Self-defeating human ecosystems[1]

By David Batten and Fabio Boschetti

INTRODUCTION

Use of agent-based simulation to assist in environmental management and decision-making has increased markedly in recent years. Simulation models representing various facets of human behaviour have emerged (Hare and Deadman, 2004). Social scientists use this kind of simulation for various purposes, including the discovery of emergent regularities. Some simulations are built on strong behavioural foundations, where the mental models of autonomous human agents are based on empirical data or stakeholders' views. In other cases, agents' strategies are selected stochastically or even arbitrarily and the outcomes tested in computational experiments.

Both practitioners and theorists have been drawn to the ability of these simulation models to generate large scale, *global* features, arising 'spontaneously' from the *local* interaction of agent populations. Realisation of the *causal* effect of local interactions on global behaviour is particularly interesting, because it shows that global system behaviour can be simulated without explicitly defining system equations 'a priori'. Macroscopic order can appear without interference from any outside agent. Such simulations show how a weakly-interactive system, in which individual agents follow only local rules, can be transformed into a strongly-interactive one displaying emergent, global order. Information can propagate globally through such models via small scale local communication.

Often overlooked in such simulation work is the fact that the intrinsic structure of the system being simulated may have a direct influence on ways in which the agents are likely to behave over time. Global features arising from the interaction process can have a *causal* effect on agents' behaviour, thereby defining a causal feedback loop. Typical examples are people caught in congested traffic, imitating one another in a community or trading in a marketplace. Collective outcomes result from the interactions among the individuals. In turn, however, the behaviour of the individuals themselves will be affected by the collective outcomes in the market. In other words, what people believe affects what happens to a market (or economy or society) and, in turn, what happens to the market affects what people believe.

Systems falling into this category are known as *self-referential* problems – situations where the forecasts made by agents serve to create the very world they are trying to forecast. In these complex systems, the "best" thing to do depends on what everyone else is doing.

---

[1] Earlier versions of this paper were presented at the CABM/HEMA Workshop, Canberra (17-18 May 2004) and the IEMSS Congress, Osnabrück (14-17 June 2004). The first author's contribution is based on a paper that will appear shortly in *Environmental Modelling and Software*.

Copyright © David Batten and Fabio Boschetti, 2005.

The purpose of this chapter is to discuss some implications arising from the self-referential nature of various social systems, with particular reference to the behaviour of human ecosystems. The work is divided into three main parts. First, we discuss some results from the Minority Game (based on the El Farol Bar problem), a self-referential system modelling the competition for limited resources which has received particular attention in recent years. Then we discuss some applications to fishing scenarios and show how this can improve our understanding of resource-depleting, human communities. Finally, we discuss relationships between self-referentiality and concepts such as optimisation and emergence.

SELF-REFERENTIAL SYSTEMS

Definition

A *self-referential* situation is one in which the forecasts made by the agents involved serve to create the world they are trying to forecast. As examples, we shall discuss two extreme cases known as the majority and the minority games. The latter has been pioneered by Arthur (1994), who placed the game in the context of agents choosing whether to attend a bar or not. This classical simulation model serves as a vignette for a larger set of problems, including stock market fluctuations, traffic congestion modelling and a suite of human ecosystems problems. We shall begin by adopting Arthur's setting. Suppose we have a set of agents who must decide whether to attend bar A or bar B. Each agent chooses with the goal of maximising its own enjoyment. As we shall see, the chosen definition of "enjoyment" will affect considerably the dynamics of the community of agents.

Majority and Minority Games

Suppose our agents are young and are keen to socialise, to see and be seen. For them, enjoyment means to be in a large crowd, the bigger the better. No agent knows which bar will have the bigger crowd, so the best thing that any agent can do is to apply the choice strategy (or mental model) that has worked best so far. Agents who happen to choose the bar with the bigger crowd (say bar A) will keep on going to that bar, since they enjoyed it. The others (who attended bar B) will be disappointed and thus will tend to change bar. Soon, the entire crowd will move to bar A. Which bar will end up having the larger crowd is not predictable at the beginning of the simulation, since it may depend on random initial fluctuations but, within a few iterations (i.e. evenings), either A or B will attract the entire crowd (see Figure 1 top). Each agent selects its future action in accordance with the behaviour of the majority of the agents. This self-referential behaviour manifests itself as a positive feedback, and soon the system will saturate, one bar becoming fully crowded and the other empty.

Suppose instead that our agents are older. For them, enjoyment means finding a nice, quiet environment where they can listen to soft music and talk. They will prefer the bar with as few attendees as possible. Now the situation changes radically; the agents who find a quiet bar (bar A, say) will be happy and stick to it. The agents who find a crowded bar (bar B) will change. However, the very act of changing bars will render bar A overcrowded, thereby defeating their choice as well as affecting the enjoyment of the other agents already attending bar A. We call this a self-defeating situation. In this case, the self-referential behaviour has a negative feedback, generating a much richer dynamics worth analysing in detail.

This problem is a metaphor for a broad class of self-referential situations. It has some interesting properties. First, if a decision model existed that agents could rely upon to forecast attendance, then a deductive solution would be possible. No such model exists. Irrespective of past attendance figures, many plausible hypotheses could be adopted to predict future attendance. Because agents' rationality is bounded, they are forced to reason *inductively*. Second, any shared expectations will be broken up. If all agents believe *most* will go, then *nobody* will go. By staying home, that common belief will be destroyed. If all agents believe *few* will go, then *all* will go, thus undermining that belief. Because any mental model that is shared by most of the agents will be *self-defeating*, agents' expectations must always differ.

Perplexed by the intractability of this problem, Arthur created a computer simulation in which his agents were given attendance figures over the past few months. Also, he created an "alphabetic soup" of several dozen predictors replicated many times. After randomly ladling out k of these to each agent, each kept track of his k different predictors and decided whether to go or not according to a preferred predictor in his set. This preferred predictor could be chosen in a variety of ways, although Arthur adopted the currently most accurate predictor for each agent in his simulations. Once decisions have been made in Arthur's simulated bar, agents learn the new attendance figure, updating the accuracy of their own set of predictors. Then decisions are made for the following week.

In this kind of problem, the set of predictors acted upon by agents – called the set of *preferred* predictors – determines the attendance. But the attendance history also determines the set of preferred predictors. We can think of this set as forming a kind of *ecology* (John Holland's term). Of interest is how this ecology evolves over time. Although, the computer-generated attendance results look more like the outcome of a random process rather than a deterministic one (see Figure 1 bottom), there is no inherently random factor governing how many people attend. Weekly attendance is a deterministic function of the individual predictions, themselves being deterministic functions of the past attendance figures.

Thus the bar problem is a relatively simple example of an emergent, self-defeating system. It is a situation where a system of interacting agents can develop collective properties that are not at all obvious from our knowledge of the agents themselves. Even if we knew all agents' individual idiosyncrasies, we are no closer to anticipating the emergent outcome. Under the influence of a sufficiently strong *attractor,* individually subjective, boundedly rational expectations self-organize to produce a kind of "collectively rational" behaviour (Arthur, 1994).

The Bar Problem Extended – Minority Games

So far the agents have been unable to communicate directly with one another. Nevertheless, a considerable amount of information processing occurs in the system. Communication among agents happens, indirectly, via the population dynamics of the bar attendance. This can be used by the agents to make a 'mild' prediction of what action to take next. That this information processing occurs is shown by the fact that the system, under specific circumstances, organises itself into a locally optimum resource exploitation (Savit et al, 1999). The implication of this will be addressed in later sections when we relate self-referentiality to optimisation and emergence.

Given that agents do communicate with each other in the real world, we wonder what may happen if they are permitted to exchange information? Bruce Edmonds' work allows communication among agents before they make their final decisions whether or not to go to the bar (Edmonds, 1999). Using a genetic programming algorithm to simulate adaptive learning, he allows each agent to competitively develop its models of what the other agents are going to do. Although the beliefs and goals of other agents are not known or represented by each agent, heterogeneity among the agents emerges in the form of non-uniform tactics and role-playing identities. These collective properties are features that emerge purely from the micro-dynamics. This reveals one of the subtleties of the minority game: if all players analyze the situation identically, all will choose the same alternative and all will lose. Therefore, players must behave heterogeneously over time. Moreover, there is always some frustration, since not all players can win at the same time [Note: this is an essential mechanism for modelling competition].

Since its introduction, this game has generated incredible interest. As there are only three parameters, the game is suitable for detailed numerical studies and analytical descriptions. Although each alternative is unspecified, for A or B one could read "*I am going to the bar*" or "*I am staying home*"; "*I am choosing the motor way*" or "*I am going by the scenic route*"; "*I am going to fish at the usual place*" or "*I am going to find a better spot*". This shows the relevance of this modelling tool in the analysis of competitive exploitation of limited resources in human ecosystems. It is worthwhile noting that the 'Tragedy of the Commons', as popularised by Hardin (1968), can be seen as a special case of the Minority Problem, in which only one choice yields a return to the agent. Interestingly, this turns the Minority game into a sort of Majority game, in which the only rational choice for all agents is to (over)exploit the resource.

Perhaps the most striking properties of the minority game are that:

- it is a model that addresses the interaction between agents and information;
- agents are able to cooperate and make use of the available information;
- there is a second order phase transition between a symmetric phase (where no information is available to agents) and an asymmetric phase (where information is available to agents)
- If agents take their impact on the game into account, there is no phase transition and this case has an exact solution (we did not mention the meaning of exact solution in the game before). Should we explain this? Or simplify?

The minority game is an abstraction of the bar problem. Like simulations of the bar problem, numerical simulations of this game have displayed a remarkably rich set of emergent, collective behaviours (Challet and Zhang, 1998). Furthermore, both the bar problem and the minority game contain the key elements of a *complex adaptive system*. First, a *medium* number of agents – a number too large for hand-calculation or intuition but too small to use statistical methods applicable to very large populations. Second, these agents are *intelligent* and *adaptive*, making decisions on the basis of rules of thumb or heuristics (like the bar predictors). Needing to modify these rules or come up with new ones if necessary, they reason *intuitively*. Finally, no single agent knows what all the others are (thinking of) doing, because each has access to limited information only.

SELF REFERENTIALITY AND OPTIMISATION

As we have seen, each agent adopts a set of strategies and, during the simulation, chooses a single strategy to employ in different scenarios in order to optimise its own return. On a larger scale, we can imagine the population of strategies as a kind of 'ecology' of strategy sets. In this perspective, a simulation can be seen as an attempt to select the sub-set of strategies that gives maximum return for the agent population. This is equivalent to a search in a large strategy space for 'optimal' strategy sets.

In optimisation theory, it is well known that the difficulty of an optimisation problem depends crucially on the form of the objective function we want to optimise (which in turns determines the shape of the solution space). As with many complex systems, we can view our system at two levels of representation. At the lower level, we have each agent trying to optimise its return by searching within its own strategy set. At the higher level, we have the population trying to optimise the global return by searching within the larger ecology of strategy sets. This situation is reminiscent of individual and collective optima in traffic systems. Each individual wants to minimize her travel time (i.e. travel cost), leading to a solution referred to in the transport literature as the "User-Equilibrium" (UE). Once congestion arises, however, the UE solution will not coincide with what is best for the system as a whole (known as the "Social Optimum or SO solution). Because such an interactive system is self-referential, the two levels influence each other, sometimes in counter-intuitive ways.

It is useful here to introduce the concept of 'alignment' (Wolpert and Turner, 2001). Two cost functions are said to be aligned if, by improving one we do not worsen the other. The majority problem is clearly aligned: by choosing the same bar, all agents improve their returns and so does the community as a whole. Aligned problems are very easy to solve and, as we saw, converge to their optimal return very quickly (all agents attend the same bar). We have two global maxima (all agents attending bar A or all attending bar B). Which solution is favoured depends subtly on the initial conditions, but one of the two global optima will be found with little search effort.

The minority problem is not aligned: the selfish choices of each agent result in non-optimal outcome for the community. ~~if all agents choose the same bar, there is no optimal return for anyone~~. Thus the minority game is much harder to solve. From a population perspective, the return is equal to the size of the minority, so the larger the minority, the greater the return. However, by definition this is not optimal for the majority of the agents, who will tend to switch strategy. This makes the problem misaligned. Not surprisingly, attendance at the bar oscillates continuously.

Here it may be worth asking the following question: Is the problem simply more difficult, or is it *fundamentally* different? Perhaps the problem simply looks more complicated to us because humans feel uncomfortable with self-defeating scenarios and non-linear oscillations. Perhaps it <u>is</u> really more complex? To try to answer this question, it is useful to note a relationship between the concept of alignment and Kauffman's definition of complexity as the 'number of conflicting constraints'

(Kauffman, 1993). In defining a fitness landscape for genetic evolution, Kauffman noticed that the more interconnections between genes (that is, the more the expression of a gene depends on the presence/absence of other genes) the more rugged is the fitness landscape that results. These dependencies and interconnections impose constraints on the expression and success of a gene. Genes' interests are not aligned and the global outcome needs to be 'compromised' among an increasing number of genes, making the solution space more and more complex.

Adopting a similar stance, we can say that the actions of any one agent in a Minority Game are constrained by many other agents; in fact, by the majority of them. Conversely, in a Majority Game the constraints are imposed by the minority, and this minority decreases in size continuously during the simulation. Informally borrowing some terminology from Statistical Mechanics, we can say that, for each agent, the Majority Game has many more winning states than the Minority game.

Misalignment also implies strong nonlinearity, and nonlinear optimisation problems often require some sort of stochastic search method. The population of agents has two choices on how to address a minority game: a) make the problem 'aligned' or b) stochastically search the strategy set;

- The easiest way to make the problem aligned is for the community to decide to share the resource and put the good of the population above the interest of the individual. In game theory, this is called a team game (Wolpert and Turner, 2001), and is equivalent to assigning a share of the population return to each agent, independently of which bar they choose. Clearly, this approach may not be easy to implement in real world situations. A variant of this approach involves properly tuning each agent's cost function, and this has been explored for the Bar problem (Wolpert and Turner, 2001) and in a simplified fishing scenario (Boschetti, 2005).

- Stochastic search strategies (genetic algorithms, simulated annealing, Swarm optimisation, Monte Carlo simulation) involve a mixture of exploration and exploitation. In optimisation jargon, exploration refers to searching in areas of the parameter space which have not been sampled, and exploitation refers to searching in the vicinity of a current local optimum (exploiting the basin of attraction currently found). When viewed within a resource management scenario, the concepts of exploration and exploitation have a striking parallel with strategies found in real communities, which is worth a closer analysis.

SELF-REFERENTIALITY, EMERGENCE AND PREDICTABILITY

Recently, the study of emergence has received considerable attention. In a practical modelling scenario, emergence can be viewed as large scale features of the dynamics that the modeller did not explicitly code into the system. In a theoretical perspective, however, emergence poses deeper questions, with the validity and necessity of the concept itself being questioned by scientists and philosophers. While the theoretical discussion is beyond the scope of this paper, it is worth noting that self-referentiality may be a good candidate under a number of current definitions of emergence (see Boschetti et al, 2005).

Some authors define emergence as the arising of 'downward causation' in a system (Goldstein 2002, Bickhard 2000, Heylighen, 1991). According to this view, a feature is emergent if it has some sort of causal power on lower level entities. While it is widely accepted that lower level entities must have an 'upward' causation on the higher level 'emergent' features (e.g. community behaviour is the result of the action of its individual members), this approach assumes a two-way causal relation (the structure of the community in turns affects the action of its members). This concept is attractive because it clearly goes beyond the reductionist approach of viewing large systems as merely the composition of smaller entities. It is also consistent with the co-evolutionary feedback effects of self-referential systems, where the system affects the agent and the agent, in turn, affects the system.

Strictly speaking, self-referentiality (i.e. determination of the success of a strategy as a function of the behaviour of the entire population) may not be an emergent feature of dynamic systems. It could be argued that this feature is explicitly coded in the model, in the definition of 'minority wins' itself. What does emerge, however, is information processing in the system, which allows for information sharing among agents. This leads to the concept of 'intrinsic emergence' (Crutchfield, 1994) and 'efficiency of prediction' (Shalizi, 2001). Within this view, the large scale patterns in the bar attendance are used by the agents to make a prediction of what bar may be majority/minority choice at the next iteration. Putting it differently, the system generates a feature (the bar attendance record) that is used by the agents (i.e. the system itself) to attempt to understand the functioning of the system, in order to improve its own predictability. As human agents need to devise effective predictive capabilities to survive in the real world, the ability to detect and exploit emergent features becomes an evolutionary requirement undergoing selective pressure[2], and therefore well worth studying.

Clearly, any prediction may be right or wrong. As we saw in the case of the majority game, the population dynamics itself provides an effective prediction and the system quickly converges towards an optimum. In the minority problem, however, arising predictions are constantly thwarted, agents are frustrated and fluctuations occur as a result. We may wonder where the advantage of making the prediction lies in the minority game. It should be seen in relative terms compared to the agents' alternative options. Agents can choose:

1) not to attempt any prediction, but to act randomly instead;
2) to use the population dynamics for prediction;
3) to try to predict the behaviour of every other agent.

Each option involves a different information processing cost. In evolutionary terms, it might be argued that each agent would tend to select the method which gives the best ratio between the predictability gain and the cost involved in the prediction (Shalizi, 2001). Current research in the minority game has attempted to model the points above, but the cost of the prediction was given only in terms of the length of the historical records available to the agents. This may be misleading, if not coupled to a formal estimation of the complexity of the historical record time series (i.e. the

[2] Our own life depends, minute by minute, on our ability (consciously or not) to predict and on the predictions being, by and large, correct.

amount of information actually needed for optimal prediction) or if we do not account for the agents being able to trade off the amount of resources saved (or employed) in the prediction. This observation seems to suggest a possible model scenario which could help us gain some insight into how relevant the concept of efficiency of prediction is to human ecosystem modelling, as well as on its relation to self-referentiality (down-ward causation). Notice for example, that a causal relation between efficiency of prediction and self-referentiality could very well be, itself, two-way.

Let us consider a modelling scenario which includes:
1) agents endowed with a vast range of strategy choices, as:

- no action;
- random actions;
- simple imitation of other agents' behaviour;
- selection from a strategy set of variable size;
- the ability to analyse the historical record time series at different levels on computational complexity;
- the ability to analyse and predict the action of other agents;

2) the ability to adaptively select among the above choices;
3) the ability to generate new strategies by searching a strategy space;
4) a different 'cost' associated with each of the above choices. We envisage, for example, no cost for random actions, low cost for pure imitation, higher cost for time series prediction and the highest cost for attempting to predict the actions or reactions of all the other agents;
5) the ability to employ the resources not invested in prediction in order to carry out some sort of useful activity;
6) the overall system would model a process which may or may not be self-referential depending on the specific configuration of the agents behaviours.

In this list, we catch a glimpse of concepts like agent's mental attitude ('shall I bother guessing", "is it worth making an effort?"), spreading of ideas (peer imitation), cost and reward of information gathering and analysis, creation of new 'ideas' (searching a strategy space) and the possible arising of one or more self-referential communities. Also, we may note that self-referentiality imposes contingency on the 'value' of an action[3]. This suggests that self-referentiality may offer an (admittedly very simple) avenue to model a very basic instantiation of the 'meaning' of an agent's actions. When viewed in this way, the development and evolution of 'meaning' becomes somewhat autonomous, since it 'emerges' from the agents' local interactions.

[3] Consider a slightly different interpretation of a Minority Game simulation. Suppose that the agents represent artists and choices A and B represent two different artistic styles. An agent (Agent Degas) choosing the minority choice (Impressionism) would be considered 'original'. In some societies, originality is given a positive value so Agent Degas would be considered 'original/interesting'. Now suppose that, at the next iteration step, the choice "Impressionism" becomes the majority choice and Agent Degas has not changed its style. Now, Agent Degas is no longer 'original' and its art is now less interesting.

HUMAN ECOSYSTEMS

In human ecosystems, human agents interact with one another and other life in a natural environment. Self-referential problems of the binary variety arise in these systems, but are seldom recognized as such. Since Hardin's influential article in Science (Hardin, 1968), most are treated purely as commons dilemmas in which agents over-exploit a scarce resource in common. Examples are the over-fishing of fisheries, the degradation of national parks and the destruction of coral reefs. The question we address in the next section is to what extent the fisheries problem may be viewed as a complex adaptive system of the self-defeating variety.

Fisheries

In an insightful paper almost two decades ago, Allen and McGlade (1986) stressed that the key elements of fishing (like ancestral hunting) are *discovery* and *exploitation*, not simply the latter as in western agriculture nowadays. As well as a desire to avoid congestion (like in the bar problem), fishing involves communication, adaptive learning and the emergence of non-uniform tactics and role-playing among fishing agents. Within such a socio-ecological system, key questions on our relationship with nature and the general problem of the management of a complex adaptive system can be explored.

Largely hidden from human view, the marine ecosystem is complex with respect to its species composition as well as to the processes occurring in it. As exploitation increases, Beddington and May (1977) showed that such an ecosystem moves towards instability and greater risk of collapse. Human responses tend to amplify random fluctuations, thereby further increasing the risk of collapse.

Ignoring the complex issue of ecosystem dynamics for the moment, let us concentrate on the strategies of the owners of the fishing vessels. From an owner's selfish viewpoint, if there are no institutional rules imposed on him, he aims to maximize his own vessel's catch. Obviously, the "best" thing for him to do will depend on what all the other fishing vessels in the vicinity of his vessel are doing. Thus our fishing problem is a self-referential one.

Allen and McGlade (1986) developed dynamic, multi-species, multi-fleet models to explore the implications of different fishing strategies and information flows among fishing vessels in a Nova Scotia fishery. They identified two fishing strategies, calling them *Cartesians* and *Stochasts*. Stochasts search randomly for better sites or use their own intuition, without resorting to information that is shared between associates. They are risk-takers seeking higher returns commensurate with the higher risks they take. At the other extreme we have the Cartesians, skippers who (for various reasons) are unwilling to take any risk and who only go to the zone promising the best *known* return. It matters not how long it takes for a Cartesian to reach their chosen site, as long as the final catch is guaranteed.

Allen and McGlade found that less information exchange among the fishing vessels ensures a more random response on the part of boats (i.e. more Stochasts among the fleet). Because the Stochasts continue to explore less visited parts of the system, the fishery as a whole has a greater chance of survival. If boats refuse to take risks and go only to where they know there are fish, the end result can be disastrous for everyone. Discovery involves risk, but to abandon it invites disaster.

There is a powerful message here that goes beyond the bounds of fisheries alone. In any mobile population, for example, we find some risk-takers and some who are risk-averse. Like the number of fishing boats turning up at the same site, the number of vehicles turning up on a specific road each day is unpredictable. Of interest are the adaptive strategies of drivers exposed to regular traffic jams. Downs (1962) identified two behavioural classes of driver: those with a low propensity to change their mode or route strategy, called *sheep*, and those with a propensity to change, called *explorers*. Explorers search for alternative options to save time. They are quick to learn and hold several heuristics in mind simultaneously. Sheep are more conservative and prone to following the same option. Empirical work in North America has confirmed the presence of sheep and explorer behaviour in real traffic (Conquest et al, 1993).

The parallels between Cartesian or Stochast fishing strategies and Sheep or Explorer driving strategies may seem striking. Yet they are less surprising if thought of as symptomatic of a more general phenomenon. In the world of technology, risk-averters and risk-takers appear under different guises: *imitators* and *innovators*. If we allow for the coexistence of imitative and innovative mechanisms in a population, both must be treated as co-evolutionary variables dependent on the unfolding of events.

Evolutionarily Stable Strategies

Is the coexistence of Cartesian (imitative) and Stochast (innovative) strategies in a human population an *evolutionarily stable strategy* (ESS)? An ESS is a strategy (or mix of strategies) which, if most members of a population adopt it, cannot be bettered by an alternative strategy. The reason is that the fitness of individuals adopting an ESS strategy is higher than the fitness of individuals adopting other strategies (Maynard-Smith, 1982). Another way of putting it is that the best strategy for any individual depends on what the majority of the population is doing (Dawkins, 1976). This last perspective helps us to grasp the idea that an ESS may be an important property of self-referential systems.

There are three classes of ESS: pure, mixed and conditional. A pure ESS is a strategy that is consistently exhibited by individuals throughout their lifetimes. A mixed ESS is a complex of two or more strategies varying either within or among individuals over time. In a conditional ESS, each individual's strategy varies under different conditions in the social or physical environment. If the evolutionary payoff for each strategy depends on what other individuals are doing and decreases as a greater proportion of the population adopts that strategy, other strategies may also be an ESS (frequency-dependent selection).

To apply the ESS idea to our fishery problem, consider the following variant of one of Maynard Smith’s simplest hypothetical cases (hawks and doves). Here we draw extensively on the simulation results in Allen and McGlade (1986). Suppose that the only two types of fishing strategy in a fleet of vessels are Stochasts and Cartesians. We want to know whether *Pure-Stochast* or *Pure-Cartesian* is an evolutionarily stable strategy. As Stochasts and Cartesians compete for returns, we must estimate payoffs to the different fleet strategies in order to find an ESS.

Serving as the "eyes" of the fleet, Stochasts search for fishing zones of high return. A fleet with a high proportion of risk-taking Stochasts will discover successive zones of high return, eventually fishing out the high-return zones in a patchwork manner. But Stochasts ignore fish from nearby zones of intermediate returns. As it takes time for the high-return zones to recover, pure Stochast strategies tend to lead to "boom-and-bust" series of good and bad years. However, this patchwork approach guarantees that the fishery system survives rather than running the risk of collapsing.

Based on the results in Allen and McGlade (1986), we can assign a *Pure-Stochast* a cumulative payoff of 76 points (equivalent to 6 large circles plus 8 small ones) over a ten-year fishing period. This includes several bad years during which the stocks must recuperate.

Cartesians refuse to take risks and rely solely on the information passed on by the Stochasts. Thus they go only to zones where they know in advance that there are some fish. A fishing fleet dominated by risk-averse Cartesians will simply direct all the boats in the short term to a few of the "best" zones for fishing. Because the information about which are the best zones is generated by a small number of Stochasts, the tendency for Cartesians is to "lock" onto one particular location as the best. In the long run, this leads to the fleet exploiting a single location instead of the whole area, leading to a small catch and a small fleet. Based on figures found in Allen and McGlade, a *Pure-Cartesian* strategy yields a payoff of 45 points over the same ten-year period.

The above figures confirm our suspicions that *Pure-Stochasts* do better than *Pure-Cartesians* over time. They are rewarded for taking more risks. *Pure-Stochasts* achieve better returns because they can maintain fishing activities over the whole area, obtaining a higher catch in the good years and maintaining a larger fishing industry overall. *Pure-Cartesians* become trapped in a few locations instead of spreading out across the whole area. They are limited each year to a smaller catch and a declining fishing fleet and industry over time.

Table 1: Strategies and Payoffs for Homogeneous Fleets

FLEET STRATEGY	10-YEAR PAYOFF
Pure-Stochasts	76 points
Pure-Cartesians	45 points

But is a *Pure-Stochast* strategy evolutionarily stable on its own? To answer this question, let us suppose that we have a fleet consisting only of *Pure-Stochasts*. Despite the bad years, they seem to do very nicely, earning the payoff given in Table 1. Now suppose that a Cartesian arises in the fleet. Being the only Cartesian, must he compete aggressively with the *Pure-Stochast?* No. Instead he could follow the *Pure-Stochasts,* letting them show him where it is best to fish. Then, by fishing at the edges of sites discovered by *Pure-Stochasts*, he does not irritate them too much. Remaining mostly unnoticed, he can "free-ride", enjoying above-average returns for very little search effort.

As long as there are sites offering intermediate returns that are left untouched by Stochasts, there are niche opportunities for Cartesians to fill. Based on figures in Allen and McGlade (1986), ten-year payoffs to a mixed fleet of Stochasts and Cartesians under various levels of information exchange are given in Table 2.

Table 2: Fleet Strategies and Payoffs when Information is Shared

FLEET STRATEGY	10-YEAR PAYOFF
If catch information is shared equally:	
Stochasts	104 points
Cartesians	83 points
If Cartesians fail to inform Stochasts:	
Stochasts	95 points
Cartesians	81 points
If Stochasts fail to inform Cartesians:	
Stochasts	109 points
Cartesians	30 points

When information is shared equally between both groups, the payoff to each is higher than to a *Pure-Stochast* or *Pure-Cartesian* fleet (see Table 1). Thus a *Pure-Stochast* strategy is not evolutionarily stable since it can be bettered by a strategy in which Stochasts and Cartesians communicate and cooperate with each other.

What happens if either group cheats by not transmitting reliable information about their catches? As Allen and McGlade noted, there is an interesting asymmetry. If Cartesians fail to inform Stochasts, there is very little effect (see Table 2). But if the information possessed by Stochasts is not transmitted to Cartesians, the latter perish (or are compelled to become Stochasts). Thus Cartesians will try to obtain catch information and Stochasts will try to withhold it. Such effects have evolved in real fishing fleets. For example, Vignaux (1996) found that trawlers fishing for *hoki* in waters off New Zealand's coastline do not share catch information, instead basing their own decisions partly on watching where other vessels fish. This amounts to spying. Other tactics include "listening in" on radios, lying about catches and spreading other misleading information.

Using agent-based simulation, a richer suite of possibilities can be explored. Recent ABM work involving Bayesian belief networks has shown that various kinds of information flow among fishing vessels has important effects on the dynamics and resource exploitation of a simulated fishery [Little et al, 2004]. As stated earlier, information flow tends to benefit the Cartesians at the expense of the Stochasts. This asymmetry confirms the importance of realistically representing the rich variety of possible fisher behaviour in any modelling framework that aims to assist with integrated management of fishery resources.

A Modified Minority Game?

The superior payoff achieved when catch information is shared reliably is in line with observed properties of the minority game. If all agents in a fleet have access to public information about catches and zones of highest return over a period of time, then the agents interact only through this public information and the system has a "mean-field" character (in the sense that no short-range interactions exist). Self-organization in such a system is achieved by allowing each agent to have several strategies from which he selects the one that seems best (to him).

In the literature on the minority game, there is a second order phase transition between a symmetric phase (where no information is available to agents) and an asymmetric phase (where information is available). It is easy to see why this phase change might occur in our fisheries context. First, if reliable catch information is passed between Stochasts and Cartesians, this may ensure that a "boom-and-bust" series of good and bad years is avoided. Stochasts can direct some Cartesians to high-return zones and others to medium-return zones, in such a way that no high-return zone will be in great danger of over-fishing. Second, the two strategies are complementary. Cartesians make good use of information, while Stochasts generate it. Together they can exploit the resource more efficiently because of their complementarity.

Ideally, each fleet needs some Stochasts (researchers) and Cartesians (producers) who cooperate within a fleet but not with competing fleets. To survive, each needs the other. What, then, is the ideal ratio of Stochasts to Cartesians? This is a difficult question to answer without considering the dynamics of the fish population and other environmental factors. The higher payoff that a Stochast may enjoy is tempered by the higher risks involved, whereas the risks taken by a Cartesian are minimal.

In a highly dynamic world like fisheries, one may expect the ratio of Stochasts to Cartesians in a fleet to vary in response to the search success of the Stochasts and their willingness to inform their Cartesian partners. Like music lovers at the El Farol bar, the ratio of Stochasts to Cartesians can oscillate forever. But, in this case, there may be no emergent regularity as the congestion level is a dynamic variable. The situation is a self-defeating one, since the "best" thing for each vessel to do depends on what everyone else is (thinking of) doing.

DISCUSSION

Discovering the emergence of global dynamics from the lower level interactions of a set of agents has always been a key aim of agent-based modelling (Batten, 2000). It is difficult to say whether self-referentiality formally qualifies as an emergent property.

In any case, this is beyond the scope of this work. What we have shown, however, is how self-referentiality – with its generation of positive and negative feedback loops – opens the door to the possibility of modelling a vast range of human behaviours more realistically, using agent-based tools that are relatively simple and inexpensive to run.

We have discussed some interesting relationships between self-referentiality and the concepts of optimisation, information processing, and predicability. This suggests that, by acting as a rudimentary proxy for instantiations of meaning, we can envisage an alternative model for the spreading of ideas and the development of culture (should we keep this one?). More importantly, each of these aspects can be followed at the two levels of representation: (1) agents and (2) community. Individual optimisation is attempted at the agent level, where agents search for their "best" strategy. At the community level, however, the aim is to identify the best mix of strategies to achieve a reasonable compromise in the face of individual goals which do not necessarily coincide.

It is worthwhile noting that the agent communities that we have discussed so far have been closed, in the sense that no communication with the environment has been possible. Because of this, the levels of description and the extent of self-referentiality are very limited. Opening up the human system to its external 'environment' could result in quite different outcomes. We could envisage one obvious (1) and two more audacious scenarios (2,3):

1) Modelling the ecological system's dynamics (e.g. fish population dynamics in the fishery example) or those of separate communities (e.g. species of fish). This would affect the evolution, optimisation, and adaptation of the agents and possibly identify more self-referential situations which, in turn, could offer novel intrinsic information processing capabilities.
2) In his pioneering work on autonomous agents, Kauffman (2000) places considerable emphasis on the inherent variety which arises from agents that use energy to carry out work to survive and make a living. This, in turn, leads to the realization that extra work can be used in order to devise tools to extract more energy and so on, in an ongoing loop. Currently, agent-based models do not include such a concept of work, nor the cost of attaining resources and information, nor what to do with the energy accumulated.
3) Wiedermann and van Leeuwen (2003) suggest that an environment enhances the computational power of agent communities. In particular, they describe the role of the environment not only as a source of continuous novelty (which allows the computation to 'go on and never halt'), but also as providing tools for storing information that can be accessed at a later stage. This combination furnishes *super-Turing* capabilities for the agents, implying a fundamentally different computational power. While this may not be relevant to relatively small systems like the minority and majority game, it may represent a crucial difference for larger systems.

CONCLUSIONS

Previous simulation work referred to in this paper has shown the importance of information-sharing and communication strategies within fishing communities (Allen and McGlade, 1986; Maury and Gascuel, 2001; Little et al, 2004). Irrespective of whether vessel owners within fishing fleets do share reliable catch information or not, agent-based models can help to clarify the potential value of information-sharing. Earlier work has shown that lower returns will be experienced by Cartesians in the absence of information flow from Stochasts. In the long run, however, Stochasts suffer "boom-and-bust" years by fishing alone because they fish out the high-return zones in a patchwork manner. Thus the evolutionarily stable strategy for a fishing fleet is a mixed one, i.e. one in which Stochasts and Cartesians communicate and cooperate with each other.

When the economic viability of a fishery is under serious threat, it is worthwhile adopting a cooperative approach – not only to encourage information sharing, but also to overcome such common pool dilemmas. If conditions are suitably conducive, a sustainable solution can even be found by the participants themselves. One successful example is the inshore fishery at Alanya in Turkey (Ostrom, 1990). Members of the local cooperative found an ingenious rotation system allocating fishing sites to local fishers that builds upon reliable and mutually beneficial information exchange.

As mentioned near the outset, the purpose of this paper was to look at a class of self-referential problems called *self-defeating* systems. A typical example is the bar problem, where the simulated population of attendees oscillates in a seemingly random manner around a critical congestion level. Another self-defeating problem is urban traffic congestion. Without a degree of cooperation and information sharing between the Sheep and Explorers (i.e. Cartesians and Stochasts) on our roads, the collective behaviour of these complex adaptive systems will continue to defy prediction or control on a daily basis.

The fishery situation is of a similar ilk. As well as a desire to avoid congestion, fishing involves communication, adaptive learning and the emergence of role-playing & non-uniform tactics among the agents involved. Human behaviour in national parks and coral reefs raises similar collective problems. In each of these ecosystems, it is impossible to achieve an eco-efficient outcome unless a threshold level of cooperation and information sharing can be achieved. Stochasts and Cartesians need to work together in order to achieve sustainability and thus avoid over-exploiting the resources therein.

Within such socio-ecological systems, the use of agent-based simulation allows key questions about our relationship with nature and the more general problem of the management of a complex adaptive system to be explored qualitatively. Rule-change experiments – like those that led to the successful fishing regime at Alanya – fall into the scientific domain of participatory, agent-based modelling. The great thing about this kind of simulation is that we can assess the likely outcomes of such rule changes ahead of their implementation.

As individual humans – in our roles within families, communities, firms, institutions and regions – we must decide how to divide our time and effort between doing what we know (with known values and payoffs) and searching for new opportunities and roles that may have superior payoffs in the future. We can choose to explore new pathways and connections or try to minimize such deviations from those pathways that we are accustomed to in our lives to date. Thus "Stochasts" and "Cartesians are two important extremes in human society. The first group take more risks, venturing into the unknown, be it by hunches, tactics or entrepreneurial activities. Their discoveries nourish a society in the long run, expanding it into the adjacent possible and assuring its long-term survival in some form. The second group prefers to devote themselves to tasks already assigned to them, seeking to undertake them as quietly and efficiently as possible. They constitute the "backbone" of a society and come much closer to our definition of "normality." The survival and sustainable prosperity of any individual, community, firm or nation, or even a human ecosystem, requires a mixture of both types of behaviour.

ACKNOWLEDGMENTS

Helpful comments on an earlier draft from Paul Atkins, Marco Janssen, participants at the joint CSIRO Agent-Based Modelling/Human Ecosystems Modelling with Agents (CABM/HEMA) Workshop, Canberra (17-18 May 2004) and the IEMSS Congress, Osnabrück (14-17 June 2004 are gratefully acknowledged. The usual caveat regarding authors' responsibility applies.

REFERENCES

Allen, P.M. and J.M. McGlade, Dynamics of discovery and exploitation: the case of Scotian Shelf fisheries, *Canadian Journal of Fisheries and Aquatic Sciences*, **43**: 1187-1200, 1986.

Arthur, W.B. Inductive behaviour and bounded rationality, *The American Economic Review,* **84**: 406-411, 1994.

Batten, D.F. *Discovering Artificial Economics: How Agents Learn and Economies Evolve*. New York: Westview Press, 2000.

Batten, D.F. and R. Bradbury, Simple memes and complex cultural dynamics, paper presented at an International Workshop on Innovation, Dynamic Regions and Regional Dynamics, Jönköping, Sweden, June, 2005.

Beddington, J.R. and R.M. May, Harvesting natural populations in a randomly fluctuating environment, *Science*, **197**: 463-5, 1977.

Bickhard, M. H., Emergence, in P. B. Andersen, C. Emmeche, N. O. Finnemann, P. V. Christiansen (Eds.), Downward Causation. (322-348). Aarhus, Denmark: University of Aarhus Press, 2000.

Boschetti, F. Improved Resource Exploitation by Collective Intelligence, *ModSim05*, Melbourne, 12-15 December, 2005.

Boschetti, F., M. Prokopenko, I. Macreadie and A. M. Grisogono, 2005, Defining and detecting emergence in complex networks, KES 2005, Melbourne, 2005.

Casti, J. Would-be worlds: the science and surprise of artificial worlds, *Computers, Environment and Urban Systems*, **23**: 193-203, 1999.

Challet, D. and Y-C. Zhang, On the minority game: analytical and numerical studies, *Physica A,* **256**: 514, 1998.

Conquest, L.J. Effect of motorist information on commuter behaviour, *Transport Research B* **18**: 101-110, 1993.

Crutchfield J., The Calculi of Emergence: Computation, Dynamics, and Induction" Physica D 75 1994: 11-54

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

Downs, A., The law of peak-hour expressway congestion, *Traffic Quarterly,* **16**: 393-409, 1962.

Edmonds, B., Gossip, sexual recombination and the El Farol bar: modelling the emergence of heterogeneity, *Journal of Artificial Societies and Social Simulation*, **2**(3): 1-21, 1999.

Goldstein, J., The Singular Nature of Emergent Levels: Suggestions for a Theory of Emergence, Nonlinear Dynamics, Psychology, and Life Sciences, Vol. 6, No. 4, 2002.

Hardin, G. The tragedy of the commons, *Science*, **162**: 1243-8, 1968.

Hare, M. and P. Deadman, Further towards a taxonomy of agent-based simulation models in environmental management *Mathematics and Computers in Simulation,* **64**: 25-40, 2004.

Heylighen, F., Modelling Emergence, World Futures: the Journal of General Eveloution, special issue on creative evolution, G. Kampis (ed), 1991.

Kauffman, S.A. *The Origins of Order.* Oxford University Press, New York, 1993.

Little, L.R., Kuikka, S., Punt, A.E., Pantus, F., Davies, C.R. and B.D. Mapstone, Information flow among fishing vessels modelled using a Bayesian network, *Environmental Modelling and Software*, **19**: 27-34, 2004.

Maury, O. and D. Gascuel, Local over-fishing and fishing tactics: theoretical considerations and consequences in stock assessment studied with a numerical simulator of fisheries, *Aquatic Living Resources*, **14**, 203-210, 2001.

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

Ostrom, E. *Governing the Commons*: *the Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press, 1990.

Savit, R., Manuca, R., and Riolo, R., Adaptive competition, market efficiency, and phase transitions. *Physical Review Letters*, **82**(10): 2203-2206, 1999.

Shalizi, C., Causal Architecture, Complexity and Self-Organization in Time Series and Cellular Automata, PhD Thesis, (http://www.cscs.umich.edu/~crshalizi/thesis/), 2001.

Vignaux, M., Analysis of vessel movements and strategies using commercial catch and effort data from the New Zealand hoki fishery, *Canadian Journal of Fisheries and Aquatic Sciences*, **53**: 2126-2136, 1996.

Wolpert, D. and K. Tumer, Optimal Payoff Functions for Members of Collectives, Advances in Complex Systems, 4(2/3):265–279, 2001.

Wiedermann, J., and J. van Leeuwen: The emergent computational potential of evolving artificial living systems. AI Commun. 15(4): 205-215, 2002.

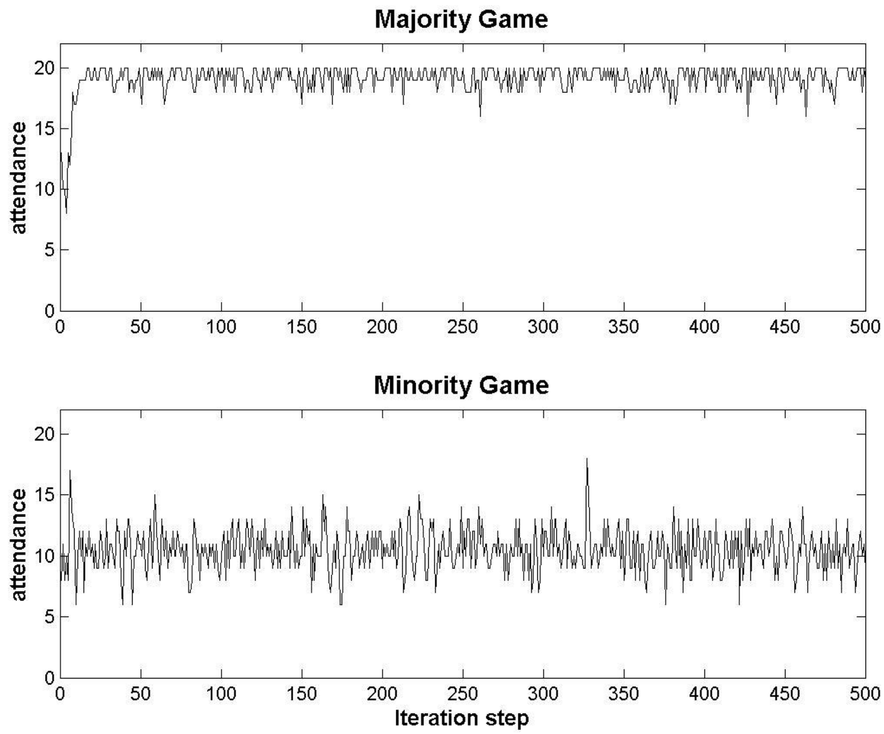

Figure 1: Example of the output of a simulation run for the Majority Game (top), and Minority Game (bottom). Both simulations involve a community of 20 agents who compete for 500 iterations.