Noble, J., & Cliff, D. (1996). On simulating the evolution of communication. In Maes, P., Mataric, M., Meyer, J.-A., Pollack, J., & Wilson, S. W. (Eds.), From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior Cambridge, MA. The MIT Press / Bradford Books.

On Simulating the Evolution of Communication

Jason Noble and Dave Cliff
School of Cognitive and Computing Sciences
University of Sussex
BRIGHTON BN1 9QH, U.K.
jasonn@cogs.susx.ac.uk; davec@cogs.susx.ac.uk

Abstract

The prospects for modelling the evolution of communication are considered, including the problem of intentional explanation, and the possibility of grounding simulation work in theoretical biology. The seminal work of MacLennan and Burghardt [16] on the evolution of cooperative communication is described, and their experiment replicated. Our results were broadly similar, in that evolved communication was observed, but specific differences are discussed. MacLennan and Burghardt's work is extended and their methodology critiqued in detail. Their experiment remains a useful demonstration, but artefactual features make their results difficult to interpret. Furthermore, we argue that too many factors are simultaneously investigated for any general principles to be extracted, and suggest an alternative program of narrowly-focused simulations.

1 Modelling Communication

What has to be happening such that we would describe two entities within a computer simulation as communicating? Can communication behaviour evolve in simulation? Do we learn anything about animal communication or the development of language from such work? MacLennan and Burghardt [16, p. 186] conclude that "even in [a] simple synthetic world, communication may evolve that exhibits some of the richness of natural communication." The intent of this paper is to investigate the prospects for modelling communication through a critical replication of their research.

1.1 Definitions

Communication can be a slippery subject, partly due to the enormous variety of signalling behaviours in nature. Looking at intra-species communication alone we find aggregational signals, alarm signals, food signals, territorial and aggressive signals, appeasement signals, courtship and mating signals, and signalling between parents and offspring [14] — the list is not exhaustive. It can be difficult to capture all this under one definition, or under

one explanatory story. Krebs and Dawkins [13] suggest that communication arises because all animals need to be good at predicting the behaviour of other animals in their environment, and that such mind-reading abilities can be exploited, through signalling, to manipulate the behaviour of the receiver. That is, communication evolves when it is beneficial to have one's behavioural intentions predicted (or falsely predicted). Some authors [7] have suggested that this places too much emphasis on dishonest communication, and a typical definition from the biology literature [14] is as follows:

> ...the transmission of a signal or signals between two or more organisms where selection has favoured both the production and the reception of the signal(s).

Note that the role of natural selection is central, and assumed as a premise (see sections 1.4 and 2.1).

1.2 Intentionality

The difficulty of defining communication leads us to the problem of intentionality. In cases of communication between animals or simulated animals, to what extent should we talk about the sender and the receiver as rational, intentional agents? Intentional (and teleological) explanations have always been risky practice in science. However, even the most rigorous accounts of communication (e.g. Shannon and Weaver [21]) are suggestive of intentionality in both senses: an agent who *intends* to communicate something to another agent, via a signal that *means* something. If we consider a phenomenon such as predator alarm calls in vervet monkeys [20], we are a long way from being able to explain this at the level of neurology, and unimaginably far from being able to explain it at the level of physics. However, we can talk usefully about what a monkey intended to achieve by calling, and what a particular call means — indeed, Harré [11] describes primatologists who eschew such intentional language, in the name of objectivity, but end up surreptitiously using it anyway.

Dennett [5] argues that we should not resist the temptation to adopt "the intentional stance" towards complex systems like vervet monkeys. He suggests that this is an

example of a more general strategy in science of "changing levels of explanation and description in order to gain access to greater predictive power or generality" [5, p. 239], and compares it to the abstraction of the concept of *food* in biology. The underlying details of physics, chemistry or whatever are legitimately ignored because they do not contribute to the usefulness of explanatory accounts at the higher level.

So where do we draw the line? Can we ascribe intentions and meaning to just anything? Dennett suggests that there will be no clear division between cases where the intentional stance is appropriate and cases where it is not. As we move to less and less complex systems, mechanistic accounts of their function will become progressively more plausible.

We are not making the bold claim that *animals have intentions*. There is insufficient evidence at this time to support either that claim or its negation. What we are saying is that the concept of communication is an intentional-level concept, and that therefore any attempt to investigate communication without using an intentional framework will be incoherent.

1.3 Work in theoretical biology, ethology, and behavioural ecology

There exists an extensive empirical and theoretical literature on animal communication. Empirical work tends to describe a particular type of signalling within one species or between two species (see e.g. [10, 13, 14] for reviews). Some of the most promising theoretical approaches (e.g. Maynard Smith [17]) involve positing a functional model of a signalling system and investigating the conditions under which that system would be evolutionarily stable. The argument is that we should only expect to find relatively stable systems in nature, and that if we are confronted with a system that appears to be otherwise, we should expect to find stability under the surface.

An example of this sort of work is that of Zahavi [25, 26], later validated mathematically by Grafen [7]. They argue that signalling in sexual selection will tend to be honest, and the signals costly. Assume that males signal their reproductive fitness to females through some phenotypic trait: whereas we might expect males to exaggerate their quality, females will be selected on the basis of their ability to discriminate between high and low quality males. The evolutionarily stable situation will be one where a signal is expensive for the male to produce (e.g. the peacock's tail). Thus deceptive communication becomes impossible: if the male can afford to display that signal, he really is of high quality. Through logical and mathematical argument, Zahavi's counter-intuitive idea — that sexual selection can be selection for a handicap — is established.

1.4 Grounding simulation

We concur with Miller [19] that such work in theoretical biology and related fields is the best starting point for those who wish to model communication and other biological phenomena *in silico*. While the simulation of adaptive behaviour (SAB) and artificial life (AL) research communities have had prima facie success at simulating evolution, or, if you prefer, generating real evolution inside a computer, it is not yet clear that their approach is a reliable route to general laws or principles concerning animal behaviour. Too often, SAB/AL work falls into the trap of circularity: a simulation is tailored to evolve a specific phenomenon, and that very phenomenon is reported as an unexpected and interesting result.

The notion that evolution is a kind of optimising or satisficing process is not really under debate (*pace* Gould and Lewontin [6]), and SAB/AL work should not be in the business of merely showing that evolution works. In all but the most open-ended simulations, the optimising principle is built in anyway: high scorers on a pre-given fitness function are more likely to reproduce. If SAB/AL is to go forward as a method of scientific investigation, "proof of concept" displays must give way to research that looks for structures underlying the observed phenomena. As Maynard Smith [18] says of the theoretical biology literature, "in using optimisation, we are not trying to confirm (or refute) the hypothesis that animals always optimise; we are trying to understand the selective forces that shaped their behaviour."

There are some recent examples of SAB/AL work on communication that start with a model or theory from biology, perhaps expressed mathematically, then validate and extend that model using iterative, computational techniques. For example, de Bourcier and Wheeler [4] look at aggressive signalling and territoriality. They state that their method of *synthetic behavioural ecology* "is pitched at an intermediate level between, on the one hand, abstract theories based on mathematical models and, on the other hand, empirical observations in complex environments" [4, p. 464].

The two seminal examples of attempts to evolve communication, however, are Werner and Dyer [22] — who simulated the evolution of a simple communication protocol that allowed immobile females to guide blind males towards them for mating — and MacLennan and Burghardt [16]. We feel that both of these papers are worthy of careful reconsideration, but for reasons of space we concentrate entirely on the latter.

2 MacLennan and Burghardt's experiment

2.1 Justification

MacLennan and Burghardt describe their method as *synthetic ethology*, contrasting it explicitly with simulation. They state that:

> Our goal in these experiments was to design a synthetic world that was as simple as possible while still permitting communication to evolve. [16, p. 165]

MacLennan and Burghardt repeatedly emphasise that their "synthetic world" is not supposed to reflect any real environment, nor are their simulated organisms like any actual species. Inspired by the synthetic psychology of Braitenberg [2], they hoped that, in comparison with empirical ethology, their stripped-down approach would be "more likely to suggest behavioral laws of great generality" [16, p. 163].

MacLennan and Burghardt were aware of the difficulty of defining communication, and of the problem of imputing intentionality. They adopted Burghardt's [3] definition of communication, which "finessed the issue of intent by the requirement that the behavior be likely to influence the receiver in a way that benefits, in a probabilistic manner, the signaler or some group of which it is a member" [16, p. 163].

They chose to investigate cooperative communication. Possibly this choice was influenced by Burghardt's definition, in that cooperative communication generally benefits both the sender *and* the group to which it belongs. They reasoned that for communication to be selected for, some of the simulated organisms must have access to information that the others in the group did not — otherwise communication would be unnecessary. The non-shared information must also be of environmental significance; it must be worth talking about. In line with their definition of communication, they designed the synthetic world such that communicating this non-shared information would tend to confer a selective advantage.

2.2 Method

MacLennan and Burghardt[1] used populations of simulated organisms that they refer to as "simorgs". The simorgs all have access to a shared global environment, and each individual has access to a private local environment. The global environment provides a medium for communication, and the local environments are a source of significant information that the simorgs may

---

[1] MacLennan and Burghardt's methodology is difficult to describe briefly, and the reader is referred to their work [16] and MacLennan's earlier article [15] for a complete account.

---

evolve to communicate *about*. Each of the environments is as simple as possible, represented by a single variable that can take on a finite number of values. It is emphasised that "there are no geometrical relations among [the simorgs]...they are not in a rectangular grid, nor are some closer than others" [16, p. 166].

MacLennan and Burghardt suggest, by way of analogy, that the global environment can be thought of as the air, capable of transmitting only one sound at a time, and the local environments can be considered exclusive hunting grounds, into which different species of prey may wander. In other words, states of the global environment have the potential to be exploited as signals, and states of the local environment are particular circumstances that it will pay simorgs to signal about.

Simorgs have only two classes of behavioural choice open to them: they can *emit* a signal (into the global environment), or they can *act* in an attempt to respond to the signal of another. The state of the global environment can be changed by any of the simorgs if that simorg emits a signal when its turn comes; the states of the simorgs' local environments are not under their control, and are periodically reset to random states.

In the synthetic world, simorgs achieve fitness by successfully cooperating with another simorg. Specifically, by responding to a signal with an action that matches the local environment state of the signaller. When this occurs, both the signaller and the respondent are rewarded with a point of fitness. Continuing their analogy, MacLennan and Burghardt suggest that this is to be regarded as two hunters bringing down a prey animal that neither could bag alone. Assuming that successful communication has taken place, note that the signal does not mean "I've got some prey here", but "I've got prey of type λ here; would you mind helping out with action-λ?" The state of another simorg's local environment is not directly knowable, and successful cooperation can only come about through a lucky guess or the employment of communication.

In order to implement their ideas in a computer program, MacLennan and Burghardt had to make a number of somewhat arbitrary practical decisions. Thus, time in the synthetic world is discrete. Once each time step, the simorgs respond (i.e., act or emit) in a fixed order; effectively they are arranged in a ring. The program keeps track of the "owner" of the symbol currently occupying the global environment. It is possible, for example, for one simorg to emit and then earn several fitness points consecutively as a series of other simorgs act in response to the same persistent signal.

Every five time steps (one environment cycle) the local environments are reset to a random value, ensuring that the simorgs must react to changing circumstances if they are to succeed. Every fifty time steps there is a breeding cycle: two fit simorgs are stochastically se-

lected as parents and, using two-point crossover with a small chance of mutation, a new simorg is generated. An unfit simorg is stochastically selected to be replaced by the child, keeping the population size constant. This arrangement is akin to the steady-state variety of genetic algorithms.

The experiments reported were run for 5000 breeding cycles, populations were of size 100, there were eight local environment states (L) and eight global environment states (G) — "just enough possible sounds to describe the possible situations" [16, p. 175] — and the mutation rate was a 0.01 probability of one mutated allelle per birth.

Finite state machines (FSMs) serve as the internal architecture of the simorgs. MacLennan and Burghardt could have used any number of architectures, and considered using neural networks, but settled on FSMs because they "are both readily understood intuitively and easy to represent in genetic strings for simulated evolution" [16, p. 167]. In the experiment described[2] the FSMs were of only one state, which reduces to a look-up table. The response a simorg would make at any one time step was completely determined by the state of the global environment and the state of its local environment. The content of each of the 64 (8×8) entries of the look-up table was a flag indicating *act* or *emit*, and an integer representing the action-type or the emitted symbol respectively. The genetic coding of the simorg was a direct mapping of this structure.

MacLennan and Burghardt included in the program a mechanism to (optionally) prevent communication from occurring: the global environment could be overwritten with a random symbol after the response of each simorg. They reasoned that if fitness increased more rapidly when communication was permitted, compared with when it was blocked, then "true communication (involving a sender)" [16, p. 172] was taking place. In a similar fashion they were interested in exploring the effect of a simple learning rule, whereby a simorg that makes an incorrect action (i.e. an action that does not correspond to the local environment state of the last emitter) in response to a signal has the appropriate entry in its look-up table altered so that it *would have* given the correct response. Thus, they report the results of subjecting the same randomly generated initial population to each of the following experimental conditions:

$C^- L^-$ communication blocked and learning disabled;

$C^+ L^-$ communication permitted and learning disabled;

$C^+ L^+$ communication permitted and learning enabled.

---

[2]MacLennan and Burghardt actually conducted two experiments; we focus entirely on the first. Experiment 2 was an attempt to evolve multiple-symbol communication and the results led them to conclude that "making the step to multiple-symbol syntax is evolutionarily hard" [16, p. 183].

In each of the conditions, they collected data on mean fitness over time. They also constructed a "denotation matrix", which recorded the number of successful communication events, arranged in a table by local and global environment states. They found that these matrices were most useful when tallied over the last 50 breeding cycles of a 5000-cycle experimental run. Under these circumstances, the matrix was interpreted by MacLennan and Burghardt as describing the evolved language of the simorgs. The degree of structure present in the matrix was indexed by co-efficient of variation and entropy statistics.

2.3 Results and conclusions

MacLennan and Burghardt report that communication did indeed evolve in the synthetic world. The results reported are for a single random initial population subjected to each of the three conditions; MacLennan and Burghardt assure us that these results are typical. In the $C^- L^-$ condition, there was only a very slight increase in fitness over the length of an experimental run, whereas in the $C^+ L^-$ condition the rate of fitness increase was an order of magnitude greater. In the $C^+ L^+$ condition, the rate of fitness increase was higher still. MacLennan and Burghardt conclude that, when it is not suppressed, communication is selected for and leads to higher levels of cooperation. The provision of the single case learning rule further increases the effectiveness of the communicative strategy.

Analyses of the denotation matrices showed that in the $C^- L^-$ condition, the pattern of symbol use was almost random. When communication was permitted the matrices were quite structured, as measured by the entropy statistic. Visual inspection of the denotation matrices made it clear that certain symbols had evolved to (almost uniquely) represent certain local states. There was ambiguity in two senses: sometimes a symbol would be used to represent two or more states, and sometimes a state was represented by two or more symbols. MacLennnan and Burghardt suggest that the ambiguity is either due to two subpopulations using different symbol dialects, or to individual simorgs using one symbol to represent two different states.

That there should be any fitness increase at all in the $C^- L^-$ condition is not obvious. MacLennan and Burghardt refer to this phenomenon as "partial cooperation through co-adaptation", and regard it as a "low-level effect" [16, p. 185]. They explain it by noting that simorgs can do better than chance if they emit a symbol only in a subset of their local situations, and guess actions within that same subset (see section 4.4 for details).

	MacLennan & Burghardt	Our results Mean	SD	p
Fitness increase				
C^-L^-	0.37	0.99	1.16	n.s.
C^+L^-	9.72	14.6	6.54	n.s.
C^+L^+	37.1	10.6	10.6	0.025
Final mean fitness				
C^-L^-	≈ 6.6	6.74	0.43	n.s.
C^+L^-	10.28	12.71	2.68	n.s.
C^+L^+	59.84	46.13	4.02	0.004

Table 1: Rate of fitness increase (determined by linear regression and measured in units $\times 10^{-4}$ breeding cycles) and final mean fitness scores. Note that mean fitness data was a moving average smoothed over 50 breeding cycles, and that final mean fitness in the C^+L^+ condition is much higher because the simorgs had four chances per environment cycle to respond after correction by the learning rule: fitness scores in this condition start at 40+ rather than the usual chance level of 6.25. Rates of increase are thus a better comparison across conditions.

3 Replication of MacLennan and Burghardt

We replicated MacLennan and Burghardt's experiment, writing our own code[3] based on the published descriptions of their procedure [15, 16].

The replication gave qualitatively similar results, in that fitness improved over time when communication was enabled, and structure developed in the denotation matrices, but the specific results in the three conditions were not reproduced. Table 1 contrasts MacLennan and Burghardt's results with our own; the rate of fitness increase per 10^4 breeding cycles and the mean final fitness are shown. MacLennan and Burghardt's results are taken directly from [16], and refer to the single run they presented as the typical case. Our own results show the mean and standard deviation across 20 runs with different random seed values. For each condition, the column labelled "p" shows the statistical significance of a two-sample t-test of the null hypothesis that MacLennan and Burghardt's result could have come from the same distribution as our data ("n.s." means not significant, i.e. $p > 0.05$).

The C^-L^- and C^+L^- conditions showed slightly higher rates of fitness increase in our own experiment. More importantly, the rate of fitness increase in the C^+L^+ condition was more than three times *smaller* in

---

[3]Each author worked independently, to allow cross-checking. Both authors used the C language; JN's version was approx. 1500 lines long (including code for various statistics not mentioned here), and DC's version was 560 lines long. A 5000 cycle C^+L^- condition ran in about 80 seconds on a Sun Sparc 20 (both versions). All source code is available on request.

	MacLennan & Burghardt	Our results Mean	SD	p
C^-L^-	5.66	4.96	0.15	< 0.001
C^+L^-	3.95	3.36	0.50	n.s.
C^+L^+	3.47	4.45	0.36	0.015

Table 2: Entropy statistics, calculated on the denotation matrix of the final 50 breeding cycles of the experiment. An entropy value of 6 would indicate a completely random matrix. A value of 3 indicates a perfectly structured matrix, with one symbol per situation.

our data than in MacLennan and Burghardt's, and this was statistically significant. Our results do not support their finding that the C^+L^+ condition, i.e. communication with learning, leads to the highest rate of fitness increase. We found communication with learning to be inferior to communication alone.

Table 2 shows the entropy of the denotation matrices over the last 50 breeding cycles of the experimental runs. Again, MacLennan and Burghardt's figures are taken directly from [16], and our own figures summarise 20 different runs. In the C^-L^- condition we found significantly *more* structure to the denotation matrices than did MacLennan and Burghardt, and in the C^+L^+ condition we found significantly *less*. Instead of the lowest entropy being associated with C^+L^+, we find it to be associated with C^+L^-. In other words, the most structured communication conventions develop in the communication only condition, and the addition of the learning rule only reduces that structure.

The differences between our findings and those of MacLennan and Burghardt should not be exaggerated. In all measurements, across all conditions, our figure was well within an order of magnitude of MacLennan and Burghardt's figure. Our interpretation of their experimental method may not reflect *exactly* their actual procedure, but at this point in time we have been unable to locate the source of the discrepancy. MacLennan and Burghardt's central result was successfully replicated: that communication, when enabled, leads to relatively high rates of fitness increase, and to the evolution of a structured "language" as evidenced by the denotation matrix.

4 Extension and critique

Having described the methods used by MacLennan and Burghardt, and noted the degree to which our results match theirs, we now wish to comment critically on certain aspects of their experiment. Several questions are raised as to what might be an appropriate methodology for studying the evolution of communication, and we hope to answer these questions in section 5.

	Mean	SD	Effect
Fitness increase			
C^-L^-	0.94	1.52	−4.5%
C^+L^-	18.6	7.05	+27.4%
C^+L^+	33.7	13.8	+218%
Final mean fitness			
C^-L^-	6.76	0.53	+0.23%
C^+L^-	14.47	2.83	+13.9%
C^+L^+	22.24	5.21	−51.8%

Table 3: Effect of random-order updating. Rate of fitness increase $\times 10^{-4}$ breeding cycles (determined by linear regression), and final mean fitness scores are shown, with means and standard deviations across 20 runs. The "effect" column compares the random-order results with our standard updating results (see table 1); note that if the updating method was not influencing the results, we would expect this value to be close to zero.

4.1 No geometry?

MacLennan and Burghardt claim that there are "no geometrical relations" [16, p. 166] among the simorgs. This is in keeping with their goal of constructing a synthetic world that is as simple as possible while still permitting communication to evolve. If the simorgs were arranged on a toroidal grid and could communicate only locally, for example, this would certainly complicate things.

However, in the current set-up, the simorgs are effectively arranged in a ring. As MacLennan and Burghardt [16, p. 170] put it, "The simorgs react one at a time in a fixed order determined by their position in a table." Thus there is at least a topology, if not a geometry: simorgs will tend to receive signals from their immediate neighbours in one direction, and send signals to their neighbours in the other direction.

The experiment could have been performed without this modest topological assumption if the simorgs were updated in a different random order at each time step. We modified our version of the program to use just such an updating procedure. Table 3 shows the rates of fitness increase and final fitness scores under this method.

There is a dramatic difference between the two updating methods. In the communication only (C^+L^-) and no communication (C^-L^-) conditions, similar performance is observed under both updating methods. The effect of the learning rule, on the other hand, depends very much on the updating method used: under random-order updating, the rate of fitness increase is much higher. Curiously, the rates of fitness increase under random-order updating come closer to the rates observed by MacLennan and Burghardt — perhaps this is a clue as to the cause of our differing findings.

Furthermore, random-order updating clears up an irksome feature of MacLennan and Burghardt's results. Fitness in the learning condition commences close to the random level of 6.25 (see the notes to table 1), which makes mean fitness directly comparable with the other conditions — note the 51.8% drop in final fitness scores. Under standard updating, a simorg will often have its look-up table corrected on the first time step of an environment cycle, then find itself in exactly the same context on the next four time steps, and score up to four "free hits". When simorgs are responding in a different random order each time step, it is no longer the case that a simorg will be communicating with the same near neighbours every time, and the learning rule loses this bonus property.

The most important point about the random updating procedure, however, is that it demonstrates that MacLennan and Burghardt's results could be dependent upon such apparently minor assumptions built in to their procedure. Their goal is to uncover general laws that can be translated back into the realm of real biology, but if the effect of learning on the evolution of communication is dependent on the updating method used, it is difficult to know what biological conclusions should be drawn. Does learning facilitate the development of a communicative system, or doesn't it?

4.2 Dialects or sub-optimal look-up tables?

MacLennan and Burghardt, noting the ambiguous symbol use evident in the denotation matrices, comment that "we cannot tell from [the denotation matrix] whether this multiple use of symbols results from two subpopulations or from individual simorgs using the symbol to denote two situations." [16, p. 179]. The idea that there could be subpopulations using different dialects seems quite plausible, especially given that the topology of the simorgs' environment (see section 4.1) ensures that simorgs will only be communicating with near neighbours. One can imagine a series of simorgs using variant A in one section of the ring, shading gradually into variant B in the opposite section, and back again.

MacLennan and Burghardt claim that the facts of the matter could easily be uncovered: given that the underlying finite state machines are available in computer memory, "there need be no mystery about how the simorgs are communicating, because the process is completely transparent." [16, p. 179]. However, they make no clear statement as to whether they in fact believe there are two or more subpopulations using variants of the evolved "language". MacLennan, in his earlier paper, is less conservative: "the differing use of symbols in various contexts makes it quite possible for every simorg to be using a different dialect of the 'language' manifest in the denotation matrix." [15, p. 653].

In an attempt to resolve this question, we used a con-

vergence statistic in our experiments. We examined each position on the genome in turn, and calculated the mean percentage of identical entries across the population of simorgs. Thus, a convergence statistic of 100% would indicate a population of simorgs with identical genomes and, thus, identical FSMs.

In runs of 5000 breeding cycles duration, the final convergence statistic was typically between 75% and 85%. This is not conclusive: it means that up to 25% of the simorgs could have been different from the norm, or that 25% of the genetic material of each simorg could be unique, and so leaves plenty of room for the possibility of different dialects. However, when the runs were extended to 2×10^4 breeding cycles or more, final convergence statistics in the C^+L^- condition were approximately 99.5%, and denotation matrices were qualitatively similar, i.e. they still showed ambiguous communication. It is implausible to suggest that there might be different dialects when the simorgs in a population are 99.5% identical to each other. We conclude that the suggestive ambiguity in the denotation matrices is nothing more than the net effect of (more or less) the whole population using a single, inefficient "language" that sometimes represents a state by more than one symbol, or uses one symbol to denote more than one state.

Despite their stated wariness about adopting any sort of intentional stance towards the simorgs [16, p. 163], we believe that MacLennan and Burghardt are not immune to the temptation to think of the simorgs as intentional agents, and that, in this case, that temptation has led them astray. Language without the scare quotes is undoubtedly the exclusive province of sophisticated intentional agents (argued in e.g. [1, 5]), but having drawn the analogy between human language and simorg communication, MacLennan and Burghardt were too ready to suspect that, like real language users, simorgs might have dialects. We are not claiming that they were wrong to draw the analogy in the first place; indeed, a central premise of adopting the intentional stance is to take such analogies very seriously. However, in dealing with simple systems like the simorgs, lower level mechanistic explanations are potentially open to us. The intentional story (dialects) can be shown, via the convergence statistic, to be inferior to the mechanistic story (less than ideal structure in the FSM look-up tables).

4.3 Consequences of the FSM look-up table approach

Imagine for a moment that you are a simorg. Disregarding the fact that simorg "decisions" are entirely determined by the look-up table, imagine that you have decided to emit a symbol. The only context that is important to you is the state of your local environment: you need to choose the right symbol to describe your situa-

tion, according to the "language" conventions that have developed. The identity of the symbol in the global environment is unimportant, because you're going to overwrite it anyway.

Similarly, if you're going to act, you don't care about the state of your local environment; you only want to interpret the global symbol in such a way as to correctly match the environment of the last emitter, and thereby score a point of fitness.

For the real simorgs of MacLennan and Burghardt, things are not this simple. There is no prior decision to emit or to act, only the consultation of a table with an entry for every possible combination of local and global environment states. As MacLennan and Burghardt put it, "finite-state machines have a rule for every possible condition." [16, p. 168].

Surprisingly, this means that the choice of the FSM architecture makes evolving a communication system harder for the simorgs than it might be under some other control architectures. For example, if during a particular run it became advantageous to reliably perform action-2 in response to symbol-7, FSM-controlled simorgs would have to ensure — through evolution or learning — that eight distinct entries in their look-up table came to be identical. That is, they would need to perform action-2 in response to symbol-7 in the context of eight different possible local environment states. By contrast, a simorg that was controlled by, for example, a classifier system (see e.g. [12, 24]) would need only to generate a single production rule: "perform action-2 in response to symbol-7". We have not yet investigated this empirically, but we offer as a hypothesis our suspicion that simorgs controlled by classifier systems would evolve significantly faster than FSM-controlled simorgs.

MacLennan and Burghardt did not believe that FSMs were the only architecture open to them, and adopted them for pragmatic reasons. However, if an arbitrary choice of control architecture is influencing their results in unexpected ways, it is again difficult to see how their conclusions can be reliably translated back to biology.

4.4 Counter-intuitive optimal strategy

The optimal strategy for the simorgs, at least at the population level, must be to act as often as possible, and to emit as infrequently as possible. This is because emitting scores no fitness points directly.

The best way for the simorgs to achieve this is to build up a link between a *single* global symbol γ and a *single* local state λ. A situation develops where simorgs always "blindly" respond with action-λ, unless they are in state λ themselves, in which case they emit γ. Imagine all the simorgs acting in this way: it is clear that they would no longer need to be concerned about the particular identity of the symbol in the global environment. They would *know* that it will always be a γ, and that it will always

reliably indicate that the last emitter (whoever that may be) is in state λ.

Assuming 8 symbols and 8 local states, this means that a successful communication will take place 7/8 of the time[4]. This would translate as a mean fitness of 87.5 — a very high value relative to the results presented in tables 1 and 3. In general the maximum fitness will be equal to $100 \times \frac{L-1}{L}$, where L is the number of local states. To date we have only seen this phenomenon evolve spontaneously when $L \leq G \leq 4$, but the principle remains.

The trouble with this result is that one presumably does not want to call it an evolved communication system or "language", even though the simorgs are ostensibly fitter than ever before. If the global environment is (almost) always in the same state, it is difficult to describe it as carrying any information. The simorgs in such a situation appear to be exploiting a loophole in the experimental design.

MacLennan and Burghardt were aware of this possibility (see section 2.3). They saw it as most relevant to the C^-L^- condition, in that it provided an explanation for the otherwise mysterious increase in fitness observed. MacLennan [15, p. 653] felt that "in most cases [it] is a low level effect that is unintrusive and can be ignored".

4.5 Fewer symbols: faster improvement

The point outlined in 4.4 has a number of implications. Given that the optimal strategy involves the utilisation of only one symbol, we hypothesised that giving the simorgs progressively *fewer* symbols to work with would steer them towards that strategy and thus *improve* their performance. This contrasts with the intuitive hypothesis that n local states will require simorgs to use n symbols to denote them. MacLennan and Burghardt seem to have assumed the truth of the intuitive hypothesis: they speak of the ideal denotation matrix as having one symbol to denote each situation, and refer to the fact that $L = G$ as meaning that "there were just enough possible sounds to describe the possible situations." [16, p. 175].

To test our hypothesis we used the C^+L^- condition, held the number of local environment states constant at $L = 8$, and varied the number of global environment states G, i.e. the number of possible symbols, from eight down to one. For each case, 20 runs of length 5000 breeding cycles were conducted. The results are shown in figure 1.

Overall, higher rates of fitness increase were associated with smaller numbers of symbols. These results certainly contradict the intuitive view. In an effort to

---

[4]Discounting for a moment the unfortunate simorg who acts in response to an out of date symbol immediately after the local environments have been randomly reset; in the "experiential world" of the simorgs, this is an infrequent event.
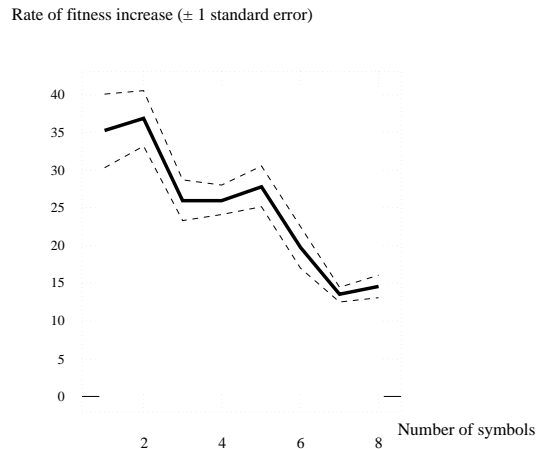

Figure 1: Mean effect on rate of fitness increase of varying G, the number of global symbols, while $L = 8$. Rate of fitness increase measured in units $\times 10^{-4}$ breeding cycles (determined by linear regression).

make a connection with real-world biology, one might argue that we are demonstrating the principle, described by Wiley [23], that small signal repertoires enhance the detectability of ritualised signals. The argument would be without merit, however. Wiley discusses the value of a small signal repertoire as a means of enhancing detection *in a noisy environment*. The simorgs' environment has no noise, and their perception of symbols is direct, immediate and reliable. Again, we claim that there is no easy biological translation of this observation concerning MacLennan and Burghardt's synthetic world.

4.6 Symbol use over time

If the simorgs were evolving a "language", with an eventual one-to-one correspondence between global symbols and local states, we should observe a fairly even distribution of the G symbols. That is, simorgs should use each symbol about equally often. This follows from the fact that the distribution of local states is random and therefore uniform.

This is not observed, however. Popular symbols tend to get more popular; in the 20 runs of the C^+L^- condition reported in table 1, the mean usage of the most popular symbol at the end of the run was 41.45%. When longer runs were conducted, the popularity of the most popular symbol was even higher. Infrequently used symbols often dropped out of use altogether. This seems to indicate that the simorgs are drifting towards the optimal strategy described in 4.4.

5   Communication reconsidered

MacLennan and Burghardt clearly succeed in establishing that communication can occur in a particular simulated environment. They express the hope that their work will suggest general laws or principles concerning animal communication, but they are aware that "if the synthetic world is too alien, we may doubt the applicability to our world of any observations made of the former." [16, p. 166]. In section 4, we have pointed out various ways in which their synthetic world is indeed alien. Regrettably, it *is* difficult to see how certain aspects of MacLennan and Burghardt's results could be translated into real-world biology. However, we are optimistic about the usefulness of evolutionary simulations generally, and we would argue that the central problem with their work is simply that it tries to do too much, too soon.

There is not an established body of literature on simulating the evolution of communication. As things stand, communication is just one of the many biological phenomena that have come under attention from those pursuing the SAB/AL programme. We think it is safe to say that there are not yet any agreed-upon methods or landmark findings; there is only the central SAB/AL premise that evolution can be captured in a computer program, and a resolution to use ideas from biology, ethology, behavioural ecology, and signalling theory. In this context we believe that, despite its apparently simple nature, MacLennan and Burghardt's work is overly complex and ambitious.

MacLennan and Burghardt are trying to do a number of things at once. Primarily, they are attempting to provide an existence proof for the synthetic evolution of communication, and they make no secret of having constructed the synthetic world so that the simorgs will be likely to reproduce only if they cooperate (i.e. communicate) in the specified way. They are also examining a process by which arbitrary symbols can evolve to denote something in a simple "language". As they put it, "beyond merely detecting the presence of communication, we are also interested in studying its structure." [16, p. 173]. Further, because the simorgs must come to know not only the correlations between symbols and local states, but also when to act and when to emit, MacLennan and Burghardt are effectively looking at the evolution of turn-taking. Finally, they are interested in the effect of learning on the evolution of communication.

With the possible exception of the basic existence proof, each of these phenomena are poorly understood, and each is worthy of a separate, narrowly-focused simulation experiment. When all of these questions of interest are thrown in together, they interfere with each other and make the extraction of general principles impossible. For instance, in trying to push the simulation towards communication, they choose to reward both the sender and the receiver of a message, and in an effort to leave things open-ended enough for spontaneous symbol-meanings to develop, they use the FSM architecture. But what is the relative importance of these factors in causing the observed results? MacLennan and Burghardt allow spontaneous strategies for emitting vs. acting to develop amongst the simorgs, presumably to leave them as unconstrained as possible, but this decision creates the loophole described in section 4.4. Would the same type of communication develop if the simorgs were constrained to be senders and then receivers in turn?

In principle, it may be that communication between simorgs is entirely dependent on their internal architecture, or on the fitness reward structure used, or some other quirk of the methodology — MacLennan and Burghardt themselves note that when the method for selecting parents was deterministic rather than stochastic, communication did not develop. It is not possible, from MacLennan and Burghardt's results alone, to determine any necessary or sufficient conditions for the evolution of communication; they are doing the equivalent of commencing the study of gravitation with a four- or five-body problem.

Of course, we are not claiming that if only the various factors bearing upon the behaviour of MacLennan and Burghardt's simorgs could be isolated, then the general principles governing naturally evolved communication would be laid bare. It is quite likely that there are complicated, non-linear interactions even in their small system. However, if we do not understand the effect of each factor alone (e.g. cost or benefit of communication, updating method, simorg architecture) then it would seem optimistic to hope to understand the complex case.

The difficulties with MacLennan and Burghardt's experiment can be seen in another light: they compare synthetic ethology favourably with empirical ethology in that experiments in the former are repeatable, and full access to all variables is possible. However, this comes at a price. MacLennan and Burghardt are forced to rigorously specify the environment and the internal nature of the simorgs, making several ad hoc decisions along the way. In a sense, they have to go down to the level of simorg genetics. This is interesting, because one of the great strengths of ethology comes from what Grafen calls the "phenotypic gambit" [8, p. 6], in which genetics is almost entirely abstracted away, and broad behavioural strategies are considered at a functional level [5]. Most of the time, the conclusions so derived are borne out in the real world. The parallel to be drawn with MacLennan and Burghardt's experiment is that there is much to be done, using simulation methods, that does not buy into

---

[5]Grafen was specifically discussing behavioural ecology, the offshoot of ethology that concentrates on Tinbergen's third question: "What is this behaviour *for*?"

the question of internal architectures, but looks at one phenotypic characteristic at a time and assesses its effect on the evolution of communication. For example, one could simulate a population of agents who were either communicators or mutes, and then allow that population to evolve under different cost and benefit regimes for communicative behaviour. We might expect that when both the sending and the receiving agent benefit from communicative behaviour, then communicators will come to dominate the population. But what about when only the receiver benefits, or when the sender's benefit is relatively small? What happens when communicators will only signal to other communicators? This sort of simulation, taking up where the mathematical arguments of biologists such as Hamilton [9] and Grafen [7] leave off, would give us a sound basis for further investigations.

The best philosophical strategy for such future work is to adopt the intentional stance with respect to simulated organisms. Despite the fact that MacLennan and Burghardt at one point go too far, in ascribing high-level intentional phenomena such as language dialects to the simorgs (see section 4.2), we agree with Dennett [5, p. 265] about intentional accounts: "...in a nutshell, *they work*. Not always, but gratifyingly often." We are also confident that mechanistic explanations can peacefully co-exist with intentional ones; in the very simple simulations we are initially proposing, no doubt mechanistic accounts will predominate, with the balance gradually shifting as real-world complexity is incrementally approached.

MacLennan and Burghardt are at pains to *avoid* intentional talk when they define communication, and MacLennan [15] criticises denotational (i.e. intentional) theories of meaning. Nevertheless they rely on an analogy featuring rational, intentional agents — the story of the hunters — to make sense of their simulation, and they use denotation matrices to index the meaning of symbols: an intentional technique, in the sense of "aboutness", if ever there was one. We contend that to seek a non-intentional account of communication is to seek an oxymoron.

References

[1] J. Bennett. *Linguistic Behaviour.* Cambridge University Press, 1976.

[2] V. Braitenberg. *Vehicles: Experiments in Synthetic Psychology.* MIT Press, Cambridge, MA, 1984.

[3] G.M. Burghardt. Defining 'communication'. In J.W. Johnston, Jr., D.G. Moulton, and A. Turk, editors, *Communication by Chemical Signals.* Appleton-Century-Crofts, New York, 1970.

[4] P. de Bourcier and M. Wheeler. Signalling and territorial aggression: An investigation by means of synthetic behavioural ecology. In D. Cliff, P. Husbands, J.-A. Meyer, and S.W. Wilson, editors, *From Animals to Animats 3: Proc Third Int Conf Simulation of Adaptive Behavior*, Cambridge, MA, 1994. Bradford / MIT Press.

[5] D.C. Dennett. *The Intentional Stance.* Bradford / MIT Press, Cambridge, MA, 1987.

[6] S.J. Gould and R. C. Lewontin. The spandrels of San Marco and the panglossian paradigm: A critique of the adaptationist programme. *Proc Roy Soc*, B 205:581–598, 1979.

[7] A. Grafen. Biological signals as handicaps. *J Theo Bio*, 144:517–546, 1990.

[8] A. Grafen. Modelling in behavioural ecology. In J.R. Krebs and N.B. Davies, editors, *Behavioural Ecology: An Evolutionary Approach*, chapter 1, pages 5–31. Blackwell, Oxford, third edition, 1991.

[9] W.D. Hamilton. The genetical evolution of social behaviour. *J Theo Bio*, 7:1–52, 1964.

[10] D.G.C. Harper. Communication. In J.R. Krebs and N.B. Davies, editors, *Behavioural Ecology: An Evolutionary Approach*, chapter 12, pages 374–397. Blackwell, Oxford, third edition, 1991.

[11] R. Harré. Vocabularies and theories. In R. Harré and V. Reynolds, editors, *The Meaning of Primate Signals*, pages 90–105. Cambridge University Press, 1984.

[12] J.H. Holland. *Adaptation in Natural and Artificial Systems.* University of Michigan Press, Ann Arbor, 1975.

[13] J.R. Krebs and R. Dawkins. Animal signals: Mind reading and manipulation. In J.R. Krebs and N.B. Davies, editors, *Behavioural Ecology: An Evolutionary Approach*, chapter 15, pages 380–402. Blackwell, Oxford, second edition, 1984.

[14] D. B. Lewis and D. M. Gower. *Biology of Communication.* Blackie, Glasgow, 1980.

[15] B.J. MacLennan. Synthetic ethology: An approach to the study of communication. In C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen, editors, *Artificial Life II.* Addison-Wesley, 1991.

[16] B.J. MacLennan and G.M. Burghardt. Synthetic ethology and the evolution of cooperative communication. *Adaptive Behavior*, 2(2):161–188, 1994.

[17] J. Maynard Smith. *Evolution and the Theory of Games.* Cambridge University Press, 1982.

[18] J. Maynard Smith. Adaptationism and satisficing. *Behav Brain Sci*, 6:370–371, 1983.

[19] G.F. Miller. Artificial life as theoretical biology: How to do real science with computer simulation. CSRP 378, School of Cognitive and Computing Sciences, University of Sussex, 1995.

[20] R. Seyfarth, D.L. Cheney, and P. Marler. Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science*, 210:801–803, 1980.

[21] C.E. Shannon and W. Weaver. *The Mathematical Theory of Communication.* University of Illinois Press, Urbana, 1949.

[22] G.M. Werner and M.G. Dyer. Evolution of communication in artificial organisms. In C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen, editors, *Artificial Life II.* Addison-Wesley, 1991.

[23] R.H. Wiley. The evolution of communication: Information and manipulation. In T.R. Halliday and P.J.B. Slater, editors, *Communication*, pages 82–113. Blackwell, Oxford, 1983.

[24] S.W. Wilson. Classifier fitness based on accuracy. *Evolutionary Computation*, 3(2), 1995.

[25] A. Zahavi. Mate selection — a selection for a handicap. *J Theo Bio*, 53:205–214, 1975.

[26] A. Zahavi. The cost of honesty (further remarks on the handicap principle). *J Theo Bio*, 67:603–605, 1977.