

The Implications of Case-Based Reasoning in Strategic Contexts

Luis R. Izquierdo & Nicholas M. Gotts

The Macaulay Institute, Craigiebuckler, AB15 8QH, Aberdeen, UK

Abstract

This paper characterises the transient dynamics and the long-term behaviour of a game theoretical model where players' decisions at any particular time are guided by a single similar situation they experienced in the past – a simple form of case-based reasoning. The transient dynamics of the model are very dependent on the process by which players learn how to play the game in any given situation. The long-run behaviour of the model varies significantly depending on whether players can occasionally explore different actions or not. When the probability of experimentation is small but non-zero, only a subset of the outcomes that are possible in the absence of experimentation persists in the long-run. In this paper we present some features that characterise such a subset of stochastically stable outcomes.

1 Introduction

This paper deals with the formal study of social interactions which can be meaningfully modelled as decision problems of strategy and, as such, using game theory as a framework. Game theory is a branch of mathematics devoted to the formal analysis of decision making in social interactions where the outcome depends on the decisions made by potentially several individuals. It is a useful framework to accurately and formally describe interdependent decision-making processes, and it also provides a collection of solution concepts that narrow the set of expected outcomes in such processes. The most widespread solution concept in game theory is the Nash equilibrium, which is a set of strategies, one for each player, such that no player, knowing the strategy of the other(s), could improve her expected payoff by unilaterally changing her own strategy.

Though extremely useful, game theory is at present somewhat limited in the sense that it is dominated by assumptions of full rationality, it generally ignores the dynamics of social processes, and it often requires disturbing and unrealistic hypotheses about players' assumptions on other players' cognitive capabilities and beliefs in order to derive specific predictions. Furthermore it is often the case that even with heroic assumptions about the computational power and beliefs that every player attributes to every other player, game theory cannot reduce the set of

expected outcomes significantly (*e.g.* consider the Folk Theorem in the repeated Prisoner's Dilemma).

While acknowledging that the work conducted in game theory up until now has been extremely useful, a growing inter-disciplinary community of scientists is starting to believe that the time has come to develop game theory past the boundaries of full rationality, common-knowledge of rationality¹, consistently aligned beliefs (Hargreaves Heap and Varoufakis 1995), static equilibria, and long-term convergence. These concerns have led many scientists to develop models of social interactions within the framework of game theory but (a) assuming players behave in ways that are cognitively more plausible than fully rational behaviour and common knowledge of rationality (*e.g.* assuming players learn from experience), and (b) paying special attention to the dynamics of such models and not only to their long-term properties. These investigations are being undertaken experimentally and formally (both analytically and using computer simulation), and special emphasis is being paid to the study of backward-looking learning algorithms, which seem to be more plausible than the forward-looking methods of reasoning employed in orthodox game theory. The latter appear to be very demanding for human agents (let alone other animals) and remain undefined in the absence of strong assumptions about other players' behaviour and beliefs.

Some of the decision-making algorithms that have attracted the attention of researchers in game theory are: reinforcement learning (with experimental studies conducted by *e.g.* Erev et al. (1999), theoretical work done by *e.g.* Bendor et al. (2001), and studies of the dynamics carried out by *e.g.* Macy and Flache (2002)), belief learning (with theoretical work on fictitious play developed by *e.g.* Fudenberg and Levine (1998)), the EWA (Experience Weighted Attraction) model (Camerer 2003), which is a hybrid of the reinforcement and belief learning, and finally, case-based reasoning (Izquierdo et al. 2004). This paper advances the work conducted by Izquierdo et al. (2004) on the implications of case-based reasoning in strategic contexts.

In particular, we study the transient and the long-run dynamics of a game theoretical model where players' decisions at any particular time are guided by a single similar situation they experienced in the past – a simple form of case-based reasoning. It is assumed in this paper that players suffer from trembling hands, *i.e.* they occasionally explore different actions with small probability. These trembles make the model slightly more realistic and reduce the set of expected outcomes in the long-term. Some outcomes that can be observed infinitely often in the model without trembles are not stable in the model with trembles no matter how unlikely trembles are as long as they are possible. This is so because the properties of the process with trembles when the probability of trembles tends to zero differ from those of the process where the probability of trembles *is* exactly zero. In this paper we present some features that characterise the set of outcomes which are stable in the presence of small trembles.

¹ Common knowledge of rationality means that every player assumes: (a) that all players are instrumentally rational, and (b) that all players are aware of other players' rationality-related assumptions (this produces an infinite recursion of shared assumptions).

2 Case-based reasoning

Case-Based Reasoning (CBR) arose out of cognitive science research in the late 1970s (Schank and Abelson 1977; Schank 1982); since then, several psychological studies have provided support for its importance as problem-solving process in human reasoning, especially for novel or difficult tasks (see Ross (1989) for a summary). Case-based reasoning is a form of reasoning by analogy within a particular domain. It consists of “solving a problem by remembering a previous similar situation and by reusing information and knowledge of that situation” (Aamodt and Plaza 1994). Case-based reasoners do not employ abstract rules as the basis to make their decisions, but instead they use similar experiences they have had in the past. Such experiences are stored in the form of cases. A case is “a contextualised piece of knowledge representing an experience that teaches a lesson fundamental to achieving the goals of the reasoner” (Kolodner 1993, p. 13). Thus, when a case-based reasoner has to solve a problem, she is reminded of a similar situation that she encountered in the past, of what she did then, and of the outcome that resulted in the recalled situation. She then uses that ‘similar past case’ as a basis to solve the problem in the present. Case-based reasoning generally consists of four main tasks (Aamodt and Plaza 1994):

1. *Retrieve* the most similar case or cases. Generally a case in CBR is rich in information and quite complex. Aamodt and Plaza (1994) say: “a feature vector holding some values and a corresponding class is not what we would call a typical case description” (because it is too trivial). Thus, performing similarity judgements is an integral part of CBR.
2. *Reuse* the information and knowledge in the retrieved case to solve the current problem. The retrieved knowledge cannot always be directly applied, so some adaptation is sometimes required.
3. *Revise* the proposed solution. This involves the evaluation of the proposed solution.
4. *Retain* the relevant information for the future – i.e. learn.

Case-based reasoning is often used as a problem-solving technique in domains where the distinction between success and failure is either fairly trivial or is made externally. However, in decision-making contexts in general, the distinction between what is satisfactory and what is not can be far from trivial, and thus, the question of whether a particular decision used in the past should be repeated, or a new decision should be explored is crucial. This dilemma naturally gives rise to Simon’s notions of satisficing, as noted by Gilboa and Schmeidler (2001).

Gilboa and Schmeidler (2001) have developed Case-Based Decision Theory (CBDT), a formal theory of decision based on past experiences which was initially inspired by case-based reasoning. Having said that, as noted by the authors, CBDT has not much in common with CBR beyond Hume’s basic argument that “from causes which appear similar we expect similar effects”. The main difference between CBR and CBDT is that while a defining feature of CBR is that “thought and action in a given situation are guided by a single distinctive prior case” (Loui

1999), in CBDT decision-makers rank available acts according to the similarity-weighted sum of utilities that resulted in *all* available cases. Anyhow, like Gilboa and Schmeidler (2001), we do not see reasoning by cases as a substitute for expected utility theory, but as a complement: reasoning by cases seems more plausible than expected utility theory 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).

To our knowledge, the implications of CBR in strategic contexts were explored for the first time by Izquierdo et al. (2004). Izquierdo et al. (2004) developed an agent-based model, CASD², where individuals use a very simple form of CBR. The main focus of the model was to investigate the implications of reasoning by a *single* distinctive past experience in strategic environments; for that reason, issues like knowledge representation and similarity assessments, which are central in CBR, are largely ignored in CASD.

Izquierdo et al. (2004) investigated the ability of case-based reasoners to cooperate in social dilemmas. Social dilemmas are especially challenging for orthodox game theory because the predictions of the theory in such contexts are often counterintuitive and have been rejected almost invariably by empirical evidence (see, for instance, work reviewed by Colman (1995) in chapters 7 and 9). They also offer a promising arena to distinguish the differences between reasoning by cases (or outcomes³) and reasoning by rules (or strategies). The following explains why. Although defining rational strategies in interdependent decision-making problems is by no means trivial, it seems sensible to assume that a) rational players choose dominant strategies, and b) rational players do not choose 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 Prisoner's Dilemma (PD) and other social dilemmas, 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.

Using their model as a "tool to think with", Izquierdo et al. (2004) developed the concept of iterative elimination of dominated outcomes, which describes a logical process through which case-based reasoners can arrive at sensible (*i.e.* Pareto optimal) outcomes in games. Dominated outcomes are outcomes which are not individually rational – *i.e.* there is at least one player who is obtaining a payoff below her *Maximin*⁴. The idea behind the process of iterative elimination of dominated outcomes is that players cannot rationally accept outcomes where they are not obtaining at least their *Maximin* (rational players are not exploitable). When

² CASD is available online under GNU General Public Licence, together with a user guide, at <http://www.macauley.ac.uk/fearlus/casd/>.

³ An outcome is a particular combination of decisions, each of them made by one player.

⁴ The largest possible payoff a player can guarantee themselves irrespective of the other players' actions.

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 dominated outcomes cannot be sustained because at least one of the players will not accept them. That inference (and the consequent disregard of dominated outcomes by every player) can make an outcome which was not previously dominated in effect be dominated. In other words, the concept of dominance can be applied to outcomes *iteratively* just as it is applied *iteratively* to strategies.

While useful as a “tool to think with”, the specific model Izquierdo et al. (2004) used was unrealistic in the sense that their simulations would necessarily end up with the agents locked in to a persistent cycle. In this paper we advance their work by developing their model further. In particular, in our model, players suffer from trembling hands (Selten 1975) – i.e. they occasionally experiment (or make mistakes) with small probability. This new functionality makes the model more realistic and allows us to make more specific predictions. In particular, we will characterise the set of outcomes where the system spends a significant proportion of time in the long-term when players experiment with very low probability. Such a set of outcomes is a subset of the set of outcomes that can be observed in the model without experimentation. As an example, we will see that in the prisoner’s dilemma, mutual cooperation belongs to the latter set but not to the former.

3 The model

The model we study here is a generalisation of CASD. In our model individuals play repeatedly a game, once per time-step. Every time they play, each player i retains a case (an experience), which comprises:

- The time-step t when the case occurred.
- The *perceived* state of the world at the beginning of time-step t , which is determined by all the decisions undertaken by each player in the game (including the case-holder) in the preceding m_i (for *memory*) time-steps. Thus, if there are n players in the game, each of whom can select one among a set of a possible actions, an agent with memory m will be able to identify $a^{n \cdot m}$ different states of the world. Thus, player i ’s perceived state of the world at the beginning of time-step t consists of the m_i preceding decisions made by every player.
- The decision made by the case-holder in that situation, in time-step t , having observed the state of the world in that same time-step.
- The payoff that the case-holder obtained after having decided in time-step t .

The number of cases that players can keep in memory is unlimited. At the time of making a decision, players decide what action to select by retrieving the most recent case which occurred in a *similar* situation for each one of the actions available to them. This set of cases, which is potentially empty, is denoted C_i . A case is perceived by the player 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 case-holder. Thus, the only function of the perceived state of the world is to determine whether two situations look *similar* to the player or not. In a certain situation (*i.e.* for a given perceived state of the world) any particular player i will face one of the following three possibilities:

- If the set C_i is empty, player i selects one action at random.
- If at least one action of those available to player i is represented in C_i but not all of them are, then player i selects randomly among those actions with the highest payoff obtained in the set C_i . If the selected action provided a payoff at least equal to player i 's *Aspiration Threshold* (AT), then the action is considered satisfactory and will be undertaken. Otherwise, one of the actions that are not represented in C_i will be selected at random.
- If every action available to player i is represented in C_i , then player i selects randomly among those actions with the highest payoff obtained in the set C_i .

Thus, players in this model satisfice in the sense that, when making decisions in situations that look familiar to them, they explore new actions if and only if the best decision taken in similar past situations did not meet their aspirations (and there are new actions to try).

As mentioned before, we also assume that players suffer from trembling hands: there is some small probability $\varepsilon \cdot \lambda_i \neq 0$ that player i selects her action randomly instead of following the algorithm above. The ratio λ_i / λ_j determines player i 's relative tendency to experiment compared with player j 's. The factor ε is a general measure of the frequency of experimentation in the whole population of players. The event that i experiments is assumed to be independent of the event that j experiments for every $i \neq j$. Different players may experiment in different ways, but it is assumed that player i 's probability of selecting any action a available to her when experimenting ($q_i(a)$) is non-zero, potentially different for different actions, and independent of time for all i ; these conditions can be relaxed to some extent. This completes the specifications of the model where players suffer from trembling hands, which will be referred to as the perturbed model.

This paper will present some mathematical results valid when the overall probability of experimentation ε tends to zero; all such results are independent of λ_i and of the particular way each of the players experiments. When presenting simulation results, it will be assumed that $\lambda_i = 1$ for all i , and that players select one of their actions randomly and without any bias when experimenting.

4 Results and Discussion

In the unperturbed model ($\varepsilon = 0$), players lock in to cycles, so they cannot experience all the different situations they would regard as different: they indefinitely go through cycles made up of a (usually small) subset of all the situations they can distinguish (Izquierdo et al. 2004). However, in the perturbed model every situation that players can distinctively recognise can occur with non-zero probability, so *eventually* every such situation happens, and it happens infinitely often.

More formally, note that both the perturbed and the unperturbed model can be formulated as finite-state discrete-time Markov chains, but there is a crucial difference between them: the unperturbed model will end up in one of many possible cycles (the period of some of these cycles is potentially equal to one), whereas the perturbed process has one unique limiting distribution. Thus, when players suffer from trembling hands, the indefinite cycles where players were locked in the unperturbed model are broken, and outcomes that occurred infinitely often in the unperturbed process (like mutual cooperation in the prisoner's dilemma (Izquierdo et al. 2004)) turn out not to be robust to small trembles. In the following two sections we study the transient and the long-run behaviour of the perturbed process.

4.1 Transient Dynamics

As one would expect, the short-term dynamics of the perturbed process – i.e. when only a few trembles have taken place – are initially similar to the dynamics of the unperturbed process. How many ‘a few trembles’ are depends on the players’ memory and aspiration thresholds; how quickly those ‘few trembles’ occur depends on the probability of trembles happening. Fig. 1 shows the proportion of outcomes where both players are cooperating (cooperation rate) in the Prisoner's Dilemma (PD) for different values of both players’ memory m and aspiration threshold AT , and for different values of the overall probability of trembles ε . The cooperation rates shown in Fig. 1 are calculated over time-steps 1001 to 1100.

A word of caution about Fig. 1 is that, because it shows the data collected at a predetermined range of time-steps (1001–1100), it represents the short-term behaviour of those series for which 1000 time-steps are not enough to approach their long-term behaviour (e.g. $m_i = 5$) but, on the other hand, it represents the long-run behaviour for some other series (e.g. those series for which 1000 time-steps are enough to reach it, like series with $m_i = 0$, and $\varepsilon \neq 0.001$). If enough number of trembles have taken place in every situation distinctively perceived by any player, then the dynamics of the perturbed model will resemble its long-run behaviour, which is independent of the players’ memory and of their aspiration thresholds. Aspiration thresholds are irrelevant in the long-term because sooner or later all players will conduct every possible action in every possible situation, so their selection algorithm will not take the value of their aspiration threshold into account. That the actual values m_i are also irrelevant in the long-term can be explained as follows: divide the set of possible situations in the game into disjoint classes such that all the situations in a certain class look similar to every agent. The decision processes occurring within each of those classes define a finite-state irreducible aperiodic discrete-time Markov chain. This Markov chain is exactly the same in every class because the decision processes are exactly the same in any one situation. Because eventually every class will be infinitely many times revisited, every class will reach its asymptotic behaviour, which is the same for every class. The values of m_i affect the number of classes, but not the decision processes within them so, in the long-run, the values of m_i are irrelevant.

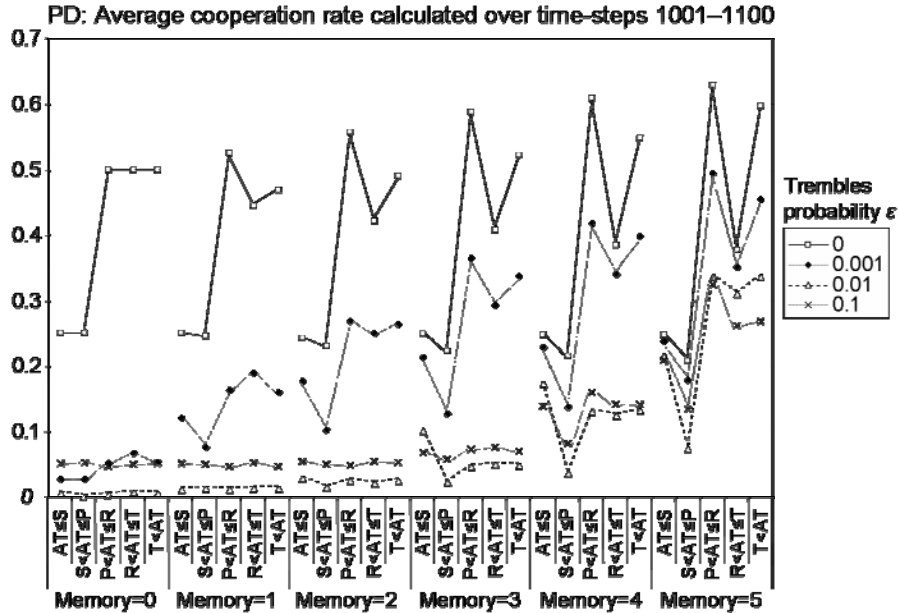


Fig. 1. Average proportion of outcomes where both players are cooperating in the Prisoner's Dilemma (PD), calculated over 100 time-steps starting at time-step 1000, and using 500 simulation runs for each data point. The payoffs in the game are represented by its initial letter: S for Suckers, P for Punishment, R for Reward, and T for Temptation

Thus, for example, the long-run cooperation rate in the PD (calculated analytically) is $4.985 \cdot 10^{-2}$ for $\varepsilon = 0.1$, $4.978 \cdot 10^{-3}$ for $\varepsilon = 0.01$, and $4.998 \cdot 10^{-4}$ for $\varepsilon = 0.001$. As we can see in Fig. 1, the series with low memory ($m_i = 0$ or $m_i = 1$) and high probability of trembles ($\varepsilon = 0.1$ or $\varepsilon = 0.01$) quickly converge to their limiting values; for those parameterisations 1000 time-steps are sufficient to closely approach the long-run behaviour of the process. If we represented the data in Fig. 1 after a sufficiently high number of time-steps, the value of every data point with $\varepsilon \neq 0$ would only depend on the probability of trembles ε (and on λ_i and $q_i(\cdot)$ generally), and it would approach the analytically calculated values presented above (calculated for $\lambda_i = 1$, and $q_i(\cdot)$ unbiased). Something which is clear in Fig. 1 is that whereas mutual cooperation usually forms part of the cycles in the unperturbed process, it cannot be sustained in the long-term when small trembles occur.

Hence the short-term behaviour of the perturbed model is a transition from a distribution similar to that corresponding to the unperturbed model to a very different distribution which is only dependent on the probabilities with which trembles occur. Thus, at the beginning of a simulation, as shown by Izquierdo et al. (2004) in the unperturbed model, the behaviour of the system does not only depend on *what* is learnt by each player in any given situation, but also, and very strongly, on *how* it is learnt, and aspiration thresholds play a major role on that learning process. Players' decisions lead them to situations which require new de-

decisions, which in turn lead players to new situations. Decisions and situations interweave in complex and highly path-dependent ways that are governed by the *process* by which players arrive at their final decisions for each situation. As time goes by, more and more trembles occur in each situation distinctively perceived by the players, and consequently the behaviour of the system starts to change. Eventually, players' aspiration thresholds and the values of m_i become irrelevant, so the initial mighty impacts of these parameters are lost forever in the mist of the past.

4.2 Long-run behaviour

Having seen that the asymptotic behaviour of the model is only dependent on the structure of trembles, a natural question is: What outcomes can be observed with probability bounded away from zero in the long-run as the probability of trembles ε tends to zero? Following Young (1993), such outcomes will be called *stochastically stable*. It turns out that whether an outcome is stochastically stable or not is independent of λ_i and of $q_i(\cdot)$ (Young 1993).

Young (1993) provides a general method to identify stochastically stable *states* in a wide range of models by solving a series of shortest path problems in a graph. In our model there are more states than outcomes, but identifying stochastically stable outcomes when the set of stochastically stable states is known is straightforward. Young's method uncovers an important feature of stochastic stability: stochastic stability selects states which are easiest to flow into from *all* possible states of the system. This contrasts with most notions of equilibrium based on full rationality. As Young (1993) notes, risk dominance "selects the equilibrium that is easiest to flow from every other equilibrium considered in isolation". Similarly, Nash stability is determined only by unilateral deviations from the equilibrium.

In this section we present some features to identify stochastically stable outcomes when reasoning is based on singletons of distinct prior outcomes. We start with a necessary condition for outcomes to be stochastically stable.

Proposition 1. Every stochastically stable outcome is individually rational.

Proof. Bearing in mind that players' memory and aspiration thresholds are not relevant in the long-term, let us focus on a perturbed model ($\varepsilon \neq 0$, $\lambda_i \neq 0$, and $q_i(\cdot) > 0$) where every agent has memory $m_i = 0$ and any arbitrary aspiration threshold. Since sooner or later all players will conduct every possible action, in studying the long-term behaviour of the system there is no point in considering states of the system where some of the actions have not been selected yet. Therefore let us define a state of the system by the payoff obtained by each player the last time they conducted each of the actions available to them. (This highlights why aspiration thresholds are irrelevant in the long-term). The set of possible states may be smaller than every possible combination of payoffs, since in some games some combinations cannot occur in the course of a simulation. The model thus defined is a finite-state irreducible aperiodic discrete-time Markov chain, which is denoted P^ε . Let P^0 be the Markov process P^ε when $\varepsilon = 0$ (which is generally reducible).

The proof rests on two arguments. The first argument, which is an immediate application of theorem 4 in Young (1993), is that every stochastically stable state is a recurrent

state of P^0 . The second argument is that the outcome succeeding any recurrent state of P^0 is necessarily individually rational. The following proves an alternative (but equivalent) formulation of the second argument: if an outcome which is not individually rational can succeed state s in P^0 , then s is a transient state of P^0 . We will prove this second argument by showing that if an outcome which is not individually rational succeeds s , then s will never be revisited. Let t be the state succeeding s when an outcome which is not individually rational has occurred. Let A be one of the players who has received a payoff below her *Maximin* after s . Let a be the action that A chose, and $p_t(A, a)$ the payoff she obtained, which is part of state t . Since A selected action a in state s , the payoff $p_s(A, a)$ that A attaches to action a in state s is at least equal to her *Maximin*. Thus, $p_s(A, a) > p_t(A, a)$. Since $p_t(A, a)$ is below A 's *Maximin*, A will never select action a ever again, so the payoff $p(A, a)$ that A will attach to action a in any subsequent state will remain unmodified. Therefore, state s , in which A attaches $p_s(A, a)$ to action a will never be revisited again. This completes the proof of the second argument, and hence the proof of the proposition.

4	2	3	4	3	4	2
4	3	2	1	3	1	2
2	1	4	1	2	1	3
A	B	C	D	E	F	

Fig. 2. Stochastically stable outcomes (highlighted in white) in various 2-player 2-strategy games. Payoffs are numeric for the sake of clarity, but only their relative order for each player is relevant

Proposition 1 is a useful condition to identify outcomes which cannot be stochastically stable but, except in very simple games (e.g. see Fig. 2A), it is not sufficient to characterise the set of stochastically stable outcomes. To try to identify features that make outcomes stochastically stable we developed a computer program⁵ that calculates the exact long-run probability that any 2-player game spends in each possible outcome when the probability of trembles tends to zero. Using this program, we came to the following conclusions:

- Stochastically stable outcomes are not necessarily Nash equilibria (e.g. see the game of Chicken in Fig. 2B).
- In fact, some players in some stochastically stable outcomes may be choosing strictly dominated strategies (e.g. see the game represented in Fig. 2C).
- Nash equilibria are not necessarily stochastically stable (e.g. see the game of Stag Hunt in Fig. 2D).
- Stochastically stable outcomes can be Pareto dominated by outcomes which are not stochastically stable (e.g. see the Prisoner's Dilemma game in Fig. 2E). However, it can be proved that stochastically stable outcomes cannot be Pareto dominated by outcomes which are one tremble away and which are not stochastically stable. Thus, in the game represented in Fig. 2C, for example, if we

⁵ This computer program is available online at <http://www.macauley.ac.uk/fearlus/casd/>.

knew that outcome (3,3) is stochastically stable, then we could infer that (4,4) would have to be stochastically stable too.

- Stochastically stable outcomes can Pareto dominate outcomes which are not stochastically stable (e.g. see game represented in Fig. 2A).

Intuitively, note that trembles can destabilise outcomes in two different ways: by giving the deviator a higher (or equal) payoff, or by giving any of the non-deviators a lower payoff⁶. The first possibility is related to the concept of Nash equilibrium, whilst the second is related to the concept of “protection” (Bendor et al. 2001). An outcome is protected if unilateral deviations by any player do not hurt any of the other players. Bendor et al. (2001) show that under a very wide range of conditions, reinforcement learning converges to individually rational outcomes which are either Pareto optimal or a protected Nash equilibrium. The same is not true for the model we study in this paper (see the game represented in Fig. 2F), but protected strict Nash equilibria are very relevant here too: if there is a protected strict Nash equilibrium in a game, then there is at least one state which is robust to any one single tremble, and the outcome that follows such state in the absence of trembles is the protected strict Nash equilibrium. In fact, it can be shown that the only stochastically stable outcome in any 2-player 2-strategy game with a (necessarily unique) protected strict Nash equilibrium is such equilibrium. The extension of this result to more general games is left for future work.

5 Conclusions

This paper has explored the implications in strategic contexts of reasoning by single and distinctive past experiences as opposed to reasoning by abstract rules (strategies). While the short-term dynamics of models where players base their decisions on past experiences are very dependent on the specifics of such models, a very wide range of models behave similarly in the long-term. In particular, a large collection of models where players experiment from time to time share the same set of stochastically stable outcomes (outcomes that persist in the long-run when trembles are very rare).

Stochastically stable outcomes are necessarily individually rational, but a clear relationship between them and Nash equilibria, or Pareto optimality, has not been found. Nash equilibria may, or may not, be stochastically stable, and stochastically stable outcomes may, or may not, be Nash equilibria. The same applies for Pareto optimal outcomes. A concept that is indeed closely related to stochastic stability is the concept of protected strict Nash equilibrium. In particular, in 2-player 2-strategy games with a protected strict Nash equilibrium (which is necessarily unique), the only stochastically stable is such an equilibrium. Future work will be devoted to investigate whether this relation holds in more general games.

⁶ Non-deviators could get a lower payoff after a tremble and still keep choosing the same action if the payoff obtained when the tremble occurs is higher than any of the payoffs that the non-deviator obtained when she last selected each of the other possible actions.

Acknowledgements

We would like to gratefully acknowledge financial support from the Scottish Executive Environment and Rural Affairs Department. We are also indebted to Segis Izquierdo and Gary Polhill for their valuable comments on this paper.

References

- Aamodt A, Plaza E (1994) Case-based reasoning: foundational issues, methodological variations, and system approaches. *AI Communications IOS Press* 7(1):39–59
- Bendor J, Mookherjee D, Ray D (2001) Aspiration-based reinforcement learning in repeated interaction games: an overview. *International Game Theory Review* 3(2-3):159–174
- Camerer CF (2003) *Behavioral game theory: experiments in strategic interaction*. Princeton University Press, New York
- Colman, AM (1995) *Game theory and its applications in the social and biological sciences*. 2nd edition. Butterworth-Heinemann, Oxford, UK
- Erev I, Bereby-Meyer Y, Roth AE (1999) The effect of adding a constant to all payoffs: experimental investigation, and implications for reinforcement learning models. *Journal of Economic Behavior & Organization* 39(1):111–128
- Fudenberg D, Levine D (1998) *The theory of learning in games*. MIT Press, Cambridge, MA
- Gilboa I, Schmeidler D (2001) *A Theory of Case-Based Decisions*. Cambridge University Press, Cambridge, UK
- Hargreaves Heap SP, Varoufakis Y (1995) *Game theory: a critical introduction*. Routledge, London
- Izquierdo LR, Gotts NM, Polhill JG (2004) Case-based reasoning, social dilemmas, and a new equilibrium concept. *Journal of Artificial Societies and Social Simulation* 7(3) <<http://jasss.soc.surrey.ac.uk/7/3/1.html>>
- Kolodner JL (1993) *Case-Based Reasoning*. Morgan Kaufman Publishers, San Mateo, USA
- Loui R (1999) Case-Based Reasoning and Analogy. In: Wilson RA, Keil FC (eds) *The MIT Encyclopedia of the Cognitive Sciences*, The MIT Press, Cambridge, USA, pp 99–101
- Macy MW, Flache A (2002) Learning dynamics in social dilemmas. *Proceedings of the National Academy of Sciences USA*, 99(3):7229–7236
- Ross B (1989) Some Psychological Results on Case-based Reasoning. In Hammond K (ed) *Proceedings of the DARPA Case-Based Reasoning Workshop*, pp 144–147, Morgan Kaufmann Publishers, San Mateo, USA
- Schank R (1982) *Dynamic Memory: a Theory of Reminding and Learning in Computers and People*. Cambridge University Press, Cambridge, UK
- Schank R, Abelson RP (1977) *Scripts, Plans, Goals and Understanding*. Lawrence Erlbaum Associates, Hillsdale, New Jersey
- Selten R (1975) Re-examination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory*, 4:25–55
- Young HP (1993) The evolution of conventions. *Econometrica*, 61(1): 57–84