	0	161	274
3	95	19	93	283	669	89	0	201
4	1	97	212	200	112	0	0	0
5	28	135	84	8	600	215	0	351
6	0	0	0	118	59	70	0	690
7	0	33	41	0	371	0	0	0

When communication was suppressed, the entropy started at approximately 6 bits, but by the end of 5000 breeding cycles had decreased to 4.95. This resulted from partial cooperation, as already discussed. An inspection of the co-occurrence matrix at the end of the experiment, and therefore reflecting the last 50 breeding cycles (Table 1), showed that there was little or no cooperation in a subset of the local-environment states (e.g., 0 and 6).

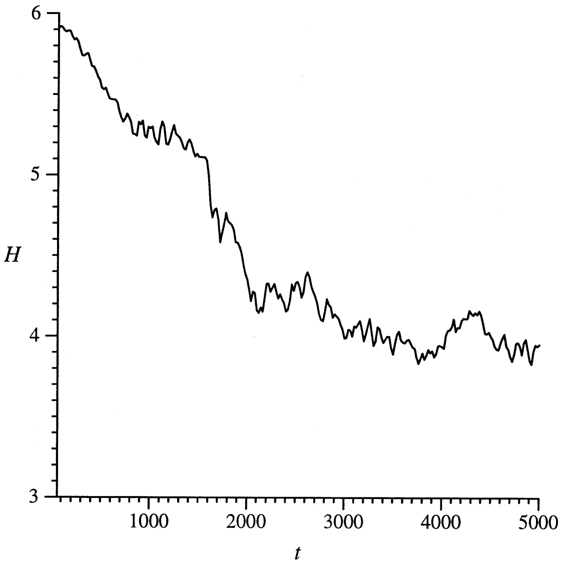

Figure 5: Entropy: Communication Permitted with Learning Disabled.

Table 2: Co-occurrence Matrix: Communication Permitted with Learning Disabled

loc.→	0	1	2	3	4	5	6	7
glob.↓								
0	0	0	2825	0	500	20	0	0
1	206	0	0	505	999	231	2	0
2	1	0	0	277	39	4935	1	2394
3	385	1	1	94	0	0	1483	1
4	0	292	0	0	19	555	0	0
5	0	0	1291	0	0	144	0	0
6	494	279	0	403	0	1133	2222	0
7	140	2659	0	202	962	0	0	0

When communication was not suppressed (but learning was disabled) the entropy decreased to 3.87 bits, which is much closer to $H_{\text{ideal}} = 3$ bits (see Fig. 5). Visual inspection of the final co-occurrence matrix (Table 2) reveals much more systematic use of the global environment as a communication medium. Sometimes a global-environment state denotes a single local-environment state almost exclusively, and vice versa. We also find examples of *ambiguity*, in which a global state denotes primarily two local states, and *synonymy*, in which two global states denote the same local state. Other complexities, typical of natural communication, also appear. The occurrence of ambiguity and synonymy in the population's communication could result from competing "dialects" in the population or from inconsistent signal use by individual agents (or by both in combination). Experiments by Noble and Cliff (1996) seem to have ruled out the dialect explanation. Rather, their research supports ambiguous signal use by individual agents, that is, the same symbol is emitted for two different local-environment states.

Table 3: Co-occurrence Matrix: Communication Permitted with Learning Enabled

loc.→	1	2	3	4	5	6	7
glob.↓							
0	29172	1287	12281	2719	1132	93	3836
1	634	107	1039	0	0	2078	0
2	1306	0	37960	85	410	7306	26611
3	410	0	0	0	126	1306	304
4	0	353	62	575	1268	420	519
5	0	46	469	0	0	0	26
6	156	0	0	0	951	0	1086
7	73	54	0	2764	135	461	102

When communication was not suppressed and learning was enabled, then the entropy achieved in this experiment was 3.91 bits, about the same as in the non-learning case (see Table 3). This is typical; sometimes it is a little greater than the non-learning entropy, sometimes a little less. Table 4 summarizes the entropy and coefficient of variation under the three experimental conditions.

Table 4: Summary of Order Measures

Measurement	Random	Communication/Learning			Ideal
		N/N	Y/N	Y/Y	
Coefficient of Variation, *V*	0	1.27	2.13	2.39	2.65
Entropy, *H* (bits)	6	4.95	3.87	3.91	3

It is worth stepping back from these experiments to stress an important point. The symbols used by the population, as recorded in the co-occurrence matrices, are meaningful *to the agents themselves*, because they are relevant to the inclusive fitness, indeed to the continued existence, of the agents. That is, the symbols exhibit intrinsic intentionality, not a derivative intentionality coming from us as designers or observers of the system. Indeed, we cannot know the meaning of the individual symbols except through empirical investigation. Thus, we may study their use as reflected, for example, in a co-occurrence matrix, and compile a "dictionary" based on observation, or we may "dissect" the agents, as described later, to see how they interpret the symbols (see MacLennan, 1990, for examples). Indeed, any meaning the symbols have for us is derived from the intrinsic meaning of the symbols to the synthetic agents.

Now certainly these agents are very simple, and they do not have any awareness of the meaning of the symbols; their response is purely mechanical. But conscious awareness is not necessary for intrinsic intentionality and meaningful communication, which can be found in microorganisms with no (or very little) awareness. For example, bacteria communicate meaningfully (e.g., Dunny & Winans, 1999) and fitness-enhancing chemotaxis shows they have intentional internal states, for example, internal representations of external chemical gradients (Dretske, 1985, p. 29), but bacteria are not conscious. Thus synthetic ethology provides a means of investigating intrinsic intentionality in its barest, simplest form, which is exactly where experimental investigations should begin.

Neuroethology seeks to understand the neural basis of a species' behavior in an evolutionary context, but neuroethological investigations are difficult because of the complexity of nervous systems, the slow pace of evolution, and the difficulty of doing controlled experiments. On the other hand, in synthetic ethology we are dealing with simpler agents and their behavioral control mechanisms are completely transparent for investigation. If some interesting behavior evolves in a population, then we can "dissect" the members of the population and determine their entire behavioral control system (see MacLennan, 1990, for examples). In particular, if, as in these experiments, the agents evolve to exhibit intrinsic intentionality, then we can completely explicate the mechanism underlying that intentionality; there can be no "ghost in the machine." Thus synthetic ethology provides a means of bridging the gap between inherently mental phenomena, such as intentionality, and the physical processes supporting them.

3.2 Other Experiments

I'll briefly review some of our other early experiments in synthetic ethology. One simple extension to the preceding experiments was to test the population's evolution of the ability to communicate by emitting and recognizing sequences of two signals

(MacLennan, 2001). To create selective pressure toward this result, we reduced the number of global-environment states to $G = 4$ while keeping the local-environment states $L = 8$; thus there were not enough global-environment states to uniquely denote the local-environment states. Of course, using sequential signals requires that the agents be able to remember the signals they have already generated as well as those they have already recognized. Therefore we increased the number of internal (memory) states to $I = 4$. As a consequence, the agents' genomes contained 128 genes with 48 alleles each. In other respects the setup was the same as our previous experiments.

Table 5: Co-occurrence Matrix: Communication Permitted with Learning Disabled (Two Symbols)

loc.→	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
glob.↓								
0/0	31	22	42	0	144	0	0	0
1/0	26	15	62	0	175	0	0	0
2/0	119	23	44	0	47	0	0	0
3/0	8	9	18	0	31	0	0	0
0/1	0	54	106	2	74	59	516	0
1/1	0	33	174	3	423	227	1979	0
2/1	0	23	65	17	139	74	125	0
3/1	0	1	24	0	48	96	51	0
0/2	50	4	4	366	7	0	8	42
1/2	35	9	0	32	1	0	6	44
2/2	52	76	0	112	7	0	13	135
3/2	52	6	1	215	2	0	2	78
0/3	0	2	13	17	0	3	0	0
1/3	0	66	19	6	0	4	0	0
2/3	0	33	61	27	0	2	0	0
3/3	0	39	38	8	0	0	0	0

Two-symbol communication was comparatively slow to evolve and never reached the same degree of organization as we observed for single-symbol communication. For example, the final co-occurrence matrix (Table 5) shows that, for the most part, the meaning is conveyed by the second (i.e., most recently received) symbol. Thus the local-environment state 5 is denoted primarily by signals 0/1, 1/1, 2/1, and 3/1. On the other hand, there are some cases in which both symbols have meaning: although 0/0, 1/0, and 3/0 denote primarily state 4, 2/0 denotes primarily state 0. Furthermore, order is significant, because 0/2 denotes primarily state 3.

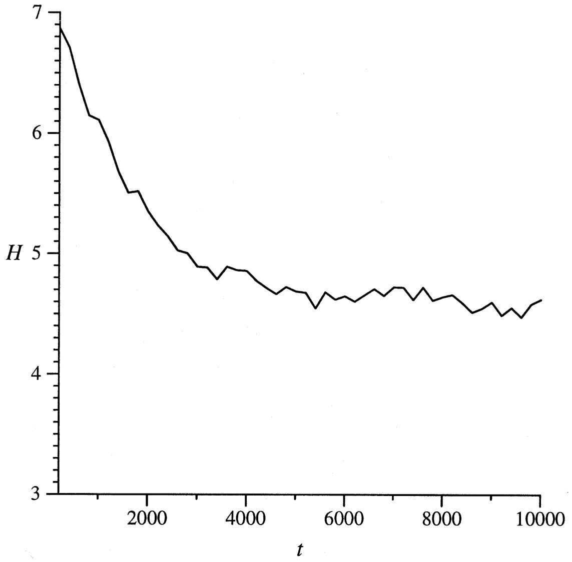

Figure 6: Entropy: Two-symbol Communication.

Although we ran experiments considerably longer than the 5000 breeding cycles used in the preceding experiments, two-symbol communication never seemed to evolve much beyond the level of organization shown in Table 5. This is displayed clearly in Fig. 6, which shows the entropy of the co-occurrence matrix. Over 5000 breeding cycles it decreases from $H_{\text{max}} = 7$ bits to about $H = 4.5$ bits, at which point it stalls, still well above $H_{\text{ideal}} = 3$ bits. Changes to the fitness calculation formula and other experimental parameters (such as population size) did not seem to have much effect on the qualitative result, nor did enabling learning.

In retrospect it is not surprising that these agents did not do better (indeed, it is somewhat remarkable they did as well as they did), for they were being asked to solve a very hard problem. Consider the problem of an agent trying to transmit its local-environment state by a two-symbol signal. Since every emission depends on the combination of local, global, and internal state, any change by another agent to the global environment will probably disrupt the first agent's emission of the second symbol. Also observe that there is no global-environment state that represents the absence of a symbol; therefore there is no way to determine the beginning or ending of a transmission, which means that it is difficult for the population to make use of the distinction between, for example, 2/0 and 0/2. Finally, since every agent must act or emit on every cycle (there is no "do nothing" operation), the only way an agent can avoid a premature attempt at cooperation after receiving the first symbol is be emitting a symbol, which will probably change the global environment and prevent any other agents from receiving the first symbol. Clearly the experiments could be designed differently to avoid these difficulties, and so we explored several alternatives.

For example, Crumpton (1994) studied a system in which (1) in each time slice the agents cycled twice, thus giving them the opportunity to emit or recognize two symbols without interference from other agents, and (2) agents could "do nothing" in addition to acting or emitting; other parameters were the same. In these experiments he observed a significantly higher use of non-repeating symbol pairs than in the preceding experiments, but the entropy level attained was about the same.

As previously mentioned, we considered alternative behavioral control mechanisms to FSMs. For example, my students Stroud and Jerke conducted preliminary experiments using artificial neural networks (ANNs) as the control mechanism (MacLennan, Jerke, Stroud, & VanHeyningen, 1990, sec. 2; MacLennan, 2001). Aside from the greater biological verisimilitude of ANNs, these experiments were motivated by two other considerations. First, in the experiments already described both the global and local environments were associated with discrete state spaces, and so the agents were not faced with the problem of dividing a continuum of states into discrete categories, which is a significant problem faced by biological agents (e.g., Wiley, 1983, pp. 163–4; MacLennan, 1992; Steels, 1997a, 1997b). Second, the behavioral mechanism used in our earlier experiments did not facilitate generalization. Therefore, for example, to attach a unique response to a global-environment state, an agent would have to acquire (through evolution or learning) that response in the context of every possible state of its local environment. Similarly, to emit a unique signal whenever its local environment was in a particular state, an agent would have to acquire that response in every possible prior global-environment state. In brief, the FSM's rules are inherently context-dependent, but the optimal signaling system, for this problem, is context-free.

To explore these issues, both the global and local environments were allowed to take real-valued states in the range [0, 1]. Thus each agent had two real-valued inputs: the state of the global environment and the state of its own local environment. Each agent also had two real-valued output neurons. One, with values in [–1, 1], indicated whether the agent was emitting or acting (as indicated by the sign of the output value). The other output, with values in [0, 1], was the value to be put into the global state in the case of an emission, or the value to be matched against the last emitter's local state in the case of an action (attempted cooperation). The match was deemed successful if the action value was within a certain distance ε of the last emitter's local-state value; to simplify comparison with the earlier experiments, they set $\varepsilon = 0.125$. There were six hidden layer neurons and the network was trained by back-propagation.

The experiments were similar in structure to the FSM-based experiments. The genetic strings determined the signs of the connections between the neurons, but not their weights, which adapted through back-propagation. That is, the genetic strings defined the initial pattern of excitatory and inhibitory connections, but back-propagation learning adapted the connection strengths.

In order to capture signaling patterns in a way comparable to the co-occurrence matrices, Jerke and Stroud divided the global- and local-state spaces into ten bins of size 0.1 and kept track of the correlation of global- and local-environment states whenever a successful cooperation occurred. Non-random signal use, similar to that in the FSM experiments, was observed, but it was partial cooperation (described above) resulting from the agents communicating about only a subset of the state spaces. See MacLennan et al. (1990, sec. 2) and MacLennan (2001) for more information about these experiments.

3.3 Related Work

As previously noted, Noble and Cliff (1996) replicated our earliest experiments on the evolution of communication (Maclennan, 1990, 1992; MacLennan & Burghardt, 1993) and extended them in several informative ways. In addition to investigating partial cooperation and ambiguous and synonymous symbols use, as already mentioned, they investigated the effect of the order in which agents were serviced in the experiments.

Werner and Dyer (1992) also demonstrated the evolution of communication by synthetic ethology. They used ANN-controlled agents of two kinds, "male" and "female," which had to find each other in order to mate. The females were immobile, but could determine the location of the males, who were blind but mobile. In their experiments the population evolved so that the females signaled the males how to find them.

The use of computers to study the origins and evolution of communication and language has developed into a rich and diverse research area. Cangelosi and Parisi (2001) is a good sampling of progress up to that time, and Wagner, Reggia, Uriagereka, and Wilkinson (2003) is a useful review of recent progress. Unfortunately, Wagner et al. incorrectly classify our experiments as an investigation of *nonsituated* communication. In situated experiments, "agents are evaluated based on their performance on a task instead of being directly evaluated on their communication abilities" (Wagner et al., 2003). Although the task performed by our agents is simple, it is distinct from any communication that might or might not be taking place (indeed, it can be accomplished to a limited degree when communication is impossible, as already explained). Agents are evaluated according to whether their actions are appropriate for other agents' local environments, which are out of the agents' control; the local-environment states are determined by the environmental physics of the synthetic world. Thus they exemplify a rudimentary degree of *symbol grounding* (Harnad, 1990), for the signals (global-state values) refer to conditions in the agents' local environments, which are external to them. Agents are causally responsive to their local environments and interact causally with the global environment. Certainly, however, the topology of the synthetic world in our experiments was too simple to permit the exploration of spatial factors, as in the experiments of Werner and Dyer (1992), Cangelosi and Parisi (1998), and Reggia, Schulz, Wilkinson, and Uriagereka (2001), for example.

4 Reconsideration and New Directions

In this section I will address some lessons that we learned from our synthetic ethology experiments and consider some important new directions.

4.1 Making Real Worlds Inside a Computer

In the spirit of keeping the experiments as simple as possible while still exhibiting the phenomena of interest, most of our synthetic ethology experiments have made use of a very simple world, in which the states about which the agents communicated were simple and unstructured. The communication system that they evolved was similarly simple, as were the agents' "psychological states" (internal representations). More structured representations and systems of communication would be expected to evolve in order to cope with a more structured environment. For example, we would not expect to observe the evolution of a language including nouns, adjectives, and verbs unless the synthetic world included objects, properties, and actions of various kinds. Nor, in the absence of these, would we expect the evolution of propositional mental states. We could, of course, build a more structured synthetic world, but there is a pitfall we must avoid in order to investigate these more complex phenomena through synthetic ethology.

As mentioned, the obvious approach to studying more complex communication, psychological states, etc. is to populate the synthetic world with various sorts of macroscopic objects (including other agents) with various sorts of properties. However, no matter how many objects and properties we build into such a system, it will still be unrealistically simple compared to any natural environment. The sum-total of things that can occur in such a world will be limited to the combinations of built-in objects and properties. This is not a problem for some investigations, but if we are trying to understand how animals parse the complexity of the natural world into meaningful categories, then by constructing our synthetic world in this way we will be begging the question, since we will have built the categories into the basic structure of the world. For example, if we build into our system two kinds of agents (which we think of as predator and prey), and we build in a behavior (which we think of as the predator killing the prey), then it is not such a surprise if our agents discover categories corresponding to **predator**, **prey**, and **predator-kills-prey**. In the natural world, in contrast, there is not such a simple relationship between the structure of the world and the categories used by humans and other animals. Even simple sensory categories are radically underdetermined (Steels, 1997a, 1997b). Therefore, if we want to use synthetic ethology to study the emergence of categories, we must be careful not to build them in from the beginning.

Certainly, the synthetic ethology experiments described in this chapter are not immune to this objection. Most of these experiments made use of unstructured discrete state spaces and the relevant categories were small finite sets (e.g., a specific value from one space paired with all possible values from the other). Even our ANN-based experiments, which use continuous state spaces, did not require significant category construction. Therefore I believe that future experiments should address the emergence of categories meaningful to the agents, as opposed to those meaningful to the experimenters.

If we don't build objects and properties into our synthetic world, how will they get there? Here we may take a hint from the natural world, in which objects and properties are emergent phenomena arising from the interaction of fundamental particles. Thus, in physical terms, when we say, “the fox sees the hare,” we are giving a high-level, approximate description of a situation in which the particles constituting the hare are interacting with the particles of the fox through the intermediary of photons. From a physical perspective, only the particles and their interactions are real in a primary sense; the fox, hare, and act of seeing have only derivative reality, as high-level descriptions of fundamental physical reality (i.e., the perspective of the *intentional stance,* Dennett, 1987). Nevertheless, the hare is meaningful to the fox (for it is relevant to its continued existence as an organized physical system). Therefore, if we want to study the *emergence* of meaning, we cannot begin with macroscopic objects and relations (e.g., **fox**, **hare**, **sees**), for then we have simply encoded *our* meanings (*our* context of relevance) into the synthetic world. Rather, we must allow meaning to emerge from underlying, fundamentally meaningless processes (i.e., processes explainable from the physical stance but not from the intentional stance), as it does in the natural world.

How can this be accomplished? One way would be to simulate the real world at the level of elementary particles (or strings) and allow meaning to emerge in macroscopic objects, just as it does in the natural world. However, aside from the computational impossibility of doing such a simulation, the shear complexity of the system (comparable to nature itself) would limit our ability to understand it. The goal of synthetic ethology is to investigate experimentally systems that are as simple as possible while still exhibiting the phenomena of interest (in this case, the emergence of meaning).

We do not want (and cannot have) the complexity of the natural world in our synthetic worlds, but we can use a simpler version of the same separation of levels. By a judicious choice of microscopic objects and interactions to build into our world, we can have emergent macroscopic objects and interactions with a rich and unpredictable structure. The agents, which are macroscopic objects whose fundamental interactions with the environment are at the microscopic level, will have to construct whatever macroscopic categories they need. In this way we will not "rig the game" by building them in from the beginning, although of course they are implicit in the microscopic objects and interactions.

It may seem implausible that we could design a synthetic world that is, on the one hand, complex enough to exhibit emergent objects, properties, and interactions, and is, on the hand, simple enough to be computationally tractable and transparent to investigation. At present, this is a topic for investigation, but I can offer one example of a possible approach.

It is well known that cellular automata (CAs), such as Conway's "Game of Life" (Gardner, 1970), can exhibit rich emergent behavior. Although the individual cells interact with their neighbors in a simple way, in "Life" we can observe the emergence of moderately-sized macroscopic objects that are able to maintain their shape, move through space, and interact with other macroscopic objects. This emergent behavior is especially characteristic of CAs whose transition rules have been designed to place them at the "edge of chaos," Wolfram's class IV (Langton, 1991; Wolfram, 1984).

Therefore we can imagine designing a synthetic world based on CAs in which agents interact with the emergent objects in terms of cell-level interactions. The agents themselves need not be modeled by CAs (although I will consider that possibility below), but their sense organs and effectors must operate at the cell level, for that is the only level of the environment that is real (in the synthetic physics). Higher-level interactions are emergent from these lower level ones, presumably through adaptive processes such as evolution and learning. Just as, according to our interests, we may categorize macroscopic "Life" configurations (e.g., as "gliders") and their group behavior (e.g., as "translation"), so we can expect synthetic agents to develop categories of objects, relations, and actions that are relevant *to them*. The behavior of the agents may be controlled by non-cellular processes, such as production rules, FSMs, or ANNs. We may construct the physics of our synthetic world so that agents must interact with the environment (or with other agents via the environment) in certain ways (defined in terms of elementary cell properties) in order to survive and reproduce.

It might be argued that CAs are not much like the physics of the natural world (although Wolfram has argued the contrary), but that is not important. Again, our goal in synthetic ethology is not to simulate any specific natural system. So for studying the emergence of meaning it does not matter that the CA does not model the fundamental physical processes of our world, or that the emergent objects do not correspond with macroscopic objects in our world. In the synthetic world within the computer, the CA is a physical environment, in which synthetic agents may behave and evolve.

4.2 Artificial Embodiment

A similar limitation of conventional simulations arises in connection with embodiment. We understand better now the essential role played by embodied interaction with an environment as a foundation for genuine intelligence (see below). Also, the fundamental test of an agent's intelligence is how well it can cope with the natural world (especially its EEA), and it is arguable whether a truly intelligent system can exist in the absence of embodiment. This had led some researchers to conclude that artificial intelligence research should be conducted in the context of autonomous robots operating in the natural world (e.g., Brooks, 1986, 1997; Steels, 1997a, 1997b; Ziemke & Sharkey, 2001). I'll briefly consider the issues.

We agree that autonomous robotics provides the fundamental benchmark of genuine intelligence, but it is a complex and difficult approach, because the building of robots may be slowed by problems of mechanical and electrical engineering and of physical construction that have no direct relevance to artificial intelligence. On the other hand, decades of AI research have shown that investigation of simulated agents in simulated worlds (micro-worlds) is inadequate to address the fundamental issues of embodied intelligence; simulated worlds do not have the complexity, unpredictability, uncertainty, openness, and genuine novelty of the natural world (Dreyfus, 1997).

For example, in a micro-world, if an agent is told **move-to** (23, 488), it can be expected to move to that location; if it tests **ahead** (**rock**) it can be expected to determine reliably whether there is an object of type **rock** in front of it; and if it executes **grasp** (**rock**) it can be expected to be in the state **holding** (**rock**). However, for an autonomous robot moving in a natural environment, all these assumptions are problematic. In attempting to move to a location, it may encounter an obstruction, get stuck, or topple over; it may be difficult to determine if there is an object ahead of it, and if it is a rock; and the attempt to grasp the rock may fail, or the rock may slip out of its grip later. These are among the myriad problems faced by real autonomous robots (and by insects crawling through the undergrowth), which are left out of micro-world simulations. Of course, we can build such hazards into the simulation: randomly distribute some simulated **obstacles**, introduce noise or a probability of misclassification, allow an object to be dropped with some probability, etc. However, the problems that occur will be just those we have built into the simulation; genuine surprises cannot arise, because the modes of failure are predefined, just like the objects, properties, and actions.

In the foregoing I have focused on the problems that are faced by autonomous robots, but that are missing from micro-world simulations. However, autonomous robots may have some advantages compared to their disembodied counterparts. There is a great deal of information that every animal knows implicitly just by virtue of having a physical body.

In both biology and situated, embodied robotics (Brooks, 1997), higher-level faculties are built upon lower-level faculties, and intelligence emerges from the interaction of less intelligent components. The foundation of this pyramid consists of low-level sensory and motor modules in direct, dynamic physical interaction with the real world. As a consequence, such intelligence is always grounded in the real world. Further, low-level sensory-motor competencies, such as the ability to perceive structure in visual and auditory inputs, and the ability to sequence and coordinate motor activities, provide a neurological basis for higher faculties such as language, propositional thought, and planning (see also Moravec, 1984).

Another advantage of embodiment is that by being situated in the real world, a robot can often avoid having a complex internal model of the world; the external world provides the only model necessary (Brooks, 1997). This is one of the ways that simple animals, such as insects, are able to accomplish complex tasks without elaborate mental models (a principle called *stigmergy*; see Camazine, Deneubourg, Franks, Sneyd, Theraulaz, and Bonabeau, 2001, pp. 56–59, ch. 19). As Simon (1969) and others have observed, complex behavior may emerge from a simple agent interacting with a complex environment, and so complex intelligent behavior may arise as much from the interaction of the agent with its environments as from the agent itself (insects, especially social insects, are again an example). Furthermore, as explained in Sec. 2, genuine semiosis may occur in very simple agents with correspondingly simple internal representations.

As a consequence of the foregoing considerations, we are faced with a research dilemma. On the one hand, we realize the inadequacy of micro-world approaches and the necessity of studying intelligence in the context of embodied agents situated in a complex, unpredictable environment. On the other, autonomous robotics research is difficult and expensive, and plagued by many engineering problems only peripherally related to the scientific study of intelligence. Therefore we would like is to bring the relative simplicity of micro-worlds to the investigation of embodied, situated intelligence.

We believe that synthetic ethology may provide such a compromise. The goal is a real but synthetic world (inside the computer) that is simpler than the natural world, but is unpredictable and open, and so can serve as an environment in which genuine intelligence may function. As before, the approach is to construct the synthetic world at the micro level, so relevant objects and behaviors emerge at the macro level; in this way we may expect genuine novelty, open-endedness, unpredictability, and uncertainty. However these objects must include the agents themselves, so that the microscopic physics of the synthetic world is sufficient for everything that takes place in it, agent behavior as well as environmental processes. The result will be a synthetic real world, simple but complete, in which we can investigate the evolution and adaptive functioning of (genuinely) intelligent agents.

It seems likely the CA model previously described could be extended to incorporate the agents. An agent would correspond to a macroscopic configuration of CA cells, which would define the agent's behavior. It is well known that the "Life" CA can be configured to compute (e.g., implementing logic gates or simulating a Turing machine), but we have not decided the best approach to use for synthetic ethology. We do not necessarily need computational universality, and the approach should be as simple as possible, for the sake of experimental control as well as computational efficiency. Further, to ensure that the agents become coupled with their environment, the synthetic world must support the evolution of the population in some form.

5 Conclusions

We have described synthetic ethology, a scientific methodology in which we construct synthetic worlds in which synthetic agents evolve and become coupled to their environment. Such a world is complete — in that it defines all the conditions for the survival and reproduction of the agents — but it is simple, which permits greater experimental control than does the natural world. As a result we can perform experiments relating the mechanisms of behavior to social phenomena in an evolutionary context. We presented several examples of such experiments, in which genuine (i.e., not simulated) meaningful communication evolved in a population of simple agents. The communication was intrinsically meaningful to the agents, but only indirectly meaningful to us, as observers. These experiments demonstrate intrinsic intentionality arising from a transparent mechanism. Finally we discussed the extension of the synthetic ethology paradigm to the problems of structured communications and mental states, complex environments, and embodied intelligence, and suggested one way in which this extension could be accomplished. Indeed, synthetic ethology offers a new tool in a comprehensive research program investigating the neuro-evolutionary basis of cognitive processes.

6 Acknowledgments

The development of synthetic ethology was furthered by the intellectual climate of the Institute for Advanced Studies of the Collegium Budapest, where the author was a Fellow in the summer of 1997. He is especially grateful for stimulating discussions there with Eörs Szathmáry, Jim Hurford, Simon Kirby, Mike Oliphant, Luc Steels, and Terrence Deacon.

7 Bibliography

Austin, J. L. (1975). *How to Do Things with Words* (2nd ed., J. O. Urmson & M. Sbisà Eds.). Cambridge: Harvard Univ. Press.

Blackburn, S. (1994). *The Oxford Dictionary of Philosophy*. Oxford: Oxford Univ. Press.

Braitenberg, V. (1984). *Vehicles: Experiments in Synthetic Psychology*. Cambridge: MIT Press.

Brooks, R. (1986). A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, RA-2, 14–23.

Brooks, R. (1997). Intelligence without representation. In J. Haugeland (Ed.), *Mind de-*

sign II: Philosophy, psychology, artificial intelligence (rev. & enlarged ed., pp. 395–420). Cambridge: MIT Press.

Burghardt, G. M. (1970). Defining ‘communication’. In J. W. Johnston, Jr., D. G. Moulton, & A. Turk (Eds.), *Communication by chemical signals* (pp. 5–18). New York: Appleton-Century-Crofts.

Bynum, W. F., Browne, E. J., & Porter, R. (Eds.). (1981). *Dictionary of the History of Science*. Princeton: Princeton Univ. Press.

Camazine, S., Deneubourg, J.-L., Franks, N. R., Sneyd, J., Theraulaz, G., & Bonabeau, E. (2001). *Self-organization in Biological Systems*. Princeton: Princeton Univ. Press.

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

Cangelosi, A., & Parisi, D. (Eds.) (2001). *Simulating the Evolution of Language*. London: Springer.

Colapietro, V. M. (1993). *Glossary of Semiotics*. New York: Paragon House.

Crumpton, J. J. (1994). *Evolution of Two-symbol Signals by Simulated Organisms*. Master’s thesis. Knoxville: Dept. of Computer Science, University of Tennessee, Knoxville.

Dennett, D. (1987). *The Intentional Stance*. Cambridge: MIT Press.

Dretske, F. (1985). Machines and the mental. In *Proceedings and Addresses of the American Philosophical Association*, 59, 23–33.

Dreyfus, H. L. (1997). From micro-worlds to knowledge representation: AI at an impasse. In J. Haugeland (Ed.), *Mind design II: Philosophy, psychology, artificial intelligence* (rev. & enlarged ed., pp. 143–182). Cambridge: MIT Press.

Dunny, G. M., & Winans, S. C. (Eds.). (1999). *Cell-cell Signaling in Bacteria*. Washington, D.C.: ASM Press.

Gardner, M. (1970). Mathematical games: The fantastic combinations of John Conway’s new solitaire game “Life.” *Scientific American*, 223(4), 120–3.

Gleick, J. (1987). *Chaos: Making a New Science*. New York: Viking.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading: Addison-Wesley.

Gregory, R. L. (Ed.). (1987). *The Oxford Companion to the Mind*. Oxford: Oxford Univ. Press.

Grice, H. P. (1957). Meaning. *Philosophical Review*, 66, 377–88.

Gutenplan, S. (Ed.). (1994). *A Companion to the Philosophy of the Mind*. Oxford: Blackwell.

Harnad, S. (1990). The symbol grounding problem. *Physica D*, 42, 335–346.

Haugeland, J. (Ed.). (1997). *Mind Design II: Philosophy, Psychology, Artificial Intelligence*. Cambridge: MIT Press.

Langton, C. G. (1991). Life at the edge of chaos. In C. G. Langton, C. Taylor, J. D. Farmer & S. Rasmussen (Eds.), *Artificial life II: The second workshop on the synthesis and simulation of living systems* (pp. 41–91). Redwood City: MIT Press.

MacLennan, B. J. (1990). *Evolution of Cooperation in a Population of Simple Machines* (Tech. Rep. CS-90-99). Knoxville: University of Tennessee, Knoxville, Dept. of Computer Science.

MacLennan, B. J. (1992). Synthetic ethology: An approach to the study of communication. In C. G. Langton, C. Taylor, J. D. Farmer & S. Rasmussen (Eds.), *Artificial life II: The second workshop on the synthesis and simulation of living systems* (pp. 631–658). Redwood City: MIT Press.

MacLennan, B. J. (2001). The emergence of communication through synthetic evolution. In V. Honavar, M. Patel & K. Balakrishnan (Eds.), *Advances in evolutionary synthesis of neural systems* (pp. 65–90). Cambridge: MIT Press.

MacLennan, B. J. (2002). Synthetic ethology: A new tool for investigating animal cognition. In M. Bekoff, C. Allen & G. M. Burghardt (Eds.), *The cognitive animal: Empirical and theoretical perspectives on animal cognition* (pp. 151–6). Cambridge: MIT Press.

MacLennan, B. J., & Burghardt, G. M. (1993). Synthetic ethology and the evolution of cooperative communication. *Adaptive Behavior*, 2, 161–188.

MacLennan, B. J., Jerke, N., Stroud, R., & VanHeyningen, M. D. (1990). *Neural Network Models of Cognitive Processes: 1990 Progress Report* (Tech. Rep. CS-90-125). Knoxville: University of Tennessee, Knoxville, Dept. of Computer Science.

Moravec, H. P. (1984). Locomotion, vision and intelligence. In M. Brady & R. Paul (Eds.), *Robotics research: The first international symposium* (pp. 215–224). Cambridge: MIT Press

Morris, C. (1964). *Signification and Significance: A Study of the Relations of Signs and Values*. Cambridge: MIT Press.

Neisser, U. (1976). *Cognition and Reality: Principles and Implications of Cognitive Psychology*. San Francisco: W. H. Freeman.

Noble, J., & Cliff, D. (1996). On simulating the evolution of communication. In P. Maes, M. Mataric, J. A. Meyer, J. Pollack, & S. W. Wilson (Eds.), *From animals to animats: Proceedings of the fourth international conference on simulation of adaptive behavior* (pp. 608–617). Cambridge: MIT Press.

Peirce, C. S. (1931–5). *Collected Papers of Charles Sanders Peirce* (Vols. 1–6, C. Hartshorne & P. Weiss, Eds.). Cambridge: Harvard Univ. Press.

Peirce, C. S. (1955). *Philosophical Writings of Peirce* (Justus Buchler, Ed.). New York: Dover.

Reggia, J. A., Schulz, R., Wilkinson, G. S., & Uriagereka, J. (2001). Conditions enabling the emergence of inter-agent signalling in an artificial world. *Artificial Life*, 7(1), 3–32.

Searle, J. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge Univ. Press.

Simon, H. A. (1969). *The Sciences of the Artificial*. Cambridge: MIT Press.

Steels, L. (1997a). Constructing and sharing perceptual distinctions. In M. van Someran & G. Widmer (Eds.), *Proceedings of the European conference on machine learning* (pp. 4–13). Berlin: Springer-Verlag.

Steels, L. (1997b). The synthetic modeling of language origins. *Evolution of Communication*, 1, 1–34.

Wagner, K., Reggia, J. A., Uriagereka, J., & Wilkinson, G. S. (2003). Progress in the simulation of emergent communication and language. *Adaptive Behavior*, 11(1), 37–69.

Werner, G. M., & Dyer, M. G. (1992). Evolution of communication in artificial organisms. In C. G. Langton, C. Taylor, J. D. Farmer & S. Rasmussen (Eds.), *Artificial life II: The second workshop on the synthesis and simulation of living systems* (pp. 659–687). Redwood City: MIT Press.

Wiley, R. H. (1983). The evolution of communication: Information and manipulation. In T. R. Halliday & P. J. B. Slater (Eds.), *Animal behavior volume 2: Communication* (pp. 156–89). New York: W. H. Freeman.

Wittgenstein, L. (1958). *Philosophical Investigations* (3rd ed.). New York: Macmillan.

Wolfram, S. (1984). Universality and complexity in cellular automata. *Physica D*, 10, 1–35.

Ziemke, T., & Sharkey, N. E. (2001). A stroll through the worlds of robots and animals: Applying Jakob von Uexküll's theory of meaning to adaptive robots and artificial life. *Semiotica*, 134, 701–46.