Case-based reasoning and social dilemmas: an agent-based simulation.

Presented at ESSA/SIMSOC-VI Conference, 19-21 September 2003, Groningen, Netherlands

L. R. Izquierdo, N. M. Gotts & J. Gary Polhill
Macaulay Institute, Craigiebuckler, Aberdeen. AB15 8QH
l.izquierdo@macaulay.ac.uk, n.gotts@macaulay.ac.uk, g.polhill@macaulay.ac.uk
http://www.macaulay.ac.uk/fearlus/

Abstract

This paper presents a study on the role of adaptation and the effects of other intrinsic motivations in addition to the economic payoffs in social dilemmas. Social dilemmas are modelled in an abstract way using a game-theoretic framework and agent-based modelling. The role of adaptation is studied by implementing agents that use case-based reasoning in their decision making algorithm. The effects of other intrinsic motivations are explored by introducing a certain preference for social approval that makes the agents be slightly more inclined to cooperate. The experiments conducted reveal a new concept of equilibrium characterised by the fact that no player can be guaranteed a higher payoff by changing their decision. It is proved that there is a broad range of adaptive decision-making algorithms for which such concept of equilibrium can be relevant.

Keywords: Social dilemmas, Case-based reasoning, Prisoner's dilemma, Agent-based simulation, Analogy.

1. Introduction

This paper presents a study on the role of adaptation and the effects of other intrinsic motivations in addition to the economic payoffs in a certain type of social dilemma. To perform the experiments outlined in this paper we have used a model constructed for that purpose, but the work reported here will feed into the process of extending the spatially explicit agent-based modelling system for land use change, FEARLUS[1] (Polhill, Gotts and Law 2001; Gotts, Polhill and Law 2003a). The FEARLUS project is aimed at using agent-based social simulation (Conte, Hegselmann, and Terna, 1997) to increase understanding of the processes underlying land use change, particularly at the regional scale and in the medium to long term. FEARLUS is being extended to deal with common-pool resource (CPR) management problems, as part of a larger project aimed at developing ways to synthesise stakeholder priorities in relation to environmental issues. This project is using the implementation of the EU Water Framework Directive (EU, 2000) as a case study, and is referred to here as the WFD project. The WFD project involves hydrological modelling, socio-economic approaches (including interviews of

[1] FEARLUS stands for 'Framework for Evaluation and Assessment of Regional Land Use Scenarios'. The existing FEARLUS model source code and user guide are available online at http://www.macaulay.ac.uk/fearlus/download.html

stakeholders which will explore their environmentally-relevant values), and agent-based social simulation.

A CPR management problem arises when multiple agents have access to a resource which is both subtractible (if one agent uses more, there is less available for others), and difficult to exclude potential users from – water resources often fall into this category. In such situations negative externalities occur: the costs of an individual's action are spread out over the whole group of appropriators whereas the benefits are often enjoyed only by the actor. Negative externalities can be so strong that the benefit obtained by any individual when increasing appropriation is lower (in absolute value) than the loss imposed to the rest of the group by that individual's action. Such circumstances set up a social dilemma: for each actor individually it is advantageous to use more, but if all do so, all end up worse than if all had restrained themselves cooperatively.

For many years, it has been accepted by many scholars that rational and self-interested individuals will not act to achieve their common interest in a social dilemma where those who don't cooperate are always better off than those who cooperate. If all participants choose not to cooperate, the collective benefit will not be produced.

However, existing empirical work on CPR management problems (Ostrom, Gardner, and Walker, 1994) indicates that while it is not easy to establish cooperation, levels of cooperation tend to be higher than would be expected if the agents concerned were the computationally unlimited and entirely self-interested agents of neoclassical economics.

Two possible reasons for this are implicit in this characterisation: real human agents:

a. Are limited in their cognitive powers and resources. They cannot calculate all the possible consequences of the actions available to them, but they do *adapt* their behaviour according to their past experience.
b. Appear to have intrinsic motivations other than personal gain. In particular, considerations of fairness frequently influence behaviour (see work reviewed in Gotts, Polhill and Law 2003b). Moreover, the salience of different drivers of behaviour may be different for each person and may be context-dependent.

The two characteristics outlined above have been modelled in a fairly abstract way, reflecting the approach that we follow in the FEARLUS team: we begin with simple models so we can undertake an adequate exploration of the parameter space and a sound interpretation of the results, and we build in additional complexity only as and when required. As the WFD project progresses, we will increase the realism of our simulations, enlightened by the work undertaken by our hydrological and socio-economic colleagues and by the results obtained from our simpler models. The focus of our research is on the identification of factors that contribute to increasing the overall level of cooperation, rather than on proposing strategies that maximise financial benefit to the individual.

This paper is structured as follows: section 2 outlines the experimental setup. Section 3 examines the effects of computationally limited but adaptive cognition. For this purpose, we have chosen to have our Agents[2] use a form of case-based reasoning: we consider it plausible as at least a partial representation of how people make use of past experience, that they recall circumstances similar to those they now face and remember what they did and with what outcome (see for example many of the chapters in Kahneman, Slovic and Tversky (1982)). Results for both the Prisoner's Dilemma and the Tragedy of the Commons game are provided in section 4. These results have revealed a new concept of equilibrium which is discussed in section 5. In section 6 we explore the additional effect of giving our Agents a second, non-financial intrinsic motivation: a preference for the good opinion of their peers, which inclines them toward cooperation. Finally, conclusions are provided in section 7.

2. The experiment setup

We model social dilemmas in an abstract way by setting our Agents to play a 2-player and a 25-player version of the Prisoner's Dilemma. Players can either cooperate or defect. These two games capture the essence of a social dilemma: every individual has strong incentives to defect, but every one is better off if all cooperate than if all defect.

Table 1 shows the 2-player Prisoner's Dilemma payoff matrix. Payoffs on the bottom left of each cell are for Player 1 and payoffs on the top right are for player 2.

		Player 2	
		Cooperate	**Defect**
Player 1	**Cooperate**	*Reward* (P2) / *Reward* (P1)	*Temptation* (P2) / *Sucker* (P1)
	Defect	*Sucker* (P2) / *Temptation* (P1)	*Punishment* (P2) / *Punishment* (P1)

Table 1. Payoff matrix for the 2-player Prisoner's Dilemma.

Because of the decision making algorithm of our Agents (explained in section 3), the actual values of the Payoffs are not relevant as long as they satisfy:

$$Temptation > Reward > Punishment > Sucker$$

In the 25-player social dilemma every agent gets a reward as long as there are no more than n defectors. In the absence of the reward, the payoff that defectors get (*Def-P*) is higher than the payoff obtained by those who cooperate (*Coop-P*). However, once the reward is offered, every player is better off if they all cooperate than if they all defect (*Coop-P* + *Reward-P* > *Def-P*). Table 2 shows the payoff matrix for a particular agent:

[2] The word 'Agent' will be written starting with a capital 'A' when it refers to the computer implementation.

	Fewer than *n* others defect	*n* others defect	More than *n* others defect
Agent cooperates	*Coop-P + Reward-P*	*Coop-P + Reward-P*	*Coop-P*
Agent defects	*Def-P + Reward-P*	*Def-P*	*Def-P*

Table 2. Payoff matrix of the "Tragedy of the Commons game" for a particular agent.

This game has been called in the literature the "Tragedy of the Commons game" (Kuhn, S., 2001) after the influential paper written by Hardin (1968). When the number of players is low, it represents a version of the "volunteer dilemma": a group needs a few volunteers, but each member is better off if others volunteer. If the number of players is large enough, the case when exactly *n* others defect is sufficiently unlikely that for all intents and purposes it can be ignored. Assuming the latter, we have a "social dilemma" as defined by Dawes (1980): "all players have dominating strategies[3] that result in a deficient equilibrium[4]". In any case, we have a "problematic social situation" (Raub, 1988) which can be defined in game theory terms as a game with Pareto inefficient[5] Nash equilibria[6].

These two games will be played repeatedly for an indefinite number of periods (until a cycle in the output of the simulation is reached) according to the following schedule of events:

1. Agents decide whether to cooperate or defect.
2. The payoff is calculated for each Agent.

The Tragedy of the Commons game differs from the two-player Prisoner's Dilemma in three important ways:

1. In the commons game, for a small number of players, the state of "minimally effective cooperation" (exactly *n* defectors) is not negligible, so there is not a dominating strategy.

2. In the commons game, using pure strategies, there are two Nash equilibria: everyone defecting (universal defection) and exactly *n* defectors (minimally effective cooperation).

3. In the two-player Prisoner's Dilemma, universal cooperation is a Pareto optimal outcome since no player can be better off without making the other player worse off. However, in the commons game the only Pareto optimal outcome is the state of minimally effective cooperation.

---

3 A certain strategy dominates another if the former provides the actor with a higher payoff for every possible combination of the other players' strategies.

4 An equilibrium is deficient if there exists another equilibrium which is preferred by every player.

5 An outcome is Pareto inefficient if there is an alternative in which at least one player is better off and none of the rest of the players is worse off.

6 If there is an outcome with the property that no individual can get a higher expected payoff by switching to another strategy while the other individuals keep their strategies unchanged, then that outcome constitutes a Nash equilibrium.

3. The decision making algorithm

This section describes the decision making algorithm of Agents who are rational in the sense that, given their belief-desire system, they optimise their choice. Defining rational behaviour in games is by no means easy. The outcome of a player's decision in a game depends on the other players' decisions, and these decisions cannot be *rationally* inferred a priori. In fact, it seems sensible to think that everyone's decisions could depend on everyone's previous decisions. To overcome that problem, a usual approach in game theory is to assume instrumental rationality and common knowledge of rationality (Aumann, 1976) and/or to focus on equilibria. We want neither to assume common knowledge of rationality nor to study only the existence of possible equilibria. While we find it useful to explore how rational agents would behave in social dilemmas, we believe that common knowledge of rationality is such a strong assumption that can hardly be sustained in a useful model of any real-world situation. Similarly, we are also interested in how different equilibria are reached (if they are indeed reached) and how probable each of them is (or, synthesising that information, how probable a certain level of cooperation is).

Once we have decided to remove any a priori assumption of other players' behaviour from our Agents' decision making algorithm, we are bound to model other types of reasoning apart from (or in addition to) deductive arguments[7]. Inferences about other players' strategies or about future payoffs can then only be made in the light of the history of the game, and therefore they can only lead to probable −rather than necessarily true− conclusions. There are many types of non-deductive arguments[8], such as generalisation by induction (from now on we will call it just induction), analogy, or abduction.

A possible decision making algorithm could consist in a) inducing general rules about future payoffs or about other players' strategies by observing several particular cases in the history of the game, and b) deducing future payoffs or the other players' possible actions in particular situations from the induced general rules. However, this would imply assuming that the other players do actually have a pre-planned fixed strategy, and a rational Agent should not necessarily have to assume that.

Our Agents use a form of analogy in their decision-making algorithm. According to the Collins English dictionary (2000), analogy is "a form of reasoning in which a similarity between two or more things is inferred from a known similarity between them in other respects". In the context of problem solving, analogy can be defined as the process of reasoning from a solved problem which seems similar to the problem to be solved

[7] Valid deductive arguments are those whose conclusion follows from its premisses with logical necessity (Copi, I. M., 1986, p. 169).

[8] Non-deductive arguments are not claimed to demonstrate the truth of their conclusions, but to support their conclusions as probable, or probably true. This type of arguments are called inductive arguments in some textbooks (e.g. (Copi, I. M., 1986, p. 404) or (Salmon, 1984, p. 32)) and should not be confused with the inference process of generalisation by induction (which is a particular type of non-deductive argument).

(Bundy, 1997, p. 4). When analogical reasoning is undertaken within a single domain it is usually called Case-Based Reasoning (Aamodt and Plaza, 1994; Bundy, 1997, p. 14)). Our Agents use Case-Based Reasoning (CBR) in their decision making algorithm. CBR arouse out of research into cognitive science in the late 1970s (Schank and Abelson, 1977; Schank, 1982). It basically consists of "solving a problem by remembering a previous similar situation and by reusing information and knowledge of that situation" (Aamodt and Plaza, 1994). The "known similarity" in the first definition of analogy is the problem situation in case-based reasoning, whereas the "inferred similarity" is the causal link between solution and outcome. The rationale is that if a solution turned out to be satisfactory when applied to a certain problem it might work in a similar situation too. There are several psychological studies that provide support for the importance of this problem-solving process in human reasoning (see Ross (1989) for a summary). There is also a number of industrial applications of CBR (Watson, 1997), particularly in domains where there is a need to solve ill-defined problems in complex situations; in such situations, it is difficult or impossible to completely specify all the rules (if they exist at all) but there are cases available.

In CBR, a case is a contextualised piece of knowledge representing an experience (Watson, 1997). In general the experience could be the Agent's own, or a neighbour's. In the latter case we would be implementing a type of social learning (Conte and Paolucci, 2001). However, in the experiments reported in this paper Agents only use cases that they have experienced themselves.

The following describes the design of an Agent who, when facing a situation in which they must choose between two alternatives, they recall the most recent similar situation in which they made each of the two possible decisions, and select the one that turned out to be better for them.

A case for an Agent, i.e. the experience they lived in time-step t, comprises:

1. The time-step t when the case occurred.

2. The state of the world at the beginning of time-step t, characterised by the factors that the Agent considers relevant to estimate the Payoff. These are:

    a. Descriptor 1 (D1): the number of other defectors.
    b. Descriptor 2 (D2): the decisions that the Agent made.

    Agents are able to remember b time-steps back (e.g. if $b = 2$, the state of the world for the Agent will be determined by the number of other defectors and the decisions made, both in time-step t-1 and in time-step t-2).

3. The decision they made in that situation, i.e. whether they cooperated or defected in time-step t, having observed the state of the world in that same time-step.

4. The Payoff that they obtained after having decided in time-step t.

The case representing the experience lived by Agent *A* in time-step *t* would then have the following structure:

t	$n_{-b} \ldots n_{-2} \;\; n_{-1}$ $d_{-b} \ldots d_{-2} \;\; d_{-1}$	d_0	p_0

where

n_i is the number of defectors (excluding agent *A*) in time-step $(t + i)$,

d_i is the decision made by Agent *A* in time-step $(t + i)$, and

p_0 is the Payoff obtained by Agent *A* in time-step *t*.

The number of cases that Agents can keep in memory is unlimited. It is also worth noting that no cases are available for any Agent until $(b + 1)$ time-steps have gone by in the simulation.

Agents make their decision whether to cooperate or not by retrieving two cases: the most recent appropriate case for each of the two decisions (each of the two possible values of d_0). A case is said to be appropriate if and only if its state of the world is a perfect match with the current state of the world observed by the Agent holding the case[9] (we will also denote this by saying that the two situations are *similar*).

In a particular situation (i.e. state of the world) an agent must face one of the following three possibilities:

1. The Agent cannot recall any previous situations that match the current state of the world. In CBR terms, the agent does not hold any appropriate cases for the current state of the world. We distinguish three different possibilities here:
   a. The Agent defects. This model will be called D-Model from now on.
   b. The Agent cooperates. This model will be called C-Model from now on.
   c. The Agent chooses at random. This model will be called R-Model from now on.

2. The Agent does not remember a previous similar situation when they made a certain decision, but they do recall at least one similar situation when they made the other decision. In CBR terms, all the appropriate cases the Agent recalls have the same value for d_0. In this situation, Agents will explore the decision that they did not make in the previous similar situation (decision by exploration).

3. The Agent remembers at least one previous similar situation when they made each of the two possible decisions. In this situation, the Agent will focus on the most recent case for each of the two decisions and choose the decision that provided them with the higher Payoff[10] (decision by comparison). In this way, Agents adapt

[9] The state of the world that different Agents perceive at a given time-step is not necessarily the same, since it includes the number of *other* defectors.

[10] A tie is impossible in any of the two games reported in this paper.

their behaviour according to the most recent feedback they got in a similar situation.

As we pointed out above, an alternative to CBR would be a rule-based system. One could induce the appropriate generalisations (rules) from the cases, and, in this view, CBR can be seen as a postponement of induction (Loui, R., 1999). However, when dealing with systems that are adaptive themselves (in the sense that they are constituted by adaptive agents), the "rules" of the system vary as the system evolves and therefore agents must constantly revise their perceptions about the system. This could be done by constantly updating the set of induced rules or by using CBR. Agents who use CBR store the original cases without building rules that summarise them. In that way, cases can suggest solutions even to ill-defined problems, such as those arising in social dilemmas, for which there may not be an adequate set of rules.

For an extensive review of agent-based simulation in the study of social dilemmas, see (Gotts, Polhill, and Law, 2003a).

4. Results and discussion

The source code of the model presented here is available upon request, along with all parameter files used to generate results in this paper.

4.1. D-Model and C-Model

Agents in the D-Model always defect when they face a completely new situation. The motivation to implement this model was to study the outcome of the game when played by agents with the following characteristics:

1. They only pursue their own benefit.
2. They are aware of the fact that other players' actions could depend on the agent's own actions. Therefore they explore the full range of possible actions under different circumstances and select the one that turned out to be best in the past.
3. In the process of exploring a new situation, they defect first, which is the decision that, given the other players' actions, offers a higher payoff under almost all circumstances.

The motivation to implement the C-model was to study the outcome of the game when played by agents with characteristics 1 and 2 above, but who prefer to start cooperating when facing a new situation.

The outcome of both the D-Model and the C-Model is stable full cooperation for both the Prisoner's Dilemma and the Tragedy of the Commons games, for any value of $b>0$ if the models are run for long enough. The explanation is simple. These two models have no random component at all, every agent has the same decision-making algorithm, and the game is symmetric. Therefore every agent will make the same decision at any given time[11]. The range of possible outcomes of a one-shot game is then reduced to either

[11] When Agents cannot recall any cases, they make the same decision by model design. When agents can recall only one case in which they either cooperated or defected, they all recall the same case, and therefore

universal cooperation or universal defection. Universal cooperation gives a higher payoff than universal defection and therefore will be preferred for any given state of the world. The situation of stable universal cooperation is then reached sooner or later (depending on the actual value of b).

It is interesting to realise that if agents tried to anticipate other agents' *actions* they would always defect, since defection is always preferred given the other agents' actions[12]. However, when agents try to anticipate the *outcome* of their actions, they are implicitly recognising that their own actions might affect other agents' decisions, and the result in this specific case is stable universal cooperation.

4.2. R-Model

The Prisoner's Dilemma (PD)

As it might be expected, the R-Model is very sensitive to the decisions that are made at random. Since the model has stochastic components, the results for a given set of parameters cannot be given in terms of assured outcomes but as a range of possible outcomes, each of them with a certain probability of happening. The probability of each outcome can either be estimated by running the model several times with different random seeds or, under certain circumstances, it can also be exactly computed.

Figure 1 shows the results obtained in the Prisoner's Dilemma when played by two agents with the same Backwards Memory b. Since the Backwards Memory is finite, the number of possible states of the world is finite and Agents must end up in a cycle. That cycle is necessarily made up of universal defections and/or universal cooperations, since if an Agent receives the sucker's payoff in any situation, they will never cooperate again in that situation. The cycle is at most 2^b time-steps long. To study how often Agents cooperate we define the 'cooperation rate' as the number of times universal cooperation is observed in a cycle divided by the length of the cycle.

A common representation of the state of the world

The difference in terms of cooperation rate between the complete representation of the state of the world (D1&D2) and the two incomplete representations of the state of the world (D1, and D2) is clear and it becomes larger the greater the value of Backwards Memory b is. When both the Agent's own decisions and the other player's decisions form the state of the world (D1&D2) the cooperation rate is much higher than when the definition of the world is incomplete. The reason for this is the following: for a certain state of the world, Agents will cooperate only when the payoff in the case when they cooperated is *Reward* and the payoff in the case they defected is *Punishment*. They will defect if they recall any other payoff for any of the two decisions. This basically means that the emergence of cooperation requires a high degree of coordination (both Agents

---

they all make the same decision. Finally, when agents can recall one case for each possible decision, those cases necessarily show a situation in which all of them cooperated or all of them defected. When given universal cooperation and universal defection as the only two possible options, they all cooperate.

[12] Except in the case of minimally effective cooperation in the "Tragedy of the Commons" game.

must have defected simultaneously and both Agents must have cooperated simultaneously for a certain state of the world). If there is lack of coordination, the Agent that receives the *Sucker* payoff will never cooperate again when they find themselves in the same state of the world. In this sense, our Agents are particularly unforgiving. When the state of the world is defined by both D1 and D2, both Agents share the same view of the state of the world in the sense that any two situations that look the same for one Agent will also look the same for the other Agent and any two situations that look different for one Agent will also look different for the other Agent. This shared view of the state of the world explains that, for a certain state of the world, the only relevant factor is the random decision that they make when they first experience that situation. If both Agents make the same decision, they will both decide to cooperate when they experience that same situation for the third time and onwards. Similarly, if they make a different decision, they will both decide to defect when they experience that same situation for the third time and onwards.

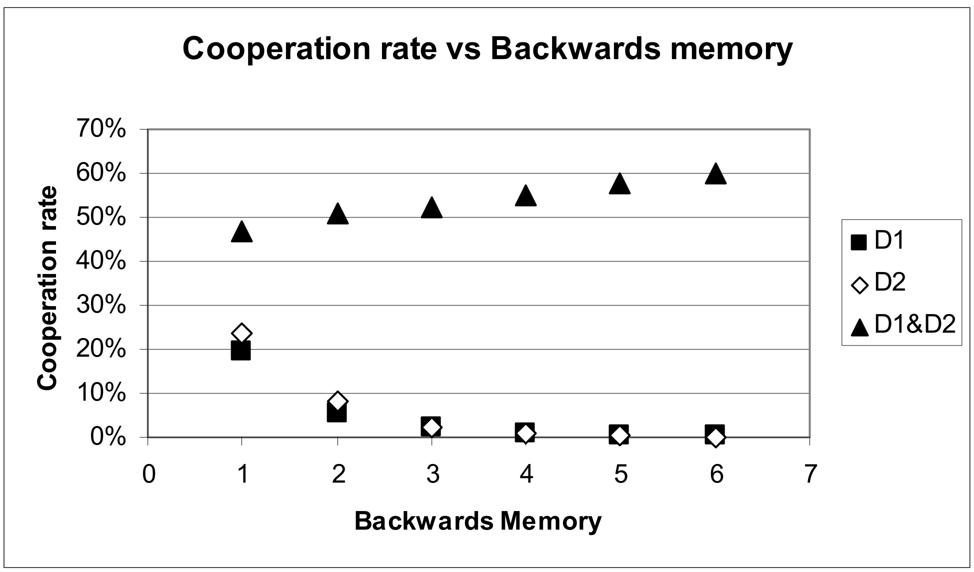

Figure 1. Cooperation rates when modelling two agents with Backwards Memory *b* and with 3 different representations of the state of the world (D1, D2, and D1&D2), playing the PD. D1 is the state of the world formed by the number of other defectors, D2 represents the state of the world constituted by the agent's own decisions, and D1&D2 represents the state of the world made up by both D1 and D2. The cooperation rate shows the probability of finding both agents cooperating once they have finished the learning period (i.e. when the run locks into a cycle). The values represented for b = 1 and b = 2 have been computed exactly (except D1&D2 and b = 2). The rest of the values have been estimated by running the model 1000 times with different random seeds. All standard error are less than 1.2%.

However, when using either of the two incomplete definitions of the state of the world, there are many situations that are represented by the same state of the world for one Agent but by a different state of the world for the other player. The number of such situations increases as the Backwards Memory *b* increases. In particular, the situations of universal defection and universal cooperation are not clearly identified by the Agents. This causes a lack of coordination that greatly reduces the chances of ending up in a cooperative outcome.

The role of Backwards Memory

After having seen that our Agents will eventually cooperate in a given situation if they happened to make the same decision the first time they experienced that situation and will defect otherwise, it comes as no surprise to see that cooperation rates are about 50% (Figure 1). What is puzzling is that cooperation rates increase as the value of Backwards Memory *b* increases, and become significantly higher than 50% for high values of *b* (the standard errors of the cooperation rate averages for $b = 5$ and $b = 6$ are less than 1%). The explanation for this is by no means obvious, but we can give some insights. Let stable universal cooperation (SUC) be the cycle in the simulation in which the two Agents are always cooperating, and let stable universal defection (SUD) be the cycle where the two Agents are always defecting. The first thing we shall do is to show that the relationship between the value of Backwards Memory and the relative frequency of the states of SUC and SUD accounts for much of the relationship between the Backwards Memory and levels of cooperation. To do that, we define the incomplete cooperation rate, which is calculated excluding the runs that ended up in either SUC or SUD from the calculation of the cooperation rate. Figure 2 shows that, although SUC and SUD are not responsible for all the relationship between Backwards Memory and levels of cooperation, they do account for a great part of it. Therefore, we will focus on these two situations in our explanation from now on.

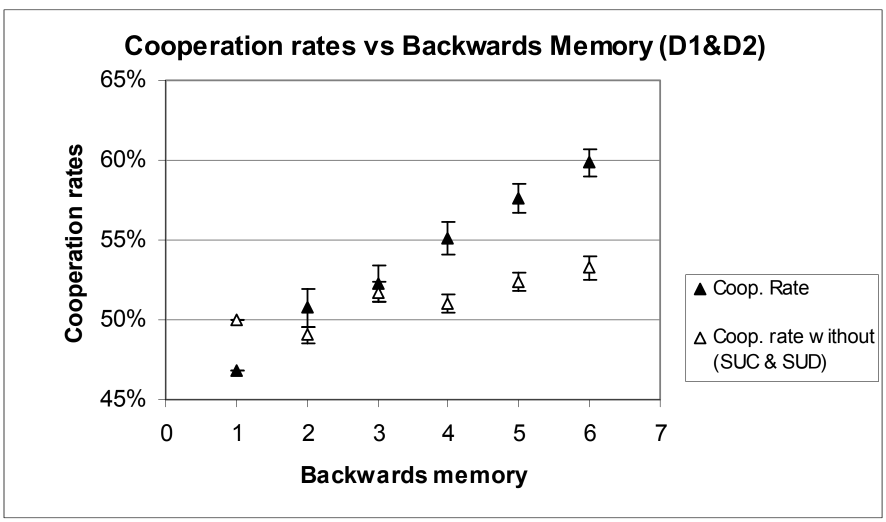

Figure 2. Cooperation rates and incomplete cooperation rates when modelling two agents with Backwards Memory *b* and D1&D2 as their representation of the state of the world, playing the PD. The incomplete cooperation rates are calculated excluding from the (complete) cooperation rates the runs that ended up in either SUC or SUD. The values shown have been estimated by running the model 1000 times with different random seeds. Standard error bars are shown.

Figure 3 shows that there is a clear relationship between the value of Backwards Memory *b* and the relative frequency of SUC and SUD. The higher the value of *b*, the more likely SUC is comparing it with SUD. For example, for $b = 6$, SUC is more than 3 times more likely than SUD.

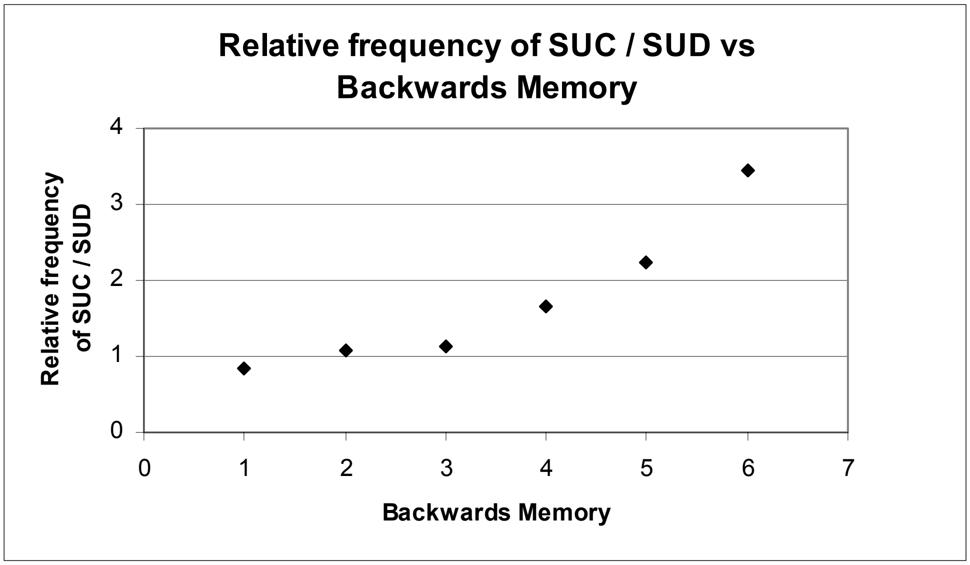

Figure 3. Number of runs that ended up in Stable Universal Cooperation (SUC) divided by the number of runs that ended up in Stable Universal Defection (SUD) for different values of Backwards Memory *b*. Conditions are the same as in Figure 2. The effect of Backwards Memory on the relative frequency of SUC / SUD is significant at the 0.001 level (chi-square hypothesis test of independence).

We will now give some insights on the reasons that make SUC more likely than SUD as the Backwards Memory increases. Let *bUC* be the state of the world formed by *b* universal cooperations and let *bUD* be the state of the world formed by *b* universal defections. As it is shown in Figure 4, in order to arrive at SUC, *bUC* must be visited three times, and the same applies for SUD and *bUD*: Agents choose randomly the first time they experience a situation, they decide by exploration the second time (switching their first decision), and by comparison the third time (repeating the more profitable decision). We now demonstrate the existence of two opposing forces: a) In general *bUD* will be visited for the first time before *bUC* is (see Figure 5), but b) assuming that both *bUD* and *bUC* have been visited once, it is more likely that *bUC* will be visited for the third time before *bUD* will. As we show below, force a) is strong for low values of *b* but loses relevance (compared to force b)) as *b* increases. On the contrary, force b) is weak for low values of *b* but gets stronger as *b* increases. This explains why cooperation rates are higher as the value of *b* increases.

Force a) is illustrated in Figure 5. If $b = 1$, *bUD* = DD and *bUC* = CC. *bUD* will be slightly more likely to be visited for the first time before *bUC* = CC because both can appear by chance with equal probability, but *bUD* can also be reached under additional circumstances; if the agents keep making (randomly and by exploration) opposing decisions, they will end up both defecting by comparison (e.g. the chain of events: DC (randomly), DC (randomly), CD (by exploration), DC (randomly), DD (by comparison)). If $b = 2$, *bUD* = DD-DD, and a similar reasoning applies. In order to get DD-DD, it is necessary to get DD first, and in order to get CC-CC it is necessary to get CC first. We have shown that DD is more likely to appear for the first time than CC is. Given that CC and DD have already appeared, it is also more likely to get DD-DD before CC-CC for the same reason as above: both *bUD* and *bUC* can be reached by chance with equal probability given DD and CC respectively, but *bUD* can also be reached by comparison if opposing decisions have been taken by the agents. The same reasoning applies to higher values of *b*, but the effect becomes less important as the Backwards Memory increases.

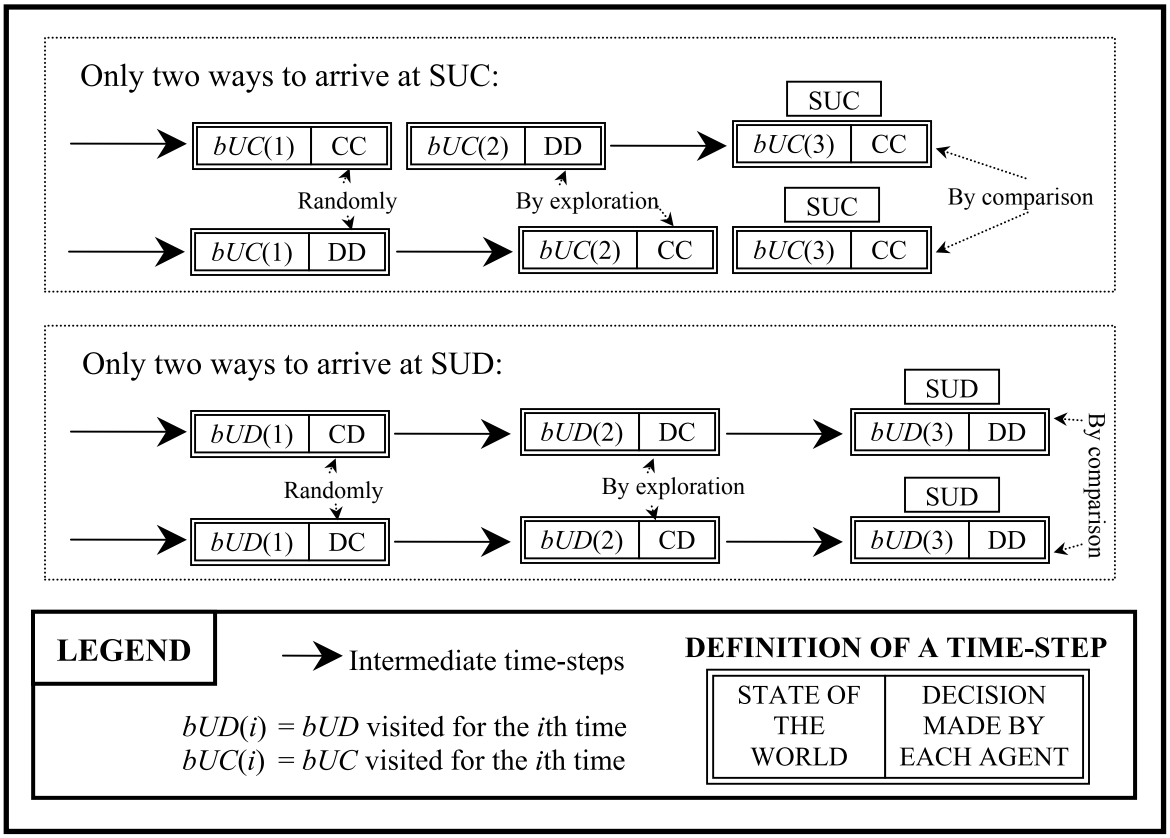

Figure 4. Different ways in which a run can reach SUC or SUD.

Force b) is more important for higher values of *b*. As it is shown in Figure 4, once *bUC* has been visited once, it only requires one set of intermediate time-steps (represented as arrows in Figure 4) to arrive at SUC (assuming that the decisions made by the agents were the same; otherwise, SUC will never happen). This is because *bUC*(t+1) necessarily follows immediately after *bUC*(t) if both Agents cooperate when observing *bUC*(t). On the other hand, once *bUD* has been visited once, it requires two sets of intermediate time-steps to arrive at SUD (assuming that the decisions made by the agents were different; otherwise, SUD will never happen). As the value of Backwards Memory *b* gets higher, the number of possible different states of the system grows exponentially, and it becomes more and more difficult to revisit a certain state. Therefore, in runs where both SUC and SUD are possible (i.e. where CC or DD have been chosen after *bUC*, and DC or CD have been chosen after *bUD*), SUC is more likely to be reached than SUD. Figure 6 illustrates this effect.

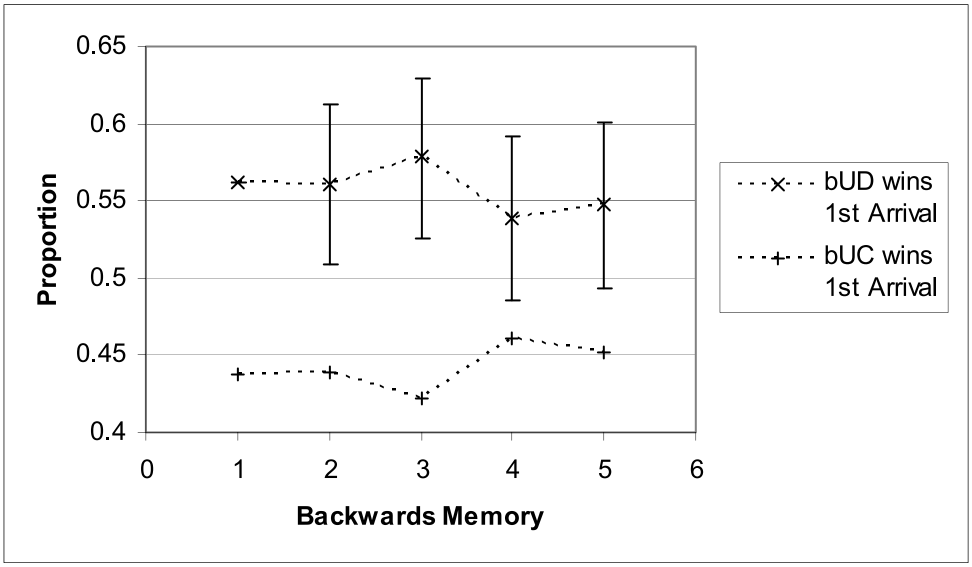

Figure 5. Among 1000 runs, of all the runs in which either *bUC* or *bUD* were visited at least once, the figure shows the proportion of them in which *bUC* was visited before *bUD* (*bUC* wins first arrival), and the proportion of them in which *bUD* was visited before (*bUD* wins first arrival). The 99.9% confidence interval for bUD is also represented.

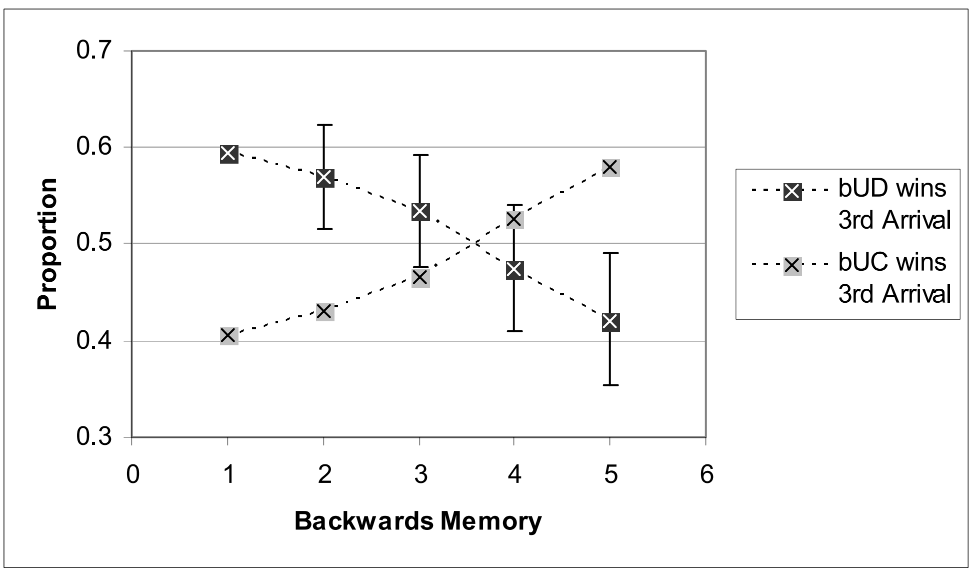

Figure 6. Among 1000 runs, of all the runs in which either *bUC* or *bUD* were visited at least three times, the figure shows the proportion of them in which *bUC* was visited for the third time before *bUD* was (*bUC* wins third arrival), and the proportion of them in which *bUD* was visited for the third time before (*bUD* wins third arrival). The 99.9% confidence interval for bUD is also represented.

Figure 7 suggests that the relative frequency of SUC and SUD does not account for all the relationship between the cooperation rate and the Backwards Memory. This indicates that cycles with more universal cooperations are more likely than cycles with more universal defections. Exploratory research suggests that this is due to the fact that before Agents defect by comparison they have made opposing decisions, which cannot be part of a cycle. However, when Agents cooperate by comparison, they have always made the same decision, so those decisions can lead to and be part of cycles more easily.

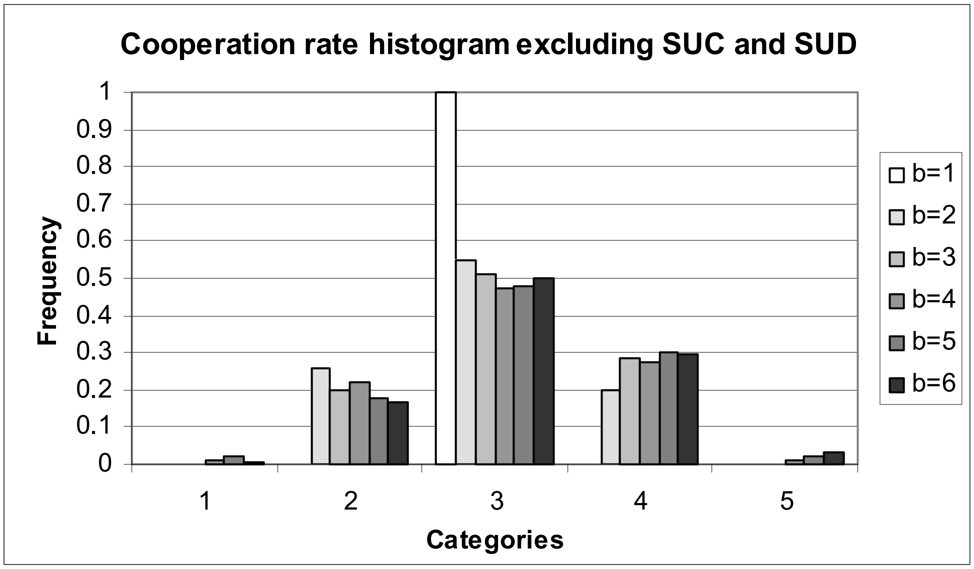

Figure 7. Histogram showing the frequency of five categories (*i* = 1, …, 5) of cooperation rates. Each category *i* includes cooperation rates between (*i*-1)·20% and *i*·20%. The frequency of cooperation rates which are equal to *i*·20% is shared between the adjacent categories. Runs which ended up in either SUC or SUD are excluded from this chart.

The Tragedy of the Commons game

Figure 8 shows the results obtained in the Tragedy of the Commons game when played by twenty five Agents with Backwards Memory b = 1, for different values of n (maximum number of defectors for which the reward is given). Again, since the Backwards Memory is finite, the number of possible states of the world is finite and Agents must end up in a cycle. The only situation in a cycle in which the reward is not given is universal defection, since if an Agent receives the *Coop-P* payoff in any situation they will never cooperate again in that situation. To study how often the reward is given we define the 'reward rate' as the number of times the reward is given in a cycle divided by the length of the cycle. Figure 9 shows the number of cooperators for some representative runs when n = 15.

As we can see in Figure 8, levels of cooperation strongly depend on the maximum number of defectors for which the reward is given (n). When the requirement is too demanding (low values of n), levels of cooperation tend to be low and the reward is not usually given. On the other hand, for moderate and high values of n, ($n \geq 15$), the reward is given almost always. Even in the case where Agents get the reward only if there are fewer defectors than half the number of participants (n = 12), the reward is given 52.3 % of the times.

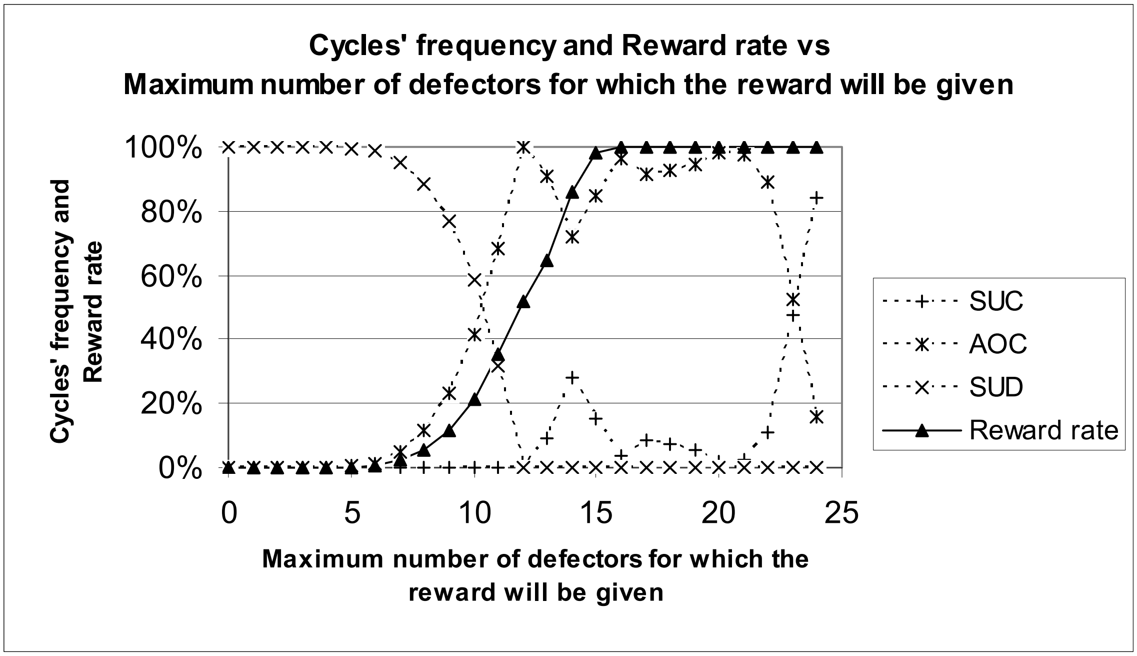

Figure 8. Reward rate (proportion of time-steps in which the reward is given once the learning process has finished) and proportion of runs that end up in Stable Universal Cooperation (SUC), Stable Universal Defection (SUD), or any other cycle (AOC), for different values of *n*, in the Tragedy of the Commons game played by 25 Agents with Backwards Memory *b* = 1. For each value of *n* the model has been run 1000 times. All standard errors for the reward rate are less than 1%.

It is clear then that cooperation can emerge from the interaction of selfish case-based reasoners.

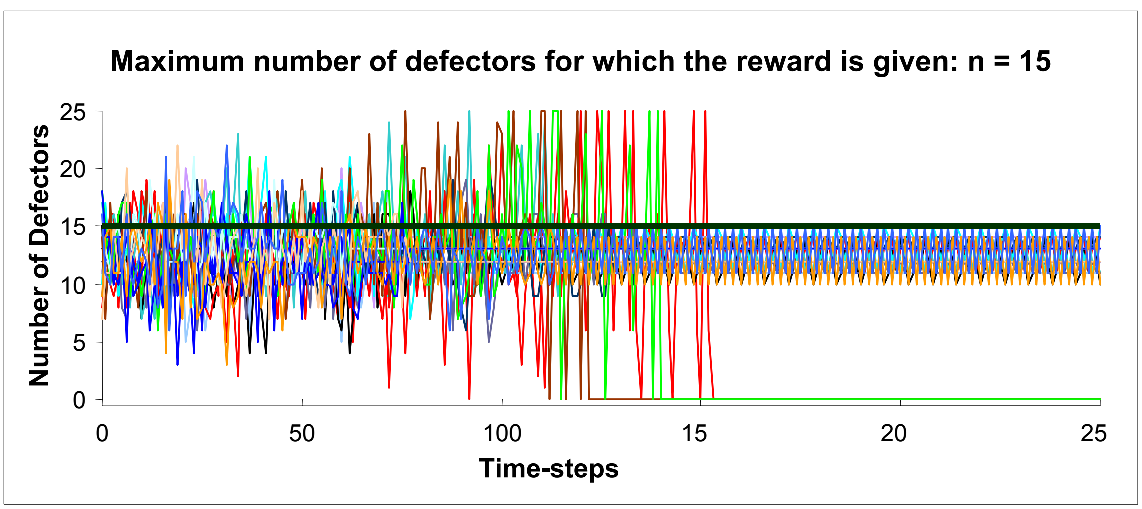

Figure 9. Number of defectors in the Tragedy of the Commons game, when played by 25 Agents with Backwards Memory *b* = 1, for different runs when *n* = 15. The straight thick bold line marks the maximum number of defectors for which the reward is given.

It is also interesting to see that for $n \geq 12$, Stable Universal Defection (SUD) does not occur in any case. In general, for a number of players p, if $n \geq (p-1)/2$ then SUD cannot occur[13]. A proof of this is given in appendix A.

5. Discussion

The results shown in section 4 have revealed a concept of equilibrium which is more relevant than the Nash equilibrium for case-based reasoners playing a repeated game. This new concept of equilibrium will be called CBR-equilibrium, and is defined for one single stage of any repeated game. Its defining property is that no player can be guaranteed a higher payoff by changing their decision[14]. The CBR-equilibrium is a weaker (i.e. less restrictive) concept than the Nash equilibrium: A Nash equilibrium is always a CBR-equilibrium but the reverse is not necessarily true. In particular, in the one-shot Prisoner's Dilemma, universal cooperation is a CBR-equilibrium while it is not a Nash equilibrium.

As opposed to the concept of Nash equilibrium (which makes the assumption that the other players will keep their strategies unchanged), the concept of CBR-equilibrium accounts for every possible action that the other players might take. The CBR-equilibrium is best defined by negation: if a certain player perceives that by changing their strategy, they will always get a higher payoff no matter the rest of the players' response, then they have a clear incentive to deviate from that outcome, so that outcome is not a CBR-equilibrium. If, on the contrary, no player has such an incentive, the outcome is a CBR-equilibrium. The CBR-equilibrium concept is coined to account for the effect that a certain player's action might have on the other players. A change of action will be undertaken necessarily only if there is certainty that it will lead to a higher outcome. Otherwise, we cannot assure that the player will change their decision. As an example, in the Tragedy of the Commons game the only CBR-equilibrium in which the reward is not given is universal defection, and all the outcomes in which the reward is given are CBR-equilibria.

We now define a broad set of decision–making algorithms used to select one action from a finite set of them (e.g. Cooperate or Defect). Let us call Cased-Based Reasoning Exploratory (CBRE) algorithms those decision-making algorithms in which an action is selected by undertaking the following steps:

1. Identify a set of possible outcomes $\mathbb{S}$ in which every action available to the decision maker is represented at least once.
2. Select an outcome that provides the decision maker with a payoff which is not lower than the decision maker's payoff in any other outcome in the set $\mathbb{S}$.

---

[13] It is assumed that $n < p$. Otherwise the reward is always given and there is no social dilemma.

[14] A slightly more restrictive definition of CBR-equilibrium would be: an outcome in which no player can be guaranteed the same or a higher payoff by changing their decision. This second definition would imply that Agents deviate from an outcome if it is certain that they will not be worse off by doing so, whereas the first definition implies that Agents move away from an outcome only if it is certain that they will be better off by doing so. The more restrictive definition of CBR-equilibrium is neither weaker nor stronger than the Nash equilibrium.

3. Return the action undertaken by the decision maker in the outcome selected in step 2.

The decision-making algorithm presented in section 3 becomes a CBRE algorithm when the simulation locks into a cycle (and it necessarily does). In that particular case, $\mathbb{S}$ is made up by the most recent past experience for each of the actions.

We are now ready to state that: a) the outcome of any game played by Agents using a CBRE algorithm is necessarily a CBR-equilibrium and b) any CBR-equilibrium can be the outcome of a game played by Agents using a CBRE algorithm. These two statements are proved in Appendix B.

As we have seen in the previous section, the actual outcome of the game can be strongly path-dependent and depends on the specific type of CBRE algorithm that players use.

Game theorists have been able to provide several concepts that narrow the set of possible solutions in many game-theoretic models: from the well-known Nash equilibrium to the notion of perfect Bayesian equilibrium, which is the strongest concept of equilibrium in game theory (Gibbons, 1992). However, in their search for a reduced set of solutions, they had to pay a very high price: they had to make disturbing assumptions such as instrumental rationality, complete information, perfect information, or common knowledge of rationality, which are rarely observed in any real-world situation.

In this paper we propose a complementary approach, assuming that people *adapt* their behaviour according to their experience and look for outcomes that have proved to be best in the past. When playing repeated games they *explore* the different outcomes of the game rather than having a pre-planned detailed strategy calculated assuming other players' knowledge and behaviour. This new assumption has led us to a new concept of equilibrium, weaker than the Nash equilibrium, but that includes all the possible outcomes that can appear in a repeated game played by a broad class of case-based reasoners.

6. Economic payoffs are not everything

Several empirical studies have shown that theoretical predictions derived from the assumption of full rationality in CPR dilemmas fail to explain observed outcomes in many situations (Ostrom, Gardner, and Walker, 1994). There are people who behave in a cooperative way even though they are aware of the fact that it is not immediately advantageous in economic terms. One way of explaining this kind of behaviour is to assume that people have *multiple* utilities or values that determine their behaviour, and some of these utilities are associated with aspects of people's behaviour other than the economic payoffs they receive (e.g. morality). Jager (2000) reviews several factors that seem to influence behaviour in a CPR dilemma besides the economic payoffs.

In our design of a more advanced socio-economic Agent we consider two drivers of behaviour: economic payoffs, as described in the previous sections, and a simple form of social approval. Agents who cooperate disapprove of the Agents who are defecting. The approving/disapproving of other Agents takes place after phase 2 in the schedule of events described in section 2. Similarly, the number of Agents that disapproved of the decision making Agent in a certain time-step will be information added to each case.

The new case structure is the following:

t	n_{-b} … n_{-2} n_{-1} d_{-b} … d_{-2} d_{-1}	d_0	p_0 dn_0

where dn_0 is the number of disapproving Agents in time-step t.

Agents make now their decisions considering the two drivers of behaviour. This is implemented by introducing a new parameter: *Peer Pressure Threshold (PPT)*. The decision making algorithm of the R-model is modified by adding an extra phase to the process explained in section 3: when Agents remember at least one previous similar situation when they made each of the two possible decisions (possibility 3), they also take into account the expected number of disapproving Agents in the following period. If the expected number of disapproving Agents is greater than the PPT, they cooperate. Otherwise they make the decision that provided them with the higher Payoff.

To study the effect of the Agents' new concern for social approval we put 25 Agents to play the Tragedy of the Commons game. Results of running the model 100 times for given values of n and PPT are shown in Figure 10.

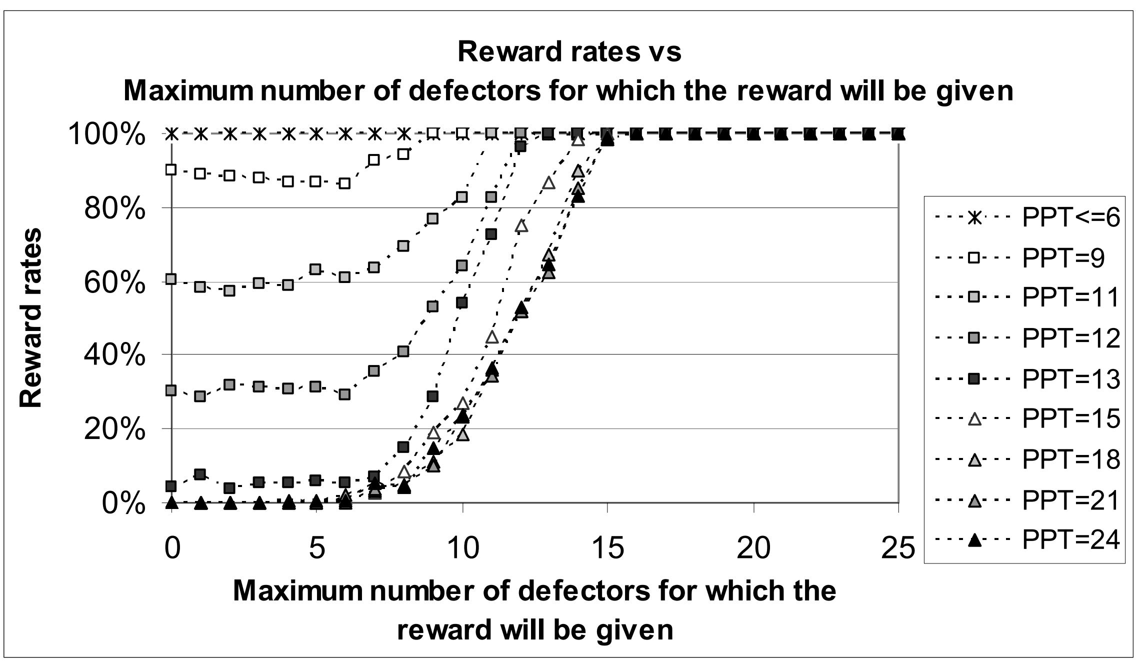

Figure 10. Reward rates for different values of *n* and *PPT* in the Tragedy of the Commons game played by 25 Agents with Backwards Memory *b* = 1. Each value has been estimated running the model 100 times.

Results indicate that there is a critical range of values of *PPT* ($9 \le PPT \le 15$) inside which a slight variation of PPT can have a major impact on the reward rate. The reward rate is quite insensitive to the value of PPT outside this critical range. For $PPT > 15$, the results are very similar to the case where Agents are completely selfish ($PPT = 24$), whereas for $PPT < 9$ the reward rate is close or equal to 100% for any value of n. When Agents are not prepared to be disapproved of by half or more of the other players ($PPT = 12$), the reward rate dramatically increases. As might have been expected, for even lower levels of PPT, the reward rate keeps increasing until PPT = 0, in which case we are forcing the Agents to cooperate as soon as they encounter the same state of the world three or more times.

7. Conclusions

We have explored the outcome of social dilemmas when played by case-based reasoners. Case-based reasoning is a method of inference that is believed to be commonly used by real people when they face ill-defined problems in which they cannot easily compute a satisfactory solution. Social dilemmas are clearly ill-defined problems since the payoff for any player depends on the other players' actions and these actions are not necessarily known by the deciding agent nor can they be rationally inferred a priori. However, when playing the game repeatedly, agents can adapt their behaviour by observing the other players' actions, and find a satisfactory solution within the constraints that the other players' actions impose. By implicitly inferring each others' behaviour, selfish case-based reasoners arrive at a cycle in which all of them can justify every decision they make by appealing to a previous past experience. In this paper we have proved that the decision they make is very often to cooperate, even though they only pursue their own benefit. The experiments conducted have revealed a new concept of equilibrium, characterised by the fact that no player can be guaranteed a higher payoff by changing their decision. There is a broad class of CBR decision-making algorithms that lead to cycles made up by such equilibria.

We have also explored the behaviour of more advanced socioeconomic Agents, which are not only motivated by personal gain, but they also take into account the other players' opinion. In our model, small doses of social approval are not enough to change the outcome of the game, but if Agents are not prepared to be disapproved of by more than half the other players, levels of cooperation dramatically increase.

Acknowledgement

This work is funded by the Scottish Executive Environment and Rural Affairs Department.

Appendix A

In the Tragedy of the Commons game, for any number of players $p>2$ following the R-model decision making algorithm (see section 3) and for any value of Backwards Memory $b>1$, it is true that:
If the maximum number of defectors for which the reward is given n ($n < p$) is greater or equal to $(p-1)/2$, then Stable Universal Defection cannot be an outcome of the model.

Proof.
Let *bUD* be the state of the world formed by b universal defections. In *bUD*,
$n_{-i} = p - 1$ for $i = 0,1, \ldots, b$
d_{-i} = defect for $i = 0,1, \ldots, b$

The state of the world *bUD* is perceived by all the Agents after b universal defections and only then. Therefore, and contrary to what happens with other states of the world, at any given time in the simulation, every agent remembers having experienced the situation defined by *bUD* the same number of times.

Let x be the number of defectors there are in the time-step immediately after the Agents experience *bUD* for the first time. Therefore $(p - x)$ is the number of defectors there are in the time-step immediately after the Agents experience *bUD* for the second time. Only one of the following three possibilities can occur in the time-step immediately after the third time that the Agents arrive at *bUD*:

a) If $(x > n)$ and $(p - x > n)$, no rewards have been given in the two previous similar situations, so agents will defect always from then on (Stable Universal Defection).
b) If $(x \le n)$ and $(p - x \le n)$, Agents have received the reward in the two previous similar situations and therefore they all defect in the following time-step. That would be then the fourth time that Agents arrive at *bUD*, but then they would remember than the last time they defected in a previous similar situation they got *Def-P* since they all defected, and the last time they cooperated they got (*Coop-P* + *Reward*). From then on, every time Agents arrive at *bUD* they will all cooperate, so Stable Universal Defection will never happen.
c) When the reward is given only in one of the two times when agents arrived at *bUD*, Agents will repeat the decision they made when they got the reward. Therefore, from then on, every time Agents arrive at *bUD* they will repeat the same decision that led them to the reward, and Stable Universal Defection will never occur.

We have proved then that $(\boldsymbol{x > n})$ and $(\boldsymbol{p - x > n})$ are two necessary conditions to arrive at Stable Universal Defection. Now we will prove that if $n \ge (p-1)/2$, those two conditions cannot be true at the same time.

$n \ge (p-1)/2 \quad \leftrightarrow \quad p - n \le n + 1$ {1}
$p - n \le n + 1$ and $\boldsymbol{p - x > n} \quad \rightarrow \quad x < n + 1$ {2} (premiss from {1})
$\boldsymbol{x > n}$ {3}

Since x and n are integers by definition, {2} and {3} cannot be simultaneously true. Therefore it is proved that if $n \geq (p\text{-}1)/2$, then Stable Universal Defection cannot be an outcome of the model.

Appendix B

Statement a):
The outcome of any game played by Agents using a CBRE algorithm is necessarily a CBR-equilibrium[15].

Proof.
If an Agent has selected a certain action using a CBRE algorithm, it means that for any other feasible action A there exists an outcome in which the decision-making Agent selects A and gets an equal or lower payoff. It is true then that no player can be guaranteed a higher payoff by switching to another action. Therefore the outcome is a CBR-equilibrium.

Statement b): Any CBR-equilibrium can be the outcome of a game played by Agents using a CBRE algorithm.

Proof.
The statement is proved by construction. At a given CBR-equilibrium, any player can identify at least one outcome for each of the non-selected actions in which their payoff would be equal or less than their payoff at the CBR-equilibrium. For each player i, let $\mathbb{T}_i$ be the set of outcomes made up by (1) the given CBR-equilibrium and (2) one outcome for each of the non-selected actions in which the payoff for the decision maker is equal or less than the payoff for the decision maker at the CBR-equilibrium. For each player i, let D_i be the decision making algorithm that uses the set of outcomes $\mathbb{T}_i$ and returns the action selected by player i in the CBR-equilibrium. For any i, D_i is a particular type of CBRE algorithm and the outcome of the game when played by Agents using D_i would be the given CBR-equilibrium.

References

Aamodt, A. and Plaza, E., 1994. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *AI Communications*. IOS Press, Vol. 7, no. 1, pp. 39-59.

[15] If we adopt the alternative definition of CBR-equilibrium presented in footnote 14, then statement a) is only true for games in which players never get the same payoff when they select different actions. The Prisoner's Dilemma and the Tragedy of the Commons game satisfy such condition.

Aumann, R., 1976. Agreeing to disagree. *Annals of Statistics*, no. 4, pp. 1236-1239.

Bundy, A. (ed.), 1997. *Artificial Intelligence Techniques: A Comprehensive Catalogue*. Springer-Verlag, fourth revised edition.

Collins, 2000. *The Collins English Dictionary*. HarperCollins Publishers.

Conte, R., Hegselmann, R., and Terna, P. (eds.) 1997. Simulating Social Phenomena. *Lecture Notes in Economics and Mathematical Systems* 456. New York: Springer-Verlag.

Conte, R. and Paolucci, M., 2001. Intelligent social learning. *JASSS - the Journal of Artificial Societies and Social Simulation*, Vol. 4, no. 1.
URL = <http://jasss.soc.surrey.ac.uk/4/1/3.html>

Copi, I. M., 1986. *Introduction to Logic*. Seventh Edition. Macmillan, New York.

Dawes, R. M., 1980. Social dilemmas. *Annual Review of Psychology*, pp. 161-193.

EU, 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 Establishing a framework for Community action in the field of water policy. *Official Journal*, L 327: pp. 0001-0073.

Gibbons, R., 1992. *A Primer in Game Theory*. FT Prentice Hall, Harlow (England).

Gotts, N. M., Polhill, J. G., and Law, A. N. R., 2003a. Aspiration Levels in a Land Use Simulation. *Cybernetics and Systems*. In Press.

Gotts, N. M., Polhill, J. G., and Law, A. N. R., 2003b. Agent-based simulation in the study of social dilemmas. *Artificial Intelligence Review*, 19, pp. 3-92.

Hardin, G., 1968. The Tragedy of the Commons. *Science* 162, pp. 1243-1248.

Jager, W., 2000. *Modelling consumer behaviour*. Doctoral thesis. Groningen: University of Groningen, Centre for Environmental and Traffic Psychology.

Kahnemann, D., Slovic, P. and Tversky, A., 1982. *Judgement Under Uncertainty: Heuristics and Biases.* Cambridge University Press.

Kuhn, S., 2001. Prisoner's Dilemma. *The Stanford Encyclopedia of Philosophy (Winter 2001 Edition)*, Edward N. Zalta (ed.),
URL = <http://plato.stanford.edu/archives/win2001/entries/prisoner-dilemma/>

Loui, R., 1999. Case-Based Reasoning and Analogy. *The MIT Encyclopedia of the Cognitive Sciences*, Wilson, R. A. and Keil, F. C. (eds.), pp. 99-101.

Ostrom, E., Gardner, R., and Walker, J. M., 1994. *Rules, games, & common-pool resources*. Ann Arbor: University of Michigan Press.

Polhill, J. G., Gotts, N. M., and Law, A. N. R., 2001. Imitative Versus Non-Imitative Strategies in a Land Use Simulation. *Cybernetics and Systems*, 32, pp. 285-307.

Raub, W., 1988. Problematic Social Situations and the 'Large-Number Dilemma' - A Game-Theoretical Analysis. *The Journal of Mathematical Sociology*, 13: 4, pp. 311-357.

Ross, B., 1989. *Some psychological results on case-based reasoning*. In Hammond, K. (ed.), Proceedings of the Case-Based Reasoning Workshop, pp. 144-147 San Mateo. DARPA, Morgan Kaufmann, Inc.

Schank, R. and Abelson, R.P., 1977. *Scripts, Plans, Goals and Understanding*. Erlbaum, Hillsdale, New Jersey, US.

Schank, R., 1982. *Dynamic memory: a theory of reminding and learning in computers and people*. Cambridge University Press, Cambridge, UK.

Salmon, M., 1984. *Introduction to Logic and Critical Thinking*. Harcourt Brace Jovanovich.

Watson, I., 1997. *Applying Case-Based Reasoning: Techniques for Enterprise Systems*. Morgan Kaufman Publishers.