Manchester Metropolitan University Business School Working Paper Series

L.R. Izquierdo, N.M. Gotts, J.G. Polhill,
and B. Edmonds*
The Macaulay Institute, Aberdeen, UK
*Centre for Policy Modelling, Manchester
Metropolitan University

Social dilemmas: What if not everybody knows that everybody knows that everybody is rational?

WPS060

November 2004

ISSN 1478-8209

Authors

L.R. Izquierdo
N.M. Gotts
J.G. Polhill
B. Edmonds

Aytoun Building, Aytoun Street, Manchester M1 3GH, United Kingdom.
Email: b.edmonds@mmu.ac.uk

Abstract

In this paper social dilemmas are modelled as two-player games. In particular we model the Prisoner's Dilemma, Chicken and Stag Hunt. When modelling these games we assume that players *adapt* their behaviour according to their experience and look for outcomes that have proved to be satisfactory in the past. These ideas are investigated by conducting several experiments with an agent-based simulation model in which agents use a simple form of case-based reasoning. It is shown that cooperation can emerge from the interaction of selfish case-based reasoners. In determining how often cooperation occurs, not only what Agents end up doing in any given situation is important, but also the process of learning what to do can crucially influence the final outcome. Agents' aspiration thresholds play an important role in that learning process. It is also found that case-based reasoners find it easier to cooperate in Chicken than in the Prisoner's Dilemma and Stag Hunt.

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

1. Introduction

Strange as it may appear, there are many social interactions out there in the real world where the outcome that results from all individuals behaving rationally is undesirable for everybody. When this actually occurs we call the behavioural result a social dilemma. In a social dilemma, decisions that seem to make perfect sense from each individual's point of view can aggregate into outcomes that are unfavourable for all. Social dilemmas are at the heart of pollution and resource depletion problems but they are by no means exclusive to these situations: in any context where collective action can lead to a common benefit we may find that individuals are tempted to undermine the collective good for their own ends.

Game theory provides us with a useful framework to study social dilemmas. Game theory is a branch of mathematics devoted to the logic of decision making in social interactions (Colman, 1995, p. 3). It is not intended to account for how people *actually* behave, but for how **instrumentally rational**[1] players *should* behave in order to attain their clearly defined goals. In a game, each player must make a choice between two or more ways of acting (usually called strategies), and the **outcome** of the game depends on the choices of every player. Players have a clearly defined set of preferences among different outcomes; these preferences are represented by payoffs. In game theory, nothing is said about the origin of preferences, which could include any motivation whatsoever. Rationality is understood as a means to achieve one's goals, which are created at a stage where rationality plays no role. Using David Hume's words in *Treatise on Human Nature*, 'passions' motivate a person to act, and 'reason' is their servant or 'slave'. In some cases payoffs are measured on interval scales (hence giving information about relative preferences), but often ordinal scales are enough to perform the analysis of the game.

---

[1] Terms in bold are defined in Appendix A.

The most elementary formalisation of a social dilemma is the two-player Prisoner's Dilemma (PD). In the PD, each player can either cooperate or defect. Given any opponent's actions, both players are better off defecting; however, they both prefer bilateral cooperation to bilateral defection. The strategic nature of the PD is present in many situations in real life. For example, it appears when two states get into an arms race, when firms set prices in an oligopoly, and when we decide how much to use of a subtractable resource or whether to contribute to the provision of a public good.

When the PD is played once by instrumentally rational agents, the expected outcome is bilateral defection: rational players do not cooperate since there is no belief that a player could hold about the other player's strategy such that it would be optimal to cooperate (the cooperative strategy is **strictly dominated** by the strategy of defecting). Considering that both players would be better off if they both cooperated, this is a striking example of how rationality can be self-defeating.

The situation is very different when the PD is played repeatedly. In that case, the rational behaviour remains undefined if no assumptions about the other player's behaviour are made. For this reason, game theory incorporates not only rationality but also **common knowledge of rationality** (CKR), hence enabling players to make inferences about their opponent's behaviour. Assuming CKR is sufficient to prove that the outcome of the PD when played repeatedly *any* finite number of times is bilateral defection at every stage. Put differently, any two strategies which are an optimal response to each other necessarily lead to a series of bilateral defections in the finitely repeated game. However, when the number of rounds is not limited in advance, not even CKR is enough to narrow significantly the set of expected outcomes. Specifically, the "Folk Theorem" states that any **individually-rational outcome** can be a **Nash equilibrium** in the infinitely-repeated PD if the discount rate of future payoffs is sufficiently close to one.

The results when the game is played repeatedly raise concerns about:

a) The validity and appropriateness of assuming CKR. CKR is unsupported by empirical evidence, it leads to conclusions that clash with widely shared intuitions and empirical results, and some authors have argued that it might be internally incoherent (see, for example, Colman (2003) and Hargreaves Heap and Varoufakis (1995)).

b) The limitations of game theory in describing the dynamics that may lead to one among many possible equilibria.

These concerns have motivated several lines of research within the framework of game theory which relax the assumption of CKR and study backward looking alternatives to the deductive, forward-looking rationality of game theory. In this paper we adopt such approach. In particular, we explore the consequences of assuming that players, who have no *a priori* beliefs about their opponent's behaviour, *adapt* their own behaviour according to experience and look for outcomes that have proved to be satisfactory in the past. These ideas have been investigated conducting several experiments with an agent-based simulation model developed by Izquierdo *et al.* (2004) in which agents use a simple form of case-based reasoning. Case-based reasoners repeat those decisions that proved to be satisfactory in a similar past situation.

Izquierdo *et al.* (2004) used their model to simulate the PD and one of its *n*-player versions. They found that the outcome of any game played by a wide range of case-based reasoners for long enough would have to yield every player at least their ***Maximin***; they also developed the concept of iterated elimination of dominated outcomes, which we explain in section 5. In this paper, we extend the results in Izquierdo *et al.* (2004) by studying the behaviour of case-based reasoners in two other social dilemma games in addition to the PD: Stag Hunt and Chicken.

2. The games

The three 2-player games that we study in this paper can be represented using the payoff matrix shown in Table 1; they differ in the players' preferences over different outcomes.

Table 1. Payoff matrix for the PD, Chicken, and Stag Hunt. 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	Player 2: Defect
Player 1	Cooperate	*Reward* (Player 2) / *Reward* (Player 1)	*Temptation* (Player 2) / *Sucker* (Player 1)
	Defect	*Sucker* (Player 2) / *Temptation* (Player 1)	*Punishment* (Player 2) / *Punishment* (Player 1)

In the three games, players prefer any outcome in which the opponent cooperates to any outcome in which the opponent defects. In particular, both players prefer mutual cooperation to mutual defection (*i.e.* mutual defection is **Pareto deficient**). However, the temptation to cheat (if *Temptation* is greater than *Reward*) or the fear to be cheated (if *Sucker* is lower than *Punishment*) can put cooperation at risk. In Chicken the problem is greed but not fear (*Temptation* > *Reward* > *Sucker* > *Punishment*); in Stag Hunt, the problem is fear but not greed (*Reward* > *Temptation* > *Punishment* > *Sucker*); and finally, both problems coincide in the paradigmatic PD (*Temptation* > *Reward* > *Punishment* > *Sucker*). Because of the decision making algorithm of our Agents (explained in section 4), the actual values of the Payoffs are not relevant as long as they satisfy the mentioned ordinal relationships.

The Nash equilibria in the one-shot games are:

- PD: bilateral defection.
- Chicken: both unilateral outcomes (and a **mixed-strategy** equilibrium).
- Stag Hunt: both bilateral outcomes (and a mixed-strategy equilibrium).

When the PD is finitely-repeated under CKR, the only possible outcome is bilateral defection at every stage. When Chicken and Stag Hunt are finitely-repeated, any sequence of stage-game Nash equilibria is a Nash equilibrium[2] of the corresponding finitely-repeated game and many more Nash equilibria can appear if further assumptions about the payoffs cardinality are made. The last statement shows the limitations of game theory to define a small set of possible equilibria in Chicken and Stag Hunt, even under the assumption of CKR. Moreover, game theory cannot say anything about the dynamics that might lead players to one of many possible equilibria.

3. Case-based reasoning

Case-based reasoning is a type of analogical reasoning. Reasoning by analogy consists in inferring a similarity between two or more things 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 (Doran, 1997). When analogical reasoning is undertaken within a single domain it is usually called Case-Based Reasoning (CBR). CBR 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 rationale behind CBR is that if a solution turned out to be satisfactory when applied to a certain problem then it might work in a similar situation too.

CBR arose out of cognitive science research in the late 1970s (Schank and Abelson, 1977), and since then several psychological studies have provided support for its

---

[2] The same statement can be made substituting sub-game perfect Nash equilibrium of the finitely repeated game for Nash equilibrium of the finitely repeated game.

importance as problem-solving process in human reasoning, especially for novel or difficult tasks (see Ross (1989) for a summary).

Within the domain of economics, a case-based decision theory has been proposed by Gilboa and Schmeidler (1995, 2001). Gilboa and Schmeidler (1995) do not see case-based decision theory (CBDT) as a substitute for expected utility theory (EUT), but as a complement. They argue that CBDT may be more plausible than EUT when dealing with novel decision problems, or in situations where probabilities cannot easily be assigned to different states of the world (uncertainty, as opposed to risk), or if such states of the world cannot be easily constructed (ignorance).

4. The model

This section describes the design of Agents which use a simple form of CBR to decide whether to cooperate or defect. In CBR, Agents record all their experiences in the form of cases. Each case is a contextualised piece of knowledge representing an experience (Watson, 1997). A case for an Agent, *i.e.* the experience they lived in time-step *t*, comprises:

a) The time-step *t* when the case occurred.

b) The *perceived* 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:

    Descriptor 1: the opponent's decision.

    Descriptor 2: the Agent's own decision.

   Agents are able to remember *ml* time-steps back (*e.g.* if *ml* = 2, the perceived state of the world for the Agent will be determined by the opponent's decisions and the Agent's own decisions, both in time-step *t*-1 and in time-step *t*-2).

c) The decision the Agent 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.

d) The Payoff that the Agent obtained after having decided in time-step *t*.

Thus the case representing the experience lived by Agent *A* in time-step *t* has the following structure:

<table>
<tr><td rowspan="2">t</td><td>$od_{t-ml} \ldots od_{t-2} \quad od_{t-1}$</td><td rowspan="2">$d_t$</td><td rowspan="2">$p_t$</td></tr>
<tr><td>$d_{t-ml} \ldots d_{t-2} \quad d_{t-1}$</td></tr>
</table>

where

od_i is the opponent's decision in time-step *i*,

d_i is the decision made by Agent *A* in time-step *i*, and

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

The number of cases that Agents can keep in memory is unlimited. Agents make their decision whether to cooperate or not by retrieving two cases: the most recent case which occurred in a *similar* situation for each of the two decisions (*i.e.* each of the two possible values of d_t). A case is perceived by the Agent to have occurred in a *similar* situation 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. The only function of the perceived state of the world is to determine whether two situations look *similar* to the Agent or not.

In a particular situation (*i.e.* for a given perceived state of the world) an Agent must face one of the following three possibilities:

1) The Agent cannot recall any previous similar situations. In CBR terms, the Agent does not hold any cases whose state of the world matches the current perceived state of the world. In this case the Agent will make an unbiased random decision.

2) The Agent does not remember any previous similar situations 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 Agent's cases whose state of the world

matches the current perceived state of the world have the same value for d_t. In this situation, Agents will explore the non-applied decision if the payoff they obtained in the last previous similar situation was below their *Aspiration Threshold AT*; otherwise they will keep the same decision they previously applied in similar situations.

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. In this way, Agents adapt their behaviour according to the most recent feedback they got in a similar situation.

In the experiments reported in this paper, all the Agents share the same Aspiration Threshold *AT* and the same *Memory Length ml*. These are the two crucial parameters in this CBR decision-making algorithm, determining when an outcome is satisfactory (so the search for solutions can stop) and when two situations are similar, respectively.

5. Iterated elimination of strictly dominated outcomes

In this section we explain a solution concept that is more relevant for case-based reasoners than the Nash equilibrium: strictly undominated outcomes (SUO). SUO are outcomes in which no player can be guaranteed a higher payoff by changing their decision (*i.e.* every player is getting at least their *Maximin*). It can be proved that simulations of the three games explained in section 2 when played by Agents using the decision making algorithm outlined in section 4 with non-trivial *AT* (*i.e. AT* greater than the minimum payoff an Agent can get) end up locked in to cycles made up of SUO.

Using the concept of SUO, Izquierdo *et al.* (2004) introduced the process of iterated elimination of strictly dominated outcomes. The idea is that a player cannot rationally accept outcomes in which the player is not getting at least their *Maximin* (a rational player is not exploitable). When players who do not accept outcomes where they get a payoff lower than *Maximin* meet, they might learn by playing the game the fact that their opponent is not exploitable either. If this occurs, it will be **mutual belief** that strictly dominated outcomes will not be sustainable because at least one of the players will not accept them. That inference (and the consequent disregard of strictly dominated outcomes by every player) can make an outcome which was not previously dominated in effect be dominated. In other words, the concept of strict dominance can be applied to outcomes *iteratively* just as it is applied *iteratively* to strategies.

As an example, consider the PD (**Figure 1**a). The only two SUO in the PD are bilateral cooperation and bilateral defection, since the other outcomes yield a payoff lower than *Maximin* to the cooperator (**Figure 1**b). If, through repeated interaction, players were able to infer that the game will not have any other outcome (because one of the players will not accept it), then they could eliminate the unilateral outcomes from their analysis and apply the concept of outcome dominance for the second time to the (two) remaining possible outcomes. For this to happen, it would have to be mutual belief that the other player is not exploitable either. When only bilateral decisions are confronted, the only strictly undominated outcome is bilateral cooperation (**Figure 1**c). When confronted with bilateral cooperation as the only alternative, bilateral defection is not strictly undominated anymore, since the two players are guaranteed a higher Payoff by changing their decision. In other words, bilateral cooperation is the only outcome that survives two steps of outcome dominance in the PD.

Figure 1. Elimination of dominated outcomes in the PD.

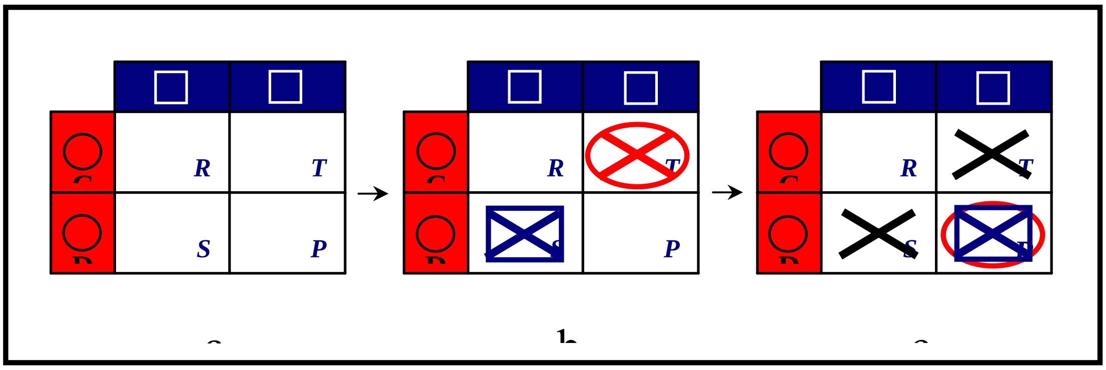


Figure b shows the remaining outcomes after having applied one step of outcome dominance. Figure c shows the remaining outcomes after having applied two steps of outcome dominance. Red circled crosses represent outcomes which are unacceptable for player Red (row), blue squared crosses represent outcomes which are unacceptable for player Blue (column), and black plain crosses represent outcomes eliminated in previous steps.

It can be shown that in any game, after applying any number of steps of strict outcome dominance, those outcomes remaining are not Pareto-dominated by any of those which have been eliminated. In particular, all the outcomes that survive two steps of outcome dominance in the PD, Chicken and Stag Hunt are **Pareto optimal**.

It is interesting to notice that when Agents decide using cases (or outcomes) as the basis of inference, the resulting outcomes seem to be 'more rational' (*i.e.* Pareto optimal) than when 'rational' (*i.e.* dominant) strategies are employed: that, as we explain below, reflects the essence of social dilemmas. Although defining rational strategies in interdependent decision-making problems is by no means trivial, it seems sensible to assume that a) rational players choose **strictly dominant strategies**, and b) rational players do not choose strictly dominated strategies. Similarly, even though defining rational outcomes cannot be done without controversy, it also seems sensible to agree that rational outcomes must be Pareto optimal. Assuming only those necessary conditions for the rationality of strategies and outcomes, we can state that, in the one-shot PD, even though there is a clear causal link between strategies and outcomes, rational strategies lead to outcomes which are not rational, whereas rational outcomes are generated by strategies which are not rational.

6. Results

The software used to conduct the experiments reported in this section was written in Objective-C using the Swarm libraries (http://wiki.swarm.org/) and is available online at http://www.macaulay.ac.uk/fearlus/casd/ under GNU General Public Licence. The program is known to work on a PC using Swarm 2.1.1 and on a Sun Sparc using Swarm 2001-12-18.

As might be expected, the 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 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, can be exactly computed.

Agents in the model make decisions at random only when they perceive a novel state of the world. Since the number of different states of the world that an Agent can perceive is finite, so is the number of random decisions the Agent can make. Therefore simulations must end up in a cycle. To study how often Agents cooperate in the three games we define the 'cooperation rate' as the number of times bilateral cooperation is observed in a cycle divided by the length of the cycle.

6.1. Prisoner's Dilemma

It is important to realise that when our Agents play the PD, Chicken or Stag Hunt, they both share the same perception of the state of the world (defined by the last *ml* moves of the two Agents) 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. Therefore, at any given time in the simulation our Agents will have visited any given state of the world the same number of times. This shared perception of the state of the world means 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.

Table 2. Decisions made by each of the two Agents playing the PD when visiting a certain state of the world for the *i*-th time. In the first column, payoffs are denoted by their initial letter. In columns 2 to 5, the first letter in each pair corresponds to the decisions of one Agent, the second letter to those of the other. C is cooperation and D is defection. The results shown in this table are independent of the value of the Memory Length.

Aspiration Thresholds (AT)	1st visit (random)	2nd visit	3rd visit	4th visit and onwards
$T < AT$	C-C	D-D	C-C	C-C
	C-D	D-C	D-D	D-D
	D-C	C-D	D-D	D-D
	D-D	C-C	C-C	C-C
$R < AT \leq T$	C-C	D-D	C-C	C-C
	C-D	D-D	D-C	D-D
	D-C	D-D	C-D	D-D
	D-D	C-C	C-C	C-C
$P < AT \leq R$	C-C	C-C	C-C	C-C
	C-D	D-D	D-C	D-D
	D-C	D-D	C-D	D-D
	D-D	C-C	C-C	C-C
$S < AT \leq P$	C-C	C-C	C-C	C-C
	C-D	D-D	D-D	D-D
	D-C	D-D	D-D	D-D
	D-D	D-D	D-D	D-D

$AT \leq S$	C-C	C-C	C-C	C-C
	C-D	C-D	C-D	C-D
	D-C	D-C	D-C	D-C
	D-D	D-D	D-D	D-D

The decision dynamics for a certain state of the world are summarised in

Table 2. Consider first the first four rows of the table ($T < AT$). These represent the case where the Aspiration Threshold (for both Agents) exceeds T. The first time any particular state of the world occurs, both Agents will choose C (Cooperate) or D (Defect) at random (column headed "1st visit"). When the same perceived state occurs a second time, the responses will be as shown in the "2nd visit" column, and so on. The table shows that by the third visit to that state, either both Agents are cooperating or both Agents are defecting, and both will then continue to make the same response. The other four sets of rows in the table show what happens when the AT is in each of four lower ranges of values.

When the simulation locks in to a cycle (and it necessarily does), the states that make up the cycle are repeatedly visited, leading to outcomes shown in the "4th visit and onwards" column in

Table 2. Looking at that column, we can identify two values for the Aspiration Threshold AT that make a particularly important difference: *Sucker* and *Punishment*.

- When $AT >$ *Sucker*, simulations lock in to cycles which are necessarily made up of bilateral decisions (the only two SUO), since if an Agent receives the S*ucker* Payoff in any situation, they will never cooperate again in that situation. In this sense our Agents are particularly unforgiving. Agents with Aspiration Thresholds greater than *Sucker* cannot be systemically exploited.

- When $AT >$ *Punishment*, there is a qualitative jump in terms of average cooperation rates. This is because if $AT >$ *Punishment*, when both Agents defect the first time they experience a certain state of the world, they will end up cooperating in that state, but they will end up defecting if $AT \leq$ *Punishment*.

Taking into account the two previous points and looking at the “4th visit and onwards” column in

Table 2, one could then think that average cooperation rates should be 25% if *AT* ≤ *Punishment* and 50% if *AT* > *Punishment* regardless of the Memory Length, but one would be wrong. Figure 2 shows the importance of Aspiration Thresholds and how they can modify the effect of the Memory Length.

Figure 2. Average cooperation rates when modelling two agents with Memory Length *ml* and Aspiration Threshold *AT*, playing the PD. The average cooperation rate shows the probability of finding both Agents cooperating once they have finished the learning period (*i.e.* when the run locks in to a cycle). The values represented for *ml* = 1 have been computed exactly. The rest of the values have been estimated by running the model 10,000 times with different random seeds. All standard errors are less than 0.5%.

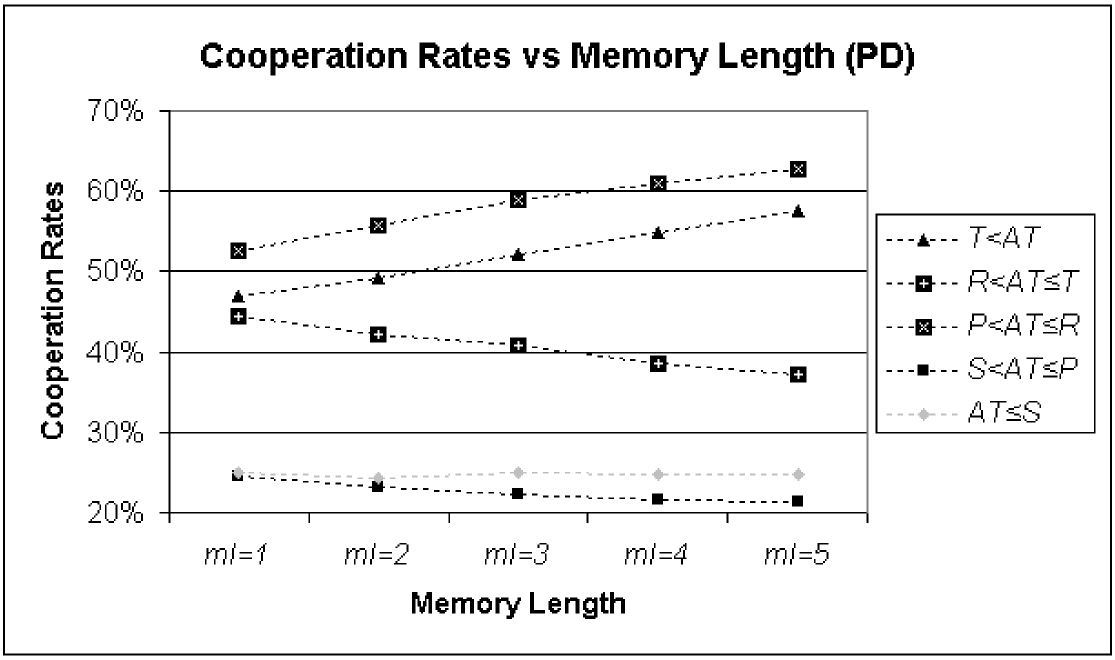


Figure 2 shows that in CBR, not only *what* is learnt, but the actual *process* of learning can have a major importance, and Aspiration Thresholds play a crucial role in that process. Consider, for example, the difference between the cases where $P < AT \leq R$ and where $R < AT \leq T$. In both cases, Agents will learn to cooperate in any given state of the world if they happen to make the same decision the first time they visit that state, and they will end up defecting in that situation otherwise. Therefore, for those two values of AT, we could expect average cooperation rates to be the same or at least similar. However, because the actual process of learning is different, differences in average cooperation rates are substantial and get larger as the Memory Length increases (see Figure 2).

6.2. Chicken

The decision dynamics for a certain state of the world in Chicken are summarised in Table **3**.

Table 3. Decisions made by each of the two Agents playing Chicken when visiting a certain state of the world for the *i*-th time. In the first column, payoffs are denoted by their initial letter. In columns 2 to 5, the first letter in each pair corresponds to the decisions of one Agent, the second letter to those of the other. C is cooperation and D is defection. The results shown in this table are independent of the value of the Memory Length.

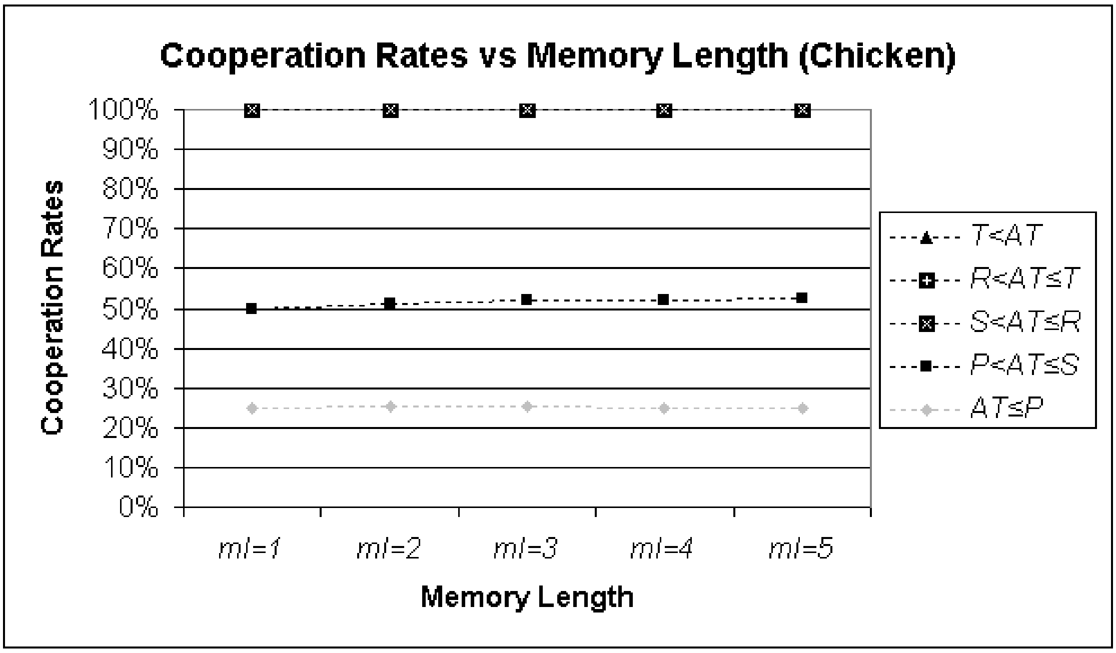
Aspiration Thresholds (AT)	1st visit (random)	2nd visit	3rd visit	4th visit and onwards
$T < AT$	C-C	D-D	C-C	C-C
	C-D	D-C	D-D	C-C
	D-C	C-D	D-D	C-C
	D-D	C-C	C-C	C-C
$R < AT \leq T$	C-C	D-D	C-C	C-C
	C-D	D-D	C-C	C-C
	D-C	D-D	C-C	C-C

	D-D	C-C	C-C	C-C
$S < AT \leq R$	C-C	C-C	C-C	C-C
	C-D	D-D	C-C	C-C
	D-C	D-D	C-C	C-C
	D-D	C-C	C-C	C-C
$P < AT \leq S$	C-C	C-C	C-C	C-C
	C-D	C-D	C-D	C-D
	D-C	D-C	D-C	D-C
	D-D	C-C	C-C	C-C
$AT \leq P$	CC	CC	CC	CC
	CD	CD	CD	CD
	DC	DC	DC	DC
	DD	DD	DD	DD

Similarly to the case where Agents played the PD, Agents playing Chicken end up in cycles made up of SUO (

Table 3). However, contrary to the PD case, where Agents could end up in any of the SUO (bilateral cooperation and bilateral defection) if $AT > S$, in Chicken if $AT > S$, Agents seem to fix on only one of the three SUO: bilateral cooperation. Figure 3 shows the results obtained running the model, and confirms Agent's fixation on bilateral cooperation. The reason for such fixation on one SUO is discussed in section 7.

Figure 3. Average cooperation rates when modelling two agents with Memory Length *ml* and Aspiration Threshold *AT*, playing Chicken. The cooperation rate is one for *AT* > *S* regardless the value of *ml*. The values represented for *ml* = 1 have been computed exactly. The rest of the values have been estimated by running the model 10,000 times with different random seeds. All standard errors are less than 0.5%.



6.3. Stag Hunt

The decision dynamics for a certain state of the world in Stag Hunt are structurally equivalent to those in the PD except for the case where $T < AT \leq R$ in Stag Hunt, which is presented in

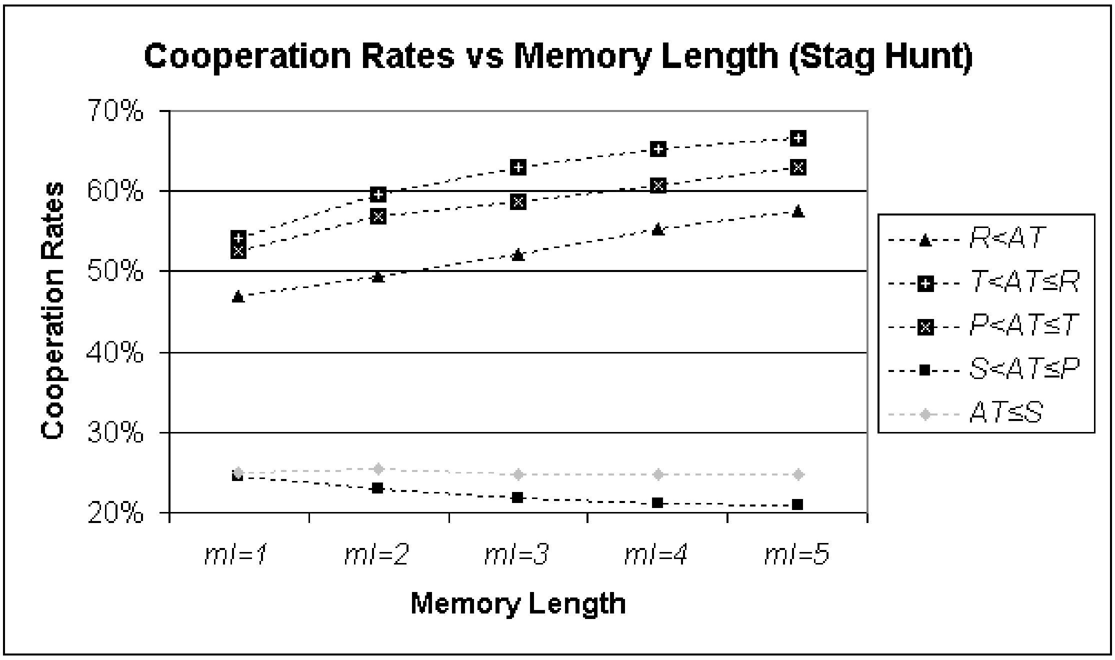
Table 4.

Figure 4 shows the results obtained running the model.

Table 4. Decisions made by each of the two Agents playing Stag hunt when visiting a certain state of the world for the *i*-th time.

Aspiration Thresholds (AT)	1st visit (random)	2nd visit	3rd visit	4th visit and onwards
$T < AT \leq R$	C-C	C-C	C-C	C-C
	C-D	D-C	D-D	D-D
	D-C	C-D	D-D	D-D
	D-D	C-C	C-C	C-C

Figure 4. Average cooperation rates when modelling two agents with Memory Length *ml* and Aspiration Threshold *AT*, playing Stag Hunt. The values represented for *ml* = 1 have been computed exactly. The rest of the values have been estimated by running the model 10,000 times with different random seeds. All standard errors are less than 0.5%.



As in the PD, it is also clear from these results that in CBR, not only *what* is learnt is important, but also *how* it is learnt, and that Aspiration Thresholds play a crucial role in that process.

7. Discussion

The experiments conducted show that cooperation can emerge from the interaction of selfish and unforgiving case-based reasoners. We have modelled a system in which Agents observe other Agents' decisions and use those observations to make further decisions. This is clearly an essential feature of any social system. Our results have shown that even in a simple system with two players the results do not only depend on *what* is eventually learnt in any given situation, but also, and very strongly, on *how* it is learnt. The results also show that Aspiration Thresholds play a major role on that learning process.

Knowing *exactly* the probability distribution of the final decisions that Agents would make in *any* situation (*i.e.* state of the world) -see Tables 2, 3, and 4- was not enough to anticipate the final outcome of the simulation. Agents' decisions lead them to situations which require new decisions, which in turn lead the Agents to new situations. Decisions and situations interweave in complex ways that are governed by the *process* by which Agents arrive at their final decisions for each situation. It is not enough to know what Agents will end up doing in any situation; the *process* of learning what to do can have major and unexpected consequences in the final outcome. In this paper we have shown how Aspiration Thresholds can alter the learning process by which final decisions are made and therefore influence the final distribution of cooperation rates in a dramatic way.

More specifically, we have found an unexpected result in Chicken. Izquierdo *et al*. (2004) proved that the case-based reasoners modelled in this paper with *AT* > *Maximin* would end up locked in to cycles consisting of SUO, and that has certainly been the case. However, whereas in the PD and Stag Hunt every SUO was visited with some probability, in Chicken Agents lock in to cycles consisting only of bilateral cooperations, even though unilateral outcomes are also SUO. What is it that makes Agents playing Chicken prefer bilateral cooperation to the other two SUO? The answer, which is explained in detail below, is that bilateral cooperation in Chicken is the outcome that occurs when Agents avoid the minimum payoff they can get.

Let us call *minimumPayoff* the minimum payoff an Agent can get in a game. As we explained before, our Agents (with *AT* > *minimumPayoff*) are particularly unforgiving in the sense that if they happen to receive *minimumPayoff*, then they will never repeat the decision that led them to that undesirable outcome in a similar situation (in a 2x2 game such behaviour is equivalent to adopting a **Maximin strategy**). In fact, in the three games with *AT* > *minimumPayoff*, if one Agent happens to receive *minimumPayoff* having observed a certain state of the world, then both Agents will end up adopting the Maximin strategy after that state of the world. In other words, if any Agent happens to receive *minimumPayoff* in a given state of the world, the outcome after that state of the world will eventually be the **Maximin equilibrium** (bilateral defection in the PD and Stag Hunt, and bilateral cooperation in Chicken).

The question now is: Under what circumstances do none of the Agents ever receive *minimumPayoff*? For any given state of the world, one of the Agents will receive *minimumPayoff* (and therefore they will end up in the Maximin equilibrium) unless they lock in to an outcome before doing so. If Agents are to accept an outcome O before one of the Agents has received *minimumPayoff*, the outcome O must satisfy one of the following two conditions:

a) Both Agents' payoffs in O must be greater than their Aspiration Threshold.

b) Both Agents can identify another outcome (excluding those in which one of the Agents received *minimumPayoff*, since they have not been visited yet by assumption) in which they took the decision they are not taking at outcome O, and they got a lower payoff.

It can be checked that the only outcome that satisfies either of these conditions when $AT > Maximin$ is bilateral cooperation in the three games.

We have shown then that when $AT > Maximin$, if any Agent happens to receive *minimumPayoff* in a given state of the world, then the outcome after that state of the world will eventually be the Maximin equilibrium; whereas if no Agent receives *minimumPayoff* in a given state of the world, then the outcome after that state of the world will eventually be bilateral cooperation. Since the Maximin equilibrium in Chicken is bilateral cooperation, the only possible outcome in Chicken when $AT > Maximin$ is bilateral cooperation. On the other hand, in the PD and Stag Hunt, the Maximin equilibrium is bilateral defection, so both SUO may occur in those games.

8. Conclusions

We have explored the outcome of social dilemmas when played by case-based reasoners. CBR is a method of inference that is believed to be commonly used by real people when they face novel or difficult problems in which they cannot easily compute a satisfactory solution (Ross, 1989). Social dilemmas are clearly ill-defined and difficult 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 their opponent's actions, and find a satisficing solution within the constraints that their opponent's actions impose. By implicitly anticipating the outcome of their actions, our 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. In determining how often cooperation occurs, not only what Agents end up doing in a given situation is important, but also the process of learning what to do can crucially influence the final outcome, and aspiration thresholds play an important role in that process.

The experiments conducted have also revealed that case-based reasoners find it easier to cooperate in the game of Chicken than in the PD or Stag Hunt. The reason is that case-based reasoners avoid outcomes where they are getting a payoff below *Maximin*, and in doing so, they often end up in the Maximin equilibrium; the Maximin equilibrium in Chicken is bilateral cooperation, whereas it is bilateral defection in the PD and Stag Hunt.

Acknowledgement

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

Appendix A

Common knowledge of rationality (CKR): CKR means that every agent assumes: (a) that all agents are instrumentally rational, and (b) that all agents are aware of other agents' rationality-related assumptions (this produces an infinite recursion of shared assumptions).

Individually-rational outcome: An outcome giving each player at least the largest payoff that they can guarantee receiving regardless the opponents' moves.

Instrumentally rational: An instrumentally rational agent acts as if they have consistent preferences and unlimited computational capacity.

Maximin: The largest possible payoff a player can guarantee themselves. This is *Punishment* in the Prisoner's Dilemma and Stag Hunt, and *Sucker* in Chicken.

Maximin equilibrium: The outcome that results when every player selects their Maximin strategy (see below).

Maximin strategy: Player A's Maximin strategy is the one that guarantees A the best outcome if the other player plays the strategy that is worst for A. The Maximin strategy is to defect in the Prisoner's Dilemma and Stag Hunt, and to cooperate in Chicken.

Mixed-Strategy: A strategy consisting of selecting each of the two possible actions (cooperate or defect) with a certain probability different from zero or one.

Mutual belief: A proposition *A* is *mutual belief* among a set of agents if each agent believes that *A*. Mutual belief by itself implies nothing about what, if any, believes anyone attributes to anyone else.

Nash equilibrium: A set of strategies such that no player, knowing the strategy of the other(s), could improve their expected payoff by changing their own.

Outcome: A particular combination of strategies, one for each player, and their associated payoffs. In the one-shot games studied in this paper, an outcome corresponds to a cell in the payoff matrix.

Pareto deficient: An outcome is Pareto deficient if there is an alternative in which at least one player is better off and no player is worse off.

Pareto optimal: An outcome is Pareto optimal if there is no other outcome in which at least one player is better off and no player is worse off.

Strictly dominant strategies: For an agent A, strategy S^*_A is strictly dominant if for each feasible combination of the other players' strategies, A's payoff from playing S^*_A is strictly more than A's payoff from playing any other strategy.

Strictly dominated strategy: For an agent A, strategy S_A is strictly dominated by strategy S^*_A if for each feasible combination of the other players' strategies, A's payoff from playing S_A is strictly less than A's payoff from playing S^*_A (Gibbons, 1992, p. 5).

9. References

Aamodt, A. and Plaza, E. (1994). "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches", *AI Communications*. IOS Press, 7 (1), 39-59.

Colman, A. M. (1995), *Game Theory and Its Applications in the Social and Biological Sciences*. 2nd edition. Oxford (UK): Butterworth-Heinemann.

Colman, A. M. (2003). "Cooperation, psychological game theory, and limitations of rationality in social interaction", *The Behavioral and Brain Sciences*, 26, 139-153.

Doran, J. (1997). "Analogical Problem Solving", in *Artificial Intelligence Techniques: A Comprehensive Catalogue*, (Ed.) Bundy A. Springer-Verlag, 4th revised edition, p. 4.

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

Gilboa, I. and Schmeidler, D. (1995). "Case-Based Decision Theory", *The Quarterly Journal of Economics*, 110, 605-639.

Gilboa, I. and Schmeidler, D. (2001), *A Theory of Case-Based Decisions*. Cambridge (UK): Cambridge University Press.

Hargreaves Heap, S. P. and Varoufakis, Y. (1995), *Game Theory: A Critical Introduction*. London: Routledge.

Izquierdo, L. R., Gotts, N. M., and Polhill, G. P. (2004). "Case-Based Reasoning, Social Dilemmas, and a New Equilibrium Concept", *Journal of Artificial Societies and Social Simulation*, forthcoming.

Ross, B. (1989). “Some Psychological Results on Case-based Reasoning”, in *Proceedings of the Case-Based Reasoning Workshop*, (Ed.) Hammond, K, 144-147. DARPA, Morgan Kaufmann, Inc.

Schank, R. and Abelson, R. P. (1977), *Scripts, Plans, Goals and Understanding*. Hillside (New Jersey): Lawrence Erlbaum Associates.

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