Advancing Learning and Evolutionary Game Theory with an Application to Social Dilemmas

Izquierdo, Luis R.

Manchester Metropolitan University, The Macaulay Institute, University of Burgos

24 April 2008

ADVANCING LEARNING AND EVOLUTIONARY GAME THEORY WITH AN APPLICATION TO SOCIAL DILEMMAS

Luis R. Izquierdo

A thesis submitted in partial fulfilment of the requirements of the
Manchester Metropolitan University for the degree of Doctor of Philosophy

Supervisors: Nick Gotts and Bruce Edmonds

Centre for Policy Modelling,
Manchester Metropolitan University Business School,
Manchester Metropolitan University
in collaboration with The Macaulay Institute

September 2007

Oral Examination: April 2008

To Segis,

for finding me whenever I've been lost,
for lifting me up whenever I've fallen down,
for using up all his bright light
to get me through my darkest nights,
and fill with colour and joy
every minute of my life.

Acknowledgements

In a perfect world, every student would have Nick Gotts as a supervisor. But Nick has been much more than a supervisor to me. He encouraged me to start this PhD when no one –not even me– believed I could do it. He is the wise wizard who opened my eyes and my mind in every sense I can think of. I will always say with immense pride and joy that I received my best education from my great and distinguished friend Nick Gotts.

To Bruce Edmonds, for showing me the beauty of philosophy, and teaching me to find great pleasure in reaching and questioning the essence of arguments.

To Gary Polhill, for walking me through the wonderland of computer science with the greatest wit and skill. For teaching me rigour, ethics and professionalism with infinite patience.

To Segis Izquierdo, for teaching me how to think, showing me how to appreciate astonishing beauty in reasoning, and helping me more than words can describe. For sharing with me his immense talent, his outstanding intelligence, his great generosity, and his wicked humour. For making me happy.

To Ryan McAllister, for his great and wonderful support, for believing in me when I most needed it, and for making me feel at home in Australia.

To Javier Perez Barbería, for being the light, the energy, the smile, and the sun of Aberdeen. For sharing with me his outstanding ability to bring out the passion in everyone he meets.

To José Manuel Galán, for being the greatest friend and the best workmate I could ever dream of.

To my great Irish-Australian friend Fiachra Kearney, for giving me so much of his amazing (latin!) passion to live and travel around the world.

To Paula López Urruchi, for giving me a million smiles and her unconditional support until she could not do it anymore.

To Steve Albon, Dick Birnie, Judith Bourassa, Gemma Camacho, Russell Dodd, Nigel Gilbert, Maggie Gill, David Hales, Cesáreo Hernández, Robin Matthews, John Milne, Graham Mills, Melanie Mitchell, Scott Moss, Ricardo del Olmo, Edoardo Pignotti, Elke Pioch, Stewart Rhind, Dale Rothman, Nacho Santos, Jörgen W. Weibull, Jeri Young, and many great friends and excellent colleagues at the Macaulay Institute, the University of Burgos, and Manchester Metropolitan University.

I could not have done anything in my life without my family.
They are everything to me.

To Isa, Marta, Segis, and Jose, because being with them, no matter where, is what makes me feel at home. For being my source of comfort, peace, and happiness.

To my most inspiring and brilliant teacher in life, for being also my greatest friend, an infinite source of admiration, and the best Dad I've ever heard of. For teaching me to find great pleasure in learning and in discovering truth, particularly when this implies being proven wrong. Above all, I will always be most grateful to my Dad for showing me that there is a whole world out there, waiting to be explored, and for giving me all the keys to open it up.

I have no words to describe what my mother has meant to me throughout all my life. She's been the leading figure in creating the part of me that I feel most proud of. She's the engine and the artist of my life. All my defects are made of the clay that I did not give her enough time to mould.

ABSTRACT

This thesis advances game theory by formally analysing the implications of replacing some of its most stringent assumptions with alternatives that –at least in certain contexts– have received greater empirical support. Specifically, this thesis makes two distinct contributions in the field of learning game theory and one in the field of evolutionary game theory. The method employed has been a symbiotic combination of computer simulation and mathematical analysis. Computer simulation has been used extensively to enhance our understanding of various formal systems beyond the current limits of mathematical tractability, and also to illustrate, complement and extend various analytical derivations.

The two extensions to learning game theory presented here abandon the orthodox assumption that players are fully rational, and assume instead that players follow one of two alternative decision-making processes –case-based reasoning or reinforcement learning– that have received strong support from cognitive science research. The formal results derived in this part of the thesis add to the growing body of work in learning game theory that supports the general principle that the stability of outcomes in games depends not only on how unilateral deviations affect the deviator but also, and crucially, on how they affect the non-deviators. Outcomes where unilateral deviations hurt the deviator (strict Nash) but not the non-deviators (protected) tend to be the most stable.

The contribution of this thesis to evolutionary game theory is a systematic study of the extent to which the assumptions made in mainstream evolutionary game theory for the sake of tractability are affecting its conclusions. Our results show that the type of strategies that are likely to emerge and be sustained in evolutionary contexts is strongly dependent on assumptions that traditionally have been thought to be unimportant or secondary (e.g. number of players, continuity of the strategy space, mutation rate, population structure…). This latter contribution is focused on the evolutionary emergence of cooperation.

Following the presentation of the main results and the discussion of their implications, this thesis provides some guidance on how the models analysed here could be parameterised and validated.

TABLE OF CONTENTS

1. Introduction

This thesis advances game theory by formally analysing the implications of some of its most stringent assumptions. The approach followed here consists in examining the consequences of replacing some of the assumptions made in game theory for the sake of mathematical tractability with alternatives that –at least in some contexts– are more plausible. The method employed to conduct this research has been a symbiotic combination of computer simulation and mathematical analysis. Our results suggest that some of the most fundamental assumptions embedded in game theory may have deeper philosophical implications than commonly assumed.

1.1. Motivation

The value of advancing game theory seems clear: it is widely agreed that game theory has become one of the cornerstones of the social sciences (Hargreaves Heap and Varoufakis, 1995). There are widespread claims that it "provides solid microfoundations for the study of social structure and social change" (Elster, 1982), and that it "may be viewed as a sort of umbrella or 'unified field' theory for the rational side of social science" (Aumann and Hart, 1992). More recently, Gintis (2000) has stated that "game theory is a universal language for the unification of the behavioral sciences". Even in the biological sciences it has been argued that some game theoretical concepts represent "one of the most important advances in evolutionary theory since Darwin" (Dawkins, 1989).

However, while extremely informative, 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 individuals' assumptions about other individuals' 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 individual attributes to every other individual, game theory cannot reduce the set of expected outcomes significantly.

Thus, whilst acknowledging that the work conducted in game theory has been tremendously useful, a growing inter-disciplinary community of scientists think the time has come to extend game theory beyond the boundaries of full rationality, common-knowledge of rationality, consistently aligned beliefs, static equilibria, and long-term convergence. These concerns have led various researchers to develop formal models of social interactions within the framework of game theory, but relaxing its most stringent assumptions. Such models are providing not only valuable insights for the specific questions they address, but also the basis to assess how robust the results obtained in classical game theory are. This thesis is a contribution to this emergent programme of research.

1.2. Aim, approach and methodology

The overall aim of this thesis is to advance non-cooperative game theory by formally studying the implications of some of its assumptions that have been made for the sake of tractability and are not generally supported by empirical evidence. This has been done following two approaches:

- The first approach consists in examining the formal implications of replacing the unsupported assumptions in mainstream non-cooperative game theory relating to individual decision-making with assumptions that stem from empirical research. In particular, this thesis abandons the assumptions of complete information, common knowledge of rationality and consistently aligned beliefs, and contemplates instead members of two classes of decision making algorithms that have received strong support from cognitive science research: reinforcement learning and case-based reasoning.
- The second approach is used to extend mainstream evolutionary game theory. It consists in exploring the implications of a wide range of competing assumptions –all of them consistent with the essence of the theory of evolution– within a common framework. The results obtained using different assumptions are then contrasted in a coherent and systematic way.

In terms of methodology, there are four features that distinguish the work conducted in this thesis from most of the previous research undertaken in the same emerging field.

- First, the contributions made in this thesis have been placed in an overall framework that can encompass, in admittedly broad terms, most of the research conducted in game theory until now. This has permitted a more transparent comparison between the assumptions investigated here and those that have been addressed so far, and also between the results derived from this research and those obtained under other assumptions.
- Secondly, in terms of method, since most of the assumptions investigated in this thesis have not been formulated to allow for mathematical tractability, but to advance our formal understanding of social interactions in real life, new methodologies have had to be employed to supplement mathematical analyses. In particular, computer simulation has been used extensively to enhance and complement mathematical derivations. These two techniques have been combined in a way that is not common in the literature of game theory or in the field of social simulation. To be specific, most of the simulations reported in this thesis are just small advances at the edge of theoretical understanding. They are advances sufficiently small so that simplified versions of them (or certain aspects of their behaviour) can be fully understood in mathematical terms –thus retaining analytical rigour–, but they are steps large enough to significantly extend our understanding beyond what is achievable using the most advanced mathematical techniques available. In this way, simulations will be shown to extend theoretical knowledge in a rigorous, formal, and almost continuous way (Probst, 1999).
- The symbiotic use of mathematical analysis and computer simulation has allowed us to characterise both the short-term and the long-term dynamics of the models investigated in this thesis. This is in contrast with most game theoretical research –which is most often concerned with the identification of asymptotic equilibria– and with most research in the field of social simulation –which is often mainly concerned with the short-term dynamics.
- Finally, a great effort has been made to ensure that all models and simulations reported in this thesis can be easily scrutinised, used, replicated and reimplemented by independent researchers. In particular, all the computer programs used to conduct the research presented here have been released under the GNU general public licence (GPL), which is one of the licences

that scores best against the criteria set out by Polhill and Edmonds (2007) for releasing scientific software. GNU GPL grants the right to inspect, copy and distribute the source code, to modify it, and also to copy and distribute any modifications. It also guarantees that any modifications will be issued under a licence that preserves these rights (i.e. copyleft protection). Furthermore, following Polhill and Edmonds' (2007) guidelines, a substantial amount of work has been devoted in this thesis to *facilitate* the process of scientific critique of this research, by carefully commenting the code, providing extensive documentation, and creating several user guides for all the developed software. All the programs and documentation are included in the Supporting Material of this thesis.

1.3. Overall framework and specific contributions

To appreciate more precisely the specific contribution of this thesis to human knowledge, it becomes necessary to formalise some terms related to game theory first. In this thesis, a clear distinction is made between game theory used *as a framework*, and the different branches of non-cooperative game theory as we know them nowadays – e.g. classical game theory, evolutionary game theory and learning game theory.

Game theory as a framework is a methodology used to build models of real-world social interactions. The result of such a process of abstraction is a formal model that typically comprises the set of individuals who interact (called *players*), the different choices available to each of the individuals (called *strategies*), and a *payoff* function that assigns a (usually numerical) value to each individual for each possible combination of choices made by every individual. In most branches of game theory, payoffs represent the preferences of each individual over each possible outcome of the social interaction. The notable exception is evolutionary game theory, where payoffs most often (but not always) represent Darwinian fitness.

The feature of the social interaction to be modelled that makes game theory a particularly useful framework to employ is its *strategic* nature: the fact that the outcome of the interaction for any individual player generally depends not only on her own choices, but also on the choices made by every other individual. Thus, game theory could arguably be defined as "the theory of interdependent decision-making" (Colman, 1995, pg. 3).

Game theory *used as a framework* provides a formal description of the social setting where the players are embedded. Importantly, it does not account for the players' behaviour, neither in a normative nor in a positive sense. It is just not the realm of game theory *as a framework* to do so. It is only when different assumptions about how players behave –or should behave– are included in the framework, that game theory *as a framework* gives rise to the different branches that compose game theory as we know it nowadays. Here we outline the main features of the three most developed branches of deductive non-cooperative game theory at this time:

<u>Classical game theory</u>: Classical game theory was chronologically the first branch to be developed (Von Neumann and Morgenstern, 1944), the one where most of the work has been focused historically, and the one with the largest representation in most game theory textbooks and academic courses. Classical game theory is a branch of mathematics devoted to the study of how instrumentally rational players should behave in order to obtain the maximum possible payoff in a formal game.

The main problem in classical game theory is that, in general, rational behaviour for any one player remains undefined in the absence of strong assumptions about other players' behaviour. Hence, in order to derive specific predictions about how rational players should behave, it is often necessary to make very stringent assumptions about everyone's beliefs (e.g. common knowledge of rationality) and their interdependent consistency. Since such strong assumptions rarely hold in the real world, it is not surprising that when game theoretical solutions have been empirically tested, disparate anomalies have been found (see, for example, work reviewed by Colman (1995) in chapters 7 and 9, Roth (1995), Ledyard (1995), and Camerer (2003)). To make matters worse, even when the most stringent assumptions are in place, it is often the case that several possible outcomes are possible, and it is not clear which –if any– may be achieved, or the process through which this selection would happen. Thus, the general applicability of classical game theory is limited. A related limitation of classical game theory is that it is an inherently static theory: it is mainly focused on the study of end-states and possible equilibria, paying hardly any attention to how such equilibria might be reached.

Evolutionary Game Theory: Some time after the emergence of classical game theory, biologists realised the potential of game theory as a framework to formally study adaptation and coevolution of biological populations (Lewontin, 1961; Hamilton, 1967). For those situations where the fitness of a phenotype is independent of its prevalence, optimisation theory is the proper mathematical tool. However, it is most common in nature that the fitness of a phenotype depends on the composition of the population (Nowak and Sigmund, 2004). In such cases game theory becomes the appropriate framework.

In evolutionary game theory, players are no longer taken to be rational. Instead, each player –most often meant to represent an individual animal– always selects the same (potentially mixed) strategy[1] –which represents its behavioural phenotype–, and payoffs are usually interpreted as Darwinian fitness. The emphasis is then placed on studying which behavioural phenotypes (i.e. strategies) are stable under some evolutionary dynamics, and how such evolutionary stable states are reached. Despite having its origin in biology, the basic ideas behind evolutionary game theory –that successful strategies tend to spread more than unsuccessful ones, and that fitness is frequency-dependent– have extended well beyond the biological realm.

The main shortcoming of mainstream evolutionary game theory is that it is founded on assumptions made to ensure that the resulting models are mathematically tractable. Most of the work assumes one single infinite and homogeneous population where players using one of a finite set of strategies are randomly matched to play an infinitely repeated 2-player symmetric game. In the last few years various alternative models (e.g. finite populations, stochastic

[1] This assumption, which is not always made in models of *cultural* evolution, is explained in detail in chapter 2.

strategies, multi-player games, structured populations) are being explored, but unsystematically.

Learning game theory: Like evolutionary game theory, learning game theory abandons the demanding assumptions of classical game theory on players' rationality and beliefs. However, unlike its evolutionary counterpart, learning game theory assumes that individual players adapt, learning over time about the game and the behaviour of others (e.g. through reinforcement, imitation, or belief updating). This learning process is *explicitly* modelled (Vega-Redondo, 2003, pg. 398). 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 classical 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 most studied classes of decision-making algorithms in the literature are: reinforcement learning (with experimental studies conducted by e.g. Erev et al. (1999), theoretical work done by e.g. Bendor et al. (2001b), 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)), and the EWA (Experience Weighted Attraction) model (Camerer and Ho, 1999), which is a hybrid of reinforcement and belief learning.

This thesis makes two specific contributions to the development of learning game theory and one in the field of evolutionary game theory. The first contribution to learning game theory is to elucidate the implications of assuming that players use a simple form of reinforcement learning as decision-making algorithm. Reinforcement learning, being one of the most widespread adaptation mechanisms in nature, has attracted the attention of many scientists and engineers for decades. This interest has led to the formulation of various different models and –when feasible– to the theoretical analysis of their dynamics. This thesis provides an in-depth analysis of the transient and asymptotic dynamics of one of the best known stochastic models of reinforcement learning (Bush and Mosteller, 1955) for 2-player 2-strategy games.

The second contribution to learning game theory is a detailed exploration of the implications of case-based reasoning as decision-making approach in strategic contexts. Case-based reasoning consists in "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 lived in the past. Such experiences are stored in the form of cases. The distinguishing feature of case-based reasoning as problem-solving mechanism is that "thought and action in a given situation are guided by a single distinctive prior case" (Loui, 1999). To our knowledge, the implications of this type of reasoning in strategic contexts have not been explored before.

Finally, the contribution of this thesis to evolutionary game theory is a systematic exploration of the impact of various assumptions made in this field; this exploration is undertaken by studying the structural robustness of evolutionary models of cooperation using a computational tool built for this specific purpose: EVO-2x2. EVO-2x2 is a computer simulation modelling framework designed to formally investigate the evolution of strategies in 2-player 2-strategy (2x2) symmetric games under various competing assumptions.

A significant part of the work conducted in this thesis is sufficiently general to be valid in a wide range of social interactions, but some of it has had to be focused on particular types of social interactions. Whenever there has been a need to select a specific type of social interaction to investigate (even if the only purpose was to illustrate the applicability of more general findings), we have always studied social dilemmas (Dawes, 1980). Social dilemmas are social interactions where individual rationality leads to outcomes for which there is at least one feasible alternative preferred by everyone. In such situations, decisions that make sense to each individual can aggregate into outcomes in which everyone suffers (Macy and Flache, 2002). The focus of this thesis has been on social dilemmas because of their importance in the social and biological sciences, and because the predictions of classical game theory in this context clash with widely shared intuitions and empirical results (see, for instance, work reviewed by Gotts et al. (2003b) and by Colman (1995) in chapters 7 and 9).

1.4. Outline of the thesis

The structure of this thesis is as follows: chapter 2 outlines the main assumptions made in game theory. We analyse each of the following branches in turn: game theory used as a framework, classical game theory, evolutionary game theory, and learning game theory. This critical review of the main assumptions made in deductive game theory will serve as a framework to clearly identify those assumptions that will be abandoned in the subsequent chapters of this thesis, and those that will be retained. Chapter 3 clarifies the scope of this thesis within game theory and explains social dilemma games in detail. It also describes the methods that have been used to formally analyse the models developed in chapters 4, 5 and 6. Chapter 4 is an in-depth analysis of the transient and asymptotic dynamics of the Bush-Mosteller reinforcement learning algorithm for 2-player 2-strategy games. Chapter 5 is an exploration of cased-based reasoning as decision-making algorithm in strategic contexts. Chapter 6 describes EVO-2x2, the modelling framework developed in this thesis to assess the impact of various assumptions made in mainstream evolutionary game theory for the sake of mathematical tractability. The use of EVO-2x2 is illustrated by conducting an investigation on the structural robustness of evolutionary models of cooperation. Chapter 7 is a general discussion of the results obtained in chapters 4, 5 and 6. We also discuss the value of the models developed in this thesis, and how they could be validated. Chapter 8 summarises the main conclusions of this work and identifies areas for further research. The proofs of the theoretical results derived in this thesis can be found in the appendices. This thesis also comprises extensive supporting material, including the source code of every computer program we have used in this research, together with user guides and instructions to replicate every experiment reported here.

2. Main assumptions in game theory

This chapter is a critical dissection of the main assumptions embedded in each of the most advanced branches of deductive game theory at this time. We distinguish between game theory as a framework (which makes no assumptions about individuals' behaviour or beliefs), classical game theory, evolutionary game theory, and learning game theory. Given the breadth and depth of game theory work, this thesis cannot present an exhaustive list of all the assumptions considered in the field. We focus on the most prevalent and relevant ones. The critical review of deductive game theory in this chapter is meant to serve as a framework where the assumptions whose impact is investigated in the subsequent chapters of this thesis can be precisely identified. It will also serve to identify what assumptions are retained in the models developed in this thesis. The last section of this chapter briefly describes some of the branches of game theory that are not purely deductive.

2.1. Game theory as a framework

Game theory as a framework is a methodology used to build models of real-world social interactions. The result of the modelling exercise is a game, i.e. a formal abstraction of the social interaction which is typically defined by[2]:

- the set of individuals who interact (called *players*),
- the different choices available to each of the individuals (called *strategies*),
- and a *payoff* function that assigns a (usually numerical) value to each individual for each possible combination of choices made by every individual.

Importantly, the abstract model developed within this framework does not make any assumptions about the players' behaviour, neither in a normative nor in a positive sense.

[2] We use here the representation of a game in strategic form for the sake of clarity. If the sequential structure of the game is complex (in terms of order of movement, players' asymmetries and information flow), the representation of the game in extensive form (which *explicitly* details the order of events, the order of moves, and the information sets) may be more adequate (see chapter 1 in Vega-Redondo (2003) for details).

Game theory as a framework is particularly useful to describe and analyse decision-making in social interactions where the outcome potentially depends on the decisions made by several individuals (i.e. interdependent decision-making processes). According to the Stanford Encyclopaedia of Philosophy, "game theory is the most important and useful tool in the analyst's kit whenever she confronts situations in which what counts as one agent's best action (for her) depends on expectations about what one or more other agents will do, and what counts as their best actions (for them) similarly depend on expectations about her" (Ross, 2006).

As with any formal model, some of the complexity of the real-world situation represented will be lost in the process of abstraction. The rationale to undertake such a process of abstraction, which implies loss of descriptive accuracy to some extent, is that it will yield insights beyond those that could be achieved without the model. Furthermore, the knowledge acquired from the analysis of the abstract formal model can still be valid in other real-world situations whose important features are captured by the same formal model even though the model was not initially developed with such situations in mind. To the extent that the formal model captures the essence of the situation under study, enables us to establish inference processes that we could not undertake otherwise, and yields insights that can be transferred to other domains, we consider that the formal model is useful (Colman, 1995, pg. 6).

Game theory as a framework makes two important assumptions. The first one is ontological and it refers to how social interactions are modelled in game theory. The framework used in game theory makes a clear distinction between structure (i.e. rules of the game) and action. The rules of the game fully constrain the set of possible actions that can be taken, i.e. there is no room for action to change structure. Obviously this is not the only ontological view that one can take when trying to distil the essence of social interactions. This clear cut between structure and action will prove useful in many circumstances, but it may not always be adequate; therefore it is important to be aware that there are many other ways of modelling social interactions (Hargreaves Heap and Varoufakis, 1995, chapter 1).

Assuming that the essence of the social interaction to be modelled is captured by the formal abstraction to a satisfactory extent (in terms of context, interplay between action and structure, history effects…) the most important assumption made when using game theory as a framework relates to the definition of the *payoff* function. In most branches of game theory, payoffs are meant to represent individuals' preferences for each possible outcome of the social interaction. The notable exception is evolutionary game theory, where payoffs most often (but not always) represent Darwinian fitness. The following two sections explain this in detail.

2.1.1. Payoffs interpreted as preferences

The payoff function for each player is effectively a preference ordering over the set of possible outcomes. Behind the concept of "payoff function" is the implicit assumption that preferences will guide action (otherwise there would not be much point in defining a payoff function). While seemingly innocuous, this underlying assumption does have certain philosophical implications which, though fascinating, fall out of the scope of this thesis (Hargreaves Heap and Varoufakis, 1995, pg. 12).

A common misconception about game theory relates to the roots of players' preferences. There is no assumption in game theory (not even as a framework) that players' preferences are formed in complete disregard of each other's interests. On the contrary, preferences in game theory are assumed to account for everything, i.e. they may include altruistic motivations, moral principles, and social constraints, for example (Colman, 1995, pg. 301; Vega-Redondo, 2003, pg. 7).

Game theory as a framework assumes that players' preference order is well defined, i.e. it satisfies the conditions of reflexivity, completeness, and transitivity (Hargreaves Heap and Varoufakis, 1995, pg. 6); and that their preference order does not change. If no further assumption is made on individuals' preferences, these are said to be *ordinal*. Ordinal preferences provide no information about the strength of preferences, so arithmetic operations on ordinal payoffs are not meaningful. An admittedly obvious point, but one which may be worth noting, is that direct comparisons of ordinal preferences between different players (e.g. "player A likes outcome X more than player B does") are meaningless.

In almost all game theoretical models, however, preferences are assumed to be cardinal, i.e. payoffs take numerical values on an interval scale. With this assumption, payoffs give a measure of the strength of the preferences, and therefore payoff differences are indeed meaningful. If nothing more than cardinality is assumed, comparisons of preferences between different players are still meaningless.

Most game theoretical models go beyond the assumption of cardinal preferences: they interpret payoffs as von Neumann-Morgenstern utilities (Colman, 1995, section 2.1; Hargreaves Heap and Varoufakis, 1995, section 1.2; Vega-Redondo, 2003, pg. 7). The benefit of making such a strong assumption is that it allows game theorists to use expected utility theory to evaluate probability distributions over possible outcomes of the game. (Note that payoffs relate to outcomes that are certain). It is important to remember that these models are –implicit or explicitly– assuming considerably more about players' preferences than just cardinality: cardinality by itself is not enough to formally justify models where individuals maximise expected payoffs. Expected payoff maximisation requires preferences to be well defined (see above) and three extra assumptions: continuity, preference increasing with probability, and independence (Hargreaves Heap and Varoufakis, 1995, pg. 10). When all these assumptions hold, payoffs embody players' attitudes to risk, and then it is true that an individual who acts on her preference ordering acts *as if* she is maximising her expected payoff (see chapter 2 in Colman (1995) for details).

Finally, the strongest assumption on preferences relates to social comparisons. There are (relatively few) models where payoffs interpreted as preferences are compared across players. This is a very strong assumption which finds its roots in the social philosophy of utilitarianism, and is not commonly observed in game theoretical models that interpret payoffs as preferences; however, it can certainly be found in the literature (see e.g. Bendor et al. (2004)). In stark contrast, it will be shown in the next section that most models in evolutionary game theory interpret payoffs as fitness, and they actually *require* comparing the payoffs obtained by different players (and often performing arithmetic operations with them).

2.1.2. Payoffs in evolutionary models

In evolutionary game theory models, the emphasis is not so much on the players, but on the strategies. In fact, it is most often understood that each player is pre-programmed to play a certain (pure or mixed) strategy, thus establishing equivalence between players and strategies. The interest then lies in studying the evolution of large *populations* of players who repeatedly interact to play a game. The aim is identifying which strategies (i.e. type of players) are most likely to thrive in this "ecosystem" and which will be wiped out by selection forces. In this context, payoffs are not interpreted as preferences, but as a value that measures the success of a strategy in relation to the others. Selection forces then act to favour strategies with higher payoffs. Thus, in models of biological (as opposed to cultural) evolution, payoffs are most often interpreted as Darwinian fitness. The crucial point here is that payoffs obtained by different players will be compared and used to determine the relative frequency of different types of players (i.e. strategies) in succeeding generations. This may not be a major assumption when modelling biological evolution, but it is one that cannot be ignored if evolution is interpreted in cultural terms.

2.2. Classical game theory

Classical game theory is devoted to the study of how instrumentally rational players should behave in order to obtain the maximum possible payoff in a formal game. Thus, as a deductive and normative branch of game theory, one could argue that classical game theory itself is incapable of being empirically tested and falsified (Colman, 1995, pg. 6). What we can clearly infer from the combination of empirical research and game theory is that, if empirical observations clash with game theoretical solutions, then (a) the observed real-world situation does not correspond to the abstracted game, or (b) at least one assumption made by game theory does not hold (or both (a) and (b)). Hence the importance of clearly identifying the assumptions made in classical game theory. The following sections analyse the two most relevant ones: complete availability of information and instrumental rationality.

2.2.1. Availability of information

A major assumption embedded in classical game theory (CGT) relates to information availability. This is a key issue, since information availability crucially affects what course of action may be regarded as rational. As an example, if players did not know anything about the game (not even its strategic nature) beyond the payoff they obtain after playing certain actions, many very simple learning models could be regarded as rational. CGT is mostly concerned with games of complete information. In these games, it is assumed that players not only know the rules of the game and their own payoffs, but also their counterparts' payoff functions. Furthermore, complete availability of information is assumed to be common knowledge. Common knowledge (CK) in game theory often comes with a certain order: zero-order CK of X is just the assumption that X prevails for every player (e.g. zero-order common knowledge of complete information (CKCI) means that every player has complete information); first-order CK is the assumption that every player knows that X prevails for every player (e.g. first-order CKCI means that every player knows that every player has complete information); in general, (n)th-order CK is the assumption that (n-1)th-order CK is known by every player. If no order is specified, it is assumed that the order is infinite (this produces an infinite recursion of shared assumptions). For different accounts of the meaning of common knowledge see Vanderschraaf and Sillari (2007).

CGT also considers games of incomplete information. As a matter of fact, if one is happy to accept certain (strong) conditions on what may count as a "rational belief", the distinction between complete and incomplete information is not essential, since games of incomplete information can be easily transformed into games of complete information (Harsanyi, 1967a, b, 1968). The basic idea behind this transformation consists in assuming that there are different "types of players", each of them with a different payoff function. Then, one must see each player's uncertainty about her counterparts' payoff functions as deriving from the player's uncertainty about which types of players her counterparts are. Finally, the transformation requires applying Harsanyi and Aumann's argument about the impossibility of players with mutual knowledge of rationality "agreeing to disagree" (Aumann, 1976). This last step ensures that rational players hold common beliefs about the probabilities that their counterparts will turn out to be of one type or another. Once this assumption is made, the analysis of the game with incomplete information is essentially the same as one of complete information.

2.2.2. Instrumental rationality

The concept of instrumental rationality in classical game theory finds its clearest roots in Hume's *Treatise on Human Nature*. In CGT rationality is understood as the capacity of identifying the actions that best satisfy the person's predefined objectives (Hargreaves Heap and Varoufakis, 1995, pg. 7), i.e. rationality plays no role in setting objectives. This basically means that instrumentally rational players have unlimited computational capacity devoted to maximise their individual payoff function, which is defined in advance. The assumption of rationality in CGT has been widely challenged. One of the alternatives that has received great attention is Simon's (1957) original concept of procedural rationality, later recast as bounded rationality (Simon, 1982) mainly for modelling purposes. Simon (1982) emphasises that people have limited knowledge of their situations, limited ability to process information, and limited time to make choices.

In any case, the main challenge within CGT comes from the fact that in most games there is no maximising strategy for any given player regardless of her counterparts' actions, i.e. rationality remains undefined in the absence of beliefs about what the other players will do. Naturally, this belief-dependency of rationality has led to different concepts of rationality based on different assumptions about what beliefs about other players' behaviour are allowed. The following sections explain the three most important approaches, namely:

1. Dominance reasoning.
2. Rationalisable strategies.
3. Consistently aligned beliefs: Nash equilibrium.

It is worth mentioning at this point that –most often– *the three* approaches outlined above make use of two extra assumptions, namely: common knowledge of complete information (CKCI; explained in the previous section), and common knowledge of rationality (CKR). Following the definition of common knowledge outlined in the previous section, first-order CKR is the assumption that every player knows that every player is rational (rationality is understood following one of the 3 interpretations mentioned above); (n)th-order CKR is the assumption that (n-1)th-order CKR is known by every player. If no order is specified, it is assumed that the order of CKR is infinite (see Aumann (1976) for a formal definition). CKCI and CKR are embedded in the definitions of approaches (2) and (3) mentioned above. Without assuming CKCI and CKR, most games are not solvable regardless of the approach taken. For the sake of clarity the following subsections will discuss the role of CKR assuming that CKCI comes with it.

Dominance reasoning

Rationality can be minimally identified with "not playing (strictly) dominated strategies"[3] (Vega-Redondo, 2003, pg. 32). This view of rationality does not require any assumption about the behaviour of other players: there is no belief that a player could hold about the other players' strategies such that it would be optimal to select a dominated strategy. In general, one has the option to reject only those strategies that are dominated by other pure strategies or, alternatively, choose to reject the (potentially larger) set of strategies that are dominated by some mixed strategy.

The elimination of dominated strategies by each player rarely leads to one single profile of strategies (the one-shot Prisoner's Dilemma is an exception for this), so CKR is usually brought into play. CKR allows the process of successive elimination of dominated strategies: with this interpretation of rationality, first-order CKR means that players assume that no player will select a dominated strategy. The elimination of certain strategies when assuming (n)th-order CKR may open the door to eliminate more strategies by assuming (n+1)th-order CKR.

[3] For a player A, strategy S_A is (strictly) dominated by strategy S^*_A if for each 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).

This iterative process goes on until no strategies can be further eliminated. When this process leads to one single strategy for every player (i.e. one single outcome) then the game is said to be dominance solvable.

Rationalisable strategies

A stronger interpretation of rationality dictates that rational players maximise their expected payoff on the basis of *some* expectations about what the others will do (clearly this interpretation prevents players from playing dominated strategies). Using this concept of rationality and assuming CKR leads to the definition of rationalisable strategies: rationalisable strategies are those that remain after making such assumptions (Bernheim, 1984; Pearce, 1984). The term rationalisable derives from the fact that every player can defend choosing such a strategy (i.e. rationalise it) on the basis of beliefs that are *consistent* with the assumption of CKR. However, given that each player may have many different rationalisable strategies (by holding different beliefs about her counterparts' beliefs), it could well be the case that once the game is played (i.e. once every player has selected a specific rationalisable strategy), some of these beliefs are proven wrong. To be clear, a set S of rationalisable strategies (one for each player) may derive from beliefs where one of the players is assuming that one of her counterparts will select a (rationalisable) strategy different from the one assigned to this counterpart in the set S itself. Informally, this would occur if one of the players presumes that one of her counterparts will "make a mistake" by expecting something that the player does not intend to do (even though this "mistaken belief" is perfectly consistent with CKR). In other words, the beliefs underlying rationalisable strategies must be consistent with the assumption of CKR for each individual player independently, but they may be inconsistent across players. Hargreaves Heap and Varoufakis (1995, pp. 51-52) give a 2-player example where both players select a rationalisable strategy on the basis of beliefs that are inconsistent across players. The following section explains that imposing consistency of beliefs across players leads to the (stronger) concept of Nash equilibrium.

Let us conclude this section by relating the concept of rationality explained here and that assumed when conducting dominance reasoning (see previous section).

As mentioned above, rationalisable strategies are necessarily undominated; a natural question is then: are iteratively undominated strategies always rationalisable? The answer to this question for 2-player games is yes (Pearce, 1984). In other words, for two player games these two concepts are equivalent. This is not true, however, for games involving more than two players. In such games, there can be iteratively undominated strategies that are not best response to any strategy profile. The subtle difference between these two concepts of rationality is brilliantly explained by Vega-Redondo (2003, pp. 66-68).

Consistently aligned beliefs: Nash equilibrium

The previous section showed that if players select rationalisable strategies, the outcome of the game may be such that the beliefs of some players are proven wrong by the choices actually made by other players. The concept of Nash equilibrium derives from imposing the additional constraint that beliefs must be consistently aligned across players. Thus, a Nash equilibrium is a set of rationalisable strategies (one for each player) whose implementation confirms the expectations of each player about the other players' choices (Hargreaves Heap and Varoufakis, 1995, pg. 53). A *corollary* of this definition is that Nash equilibria are formed by sets of strategies that are best replies to each other. This simple shortcut through the cumbersome web of players' beliefs over their counterparts' beliefs is probably one of the main factors that explain the success of the Nash equilibrium (NE) in the social sciences. Another reason is that NEs can be interpreted in a number of meaningful and useful ways (Holt and Roth, 2004). The concept of NE, however, is not free from problems. There are many games without any NE in pure strategies, and many others with more than one. In these cases, the assumption of consistently aligned beliefs is particularly problematic. How can players coordinate their beliefs in the absence of communication? The problem of multiple NE is particularly acute in repeated games, as illustrated by the extensive variety of "folk theorems" available in the literature. In broad terms, "folk theorems" demonstrate that repeated interactions typically allow for a wide range of equilibrium behaviour. Vega-Redondo (2003, chapter 8) reviews several "folk theorems", differing in their time horizon (finite or infinite), information conditions (complete or incomplete information, and perfect or imperfect observability), and equilibrium concept (Nash or subgame perfect).

Let us conclude this section by stating that the concept of NE is significantly stronger than that of rationalisable strategies. In particular, Bernheim (1984) showed by example that one can find rationalisable strategies that are not part of any NE (i.e. there is no NE that assigns a positive weight to them). In other words, there are outcomes where all players are selecting a rationalisable strategy, and which cannot be interpreted as the result of a mis-coordination among players that were hoping to arrive at a NE. This clearly indicates that the notion of rationalisability embodies something broader than equilibrium mis-coordination (Vega-Redondo, 2003, pg. 65).

Refinements of Nash equilibrium

The problem of multiple Nash equilibria outlined in the previous section has led to the proposal of countless refinements aimed at eliminating those NEs that are not considered plausible or desirable for not fulfilling some additional condition (see van Damme (1987) for a comprehensive study). Unfortunately, so many refinements have been developed by now that "in many games which have multiple Nash equilibria, each equilibrium could be justified by some refinement present in the literature" (Alexander, 2003). In this section we briefly present only one, namely "trembling hand perfection" in its strategic-form version (see Vega-Redondo, 2003, chapter 4), since the idea underlying this refinement will be used extensively in this thesis.

The "trembling hand perfect" refinement, which was proposed by Selten (1975), eliminates those Nash equilibria that are not robust to small mistakes. The refinement process assumes that players' hands may tremble, i.e. players may select an unintended action (i.e. deviate from the equilibrium) with small probability. An alternative view of the same phenomenon is that players may experiment with small probability. Some NEs may resist the possibility of these trembles and some may not: those NEs that do not survive arbitrarily small trembles are eliminated. Slightly more formally, the set of trembling hand perfect equilibria in a game is the limit of the sequence of Nash equilibria in perturbed versions of the game (i.e. versions of the game played with trembles) as the probability of trembles goes to zero. In 2-player strategic-form games, an equilibrium is perfect if and only if it is a Nash equilibrium that involves no weakly dominated strategies by either player (Van Damme, 1987, Theorem 3.2.2). The reasoning behind this refinement will prove to be very useful to reduce the set of possible outcomes of the game in the models developed in chapters 4 and 5 of this thesis.

2.3. Evolutionary game theory

Many biological and socio-economic systems are governed, at least to some extent, by evolutionary pressures. Such evolutionary systems may be composed of entities of very different nature, such as molecules, cells, genes, animals, organisations, ideas, behaviours… but they all share the three common features that characterise any evolutionary system: diversity, selection, and replication.

Diversity: entities in the system are not all the same; they show dissimilarities that affect their so-called individual fitness. Fitness is just a measurable indicator that determines how a population of entities evolves: entities with higher fitness will tend to spread relatively more than those with lower fitness. The precise mechanism that links current fitness with future population composition is the selection mechanism, which is explained in the next point. Note that in general this selection mechanism reduces the diversity of the system, since it favours some existing entities over others. There may be, however, mechanisms that tend to preserve the heterogeneous nature of the system: most evolutionary systems are subject to processes that create and maintain diversity. This diversity-generating mechanism acts in the opposite direction to the selection force, and it is the only mechanism that may preclude the system from locking-in. In biological systems, diversity generally stems from genetic mutations whereas in many socio-economic systems, it is innovations, asymmetries in the flow of information, or even simple mistakes, which are often responsible for the incessant appearance of different forms of behaviour. The process by which new entities appear in an evolutionary system is usually called mutation in biological contexts and experimentation or innovation in socio-economic contexts.

Selection: The mechanism of selection is a discriminating force that favours some specific entities rather than others. By selecting only certain entities from the population, this selection force diminishes the heterogeneity of the system. As mentioned above, the criterion by which some entities are selected among the population rather than others is usually called fitness. In evolutionary game theory strategies (which may be seen as behavioural phenotypes) are selected on the basis of the payoff they obtain, i.e. the relative frequency of strategies which obtained higher payoffs in the population will increase at the expense of those which obtained relatively lower payoffs.

Replication / Inheritance / Preservation: The properties of the entities in the system (or the entities themselves) are preserved, replicated or inherited from one generation to the next at least to some extent. Replication mechanisms can be carried out through a range of processes, from genetic transmission in biological systems to social learning processes such as imitation in some socio-economic contexts.

The main assumption underlying evolutionary thinking is that the entities which are *more successful*[4] at a particular time will have the best chance of being present in the future. In biological and economic contexts, this assumption often derives from competition among entities for scarce resources or market shares. In social contexts, evolution is often understood as *cultural* evolution, where this refers to changes in behaviour, beliefs, or social norms over time (Alexander, 2003), and may be justified by "the tendency of human behaviour to adjust in response to persistent differentials in material incentives" (Sethi and Somanathan, 1996, pg. 783).

Evolutionary game theory (EGT) is devoted to the study of the evolution of strategies. In biological systems, players are most often assumed to be pre-programmed to play one given strategy, so studying the evolution of a population of strategies becomes formally equivalent to studying the evolution of a population of players. By contrast, in socio-economic models, players are usually assumed to live forever, and switch their strategy following evolutionary pressures. The role of players relative to the role of strategies is irrelevant for the formal analysis of the system, where –in both cases– it is strategies that are

[4] Note that this is a measure of *relative* performance.

actually subjected to evolutionary pressures. Thus, without loss of generality and for the sake of clarity, we take here the biological stand and assume that players select always the same strategy.

Thus, EGT is devoted to the study of large *populations* of players who repeatedly interact to play a game. Strategies are subjected to selection pressures in the sense that the relative frequency of strategies which obtain higher payoffs in the population will increase at the expense of those which obtain relatively lower payoffs. The aim is to identify which strategies (i.e. type of players) are most likely to thrive in this "evolving ecosystem of strategies" and which will be wiped out by selective forces. As mentioned before, payoffs in evolutionary contexts are not interpreted as preferences; instead they provide the value that is used to measure the relative success of one strategy in relation to the others.

2.3.1. Evolutionary stability: evolutionary stable strategies

The study of dynamic systems often begins with the identification of their stable states. This is often called static analysis, as it does not consider the dynamics of the system explicitly, but only its rest points. The most important concept in the static analysis of EGT is the concept of Evolutionary Stable Strategy (ESS), proposed by Maynard Smith and Price (1973). Very informally, a population playing an ESS is *uninvadable* by any other strategy (Weibull, 2002). To be more precise, consider a very large population of players who are repeatedly drawn at random to play a 2-player symmetric game. Initially all players are selecting the same (incumbent) strategy. That strategy is an ESS if there exists a positive invasion barrier such that for any given mutation that may occur and assuming that the population share of individuals playing the mutant strategy falls below this barrier, the incumbent strategy earns a higher payoff than the mutant strategy (Weibull, 1995, pg. 33). The original concept of ESS has proven to be tremendously useful, but it is important to be aware of the assumptions underpinning its theoretical framework: the ESS is derived for a system composed of a *single infinite* population of individuals who are repeatedly *randomly* drawn to play a *2-player symmetric* game; furthermore, it only considers *monomorphic* populations (all individuals are playing the same strategy) which can be invaded by only *one type of mutant strategy at a time*.

In particular, the assumption of one single *infinite* population has a number of important implications. For a start, this assumption is in effect a mean-field approximation used to equate the average payoff actually obtained by a population with the expected value of a probability distribution of payoffs (which would be obtained by explicitly modelling players' interactions). It is also the assumption that justifies treating as equivalent a mixed strategy and a population profile where pure strategies are played in the population with the frequency induced by the corresponding probability in the mixed strategy (see Vega-Redondo, 2003, pp. 356-7). Finally, it effectively eliminates the impact of arbitrarily small invasions on the incumbent population. This last point is best explained with a simple example. Consider a 2-player population where player i can impose a punishment of magnitude P on player j at a cost of $C < P$. Clearly, punishing j would give a relative advantage to i over j, so this behaviour would be evolutionary favoured. Now consider a large population of potentially punishable players j, and think of the effect of the same single punishment conducted by one mutant i on one of the players in the incumbent population. Player i will incur the cost C, but the average payoff of the incumbent population will only decrease in P divided by the size of the population n. If n is infinite, then the effect of i's punishment on the incumbent population is just zero. This reasoning is important because it is behind the (correct) argument that the concept of ESS is a refinement of (symmetric 2-player games) Nash equilibrium. Without the assumption of infinite populations, the argument does not necessarily hold (see Galán and Izquierdo (2005) for an illustration). To avoid this issue without having to impose infinite populations, an alternative is to make sure that the smallest invasion barrier expressed as a population share exceeds $1/n$ (Weibull, 1995, pp. 33-34).

2.3.2. Evolutionary dynamics: the replicator dynamics

Naturally, to study the dynamics of an evolutionary system explicitly (i.e. beyond the analysis of its rest points), it becomes necessary to specify the particular process that governs such dynamics. The most extensively studied dynamic process in EGT is the replicator dynamics, proposed by Taylor and Jonker (1978). In the replicator dynamics (RD), payoffs are interpreted as the number of viable offspring that inherit the same behavioural phenotype (i.e. strategy) as their (single) parent. The theoretical model underpinning the basic RD also assumes a *single infinite* population of individuals who are repeatedly *randomly* drawn to play a *2-player symmetric* game. Furthermore, individuals can only play *one out of a finite set of pure strategies*, and *mutations (and random drift) are not allowed*[5]. This set of assumptions is enough to fully determine a *deterministic* dynamic process in which the rate of change in the frequency of any given strategy is equal to the relative difference between its average payoff and the average payoff obtained across all strategies in the population. Most often, time is treated as a continuous variable, and this allows the formalisation of the dynamic process as a system of ordinary differential equations.

With these assumptions in place, game theorists have been able to derive a chain of useful mathematical results that link the concept of ESS, the dynamics of the basic RD and the concept of NE. The logical chain is as follows: the population profile induced by an ESS is asymptotically stable in terms of the RD (Hofbauer et al., 1979); the mixed strategy corresponding to an asymptotically stable equilibrium of the RD is in (symmetric) perfect Nash equilibrium with itself (see proof in e.g. Weibull, 1995, section 3.4); and finally, a mixed strategy played at a symmetric Nash equilibrium (in a 2-player symmetric game with a finite set of pure strategies) induces a stationary population state of the RD (see proof in e.g. Vega-Redondo, 2003, pg. 367).

2.3.3. Further developments

While undoubtedly extremely useful, the assumptions embedded in the original concept of ESS and in the basic RD limit the applicability of the analytical results obtained with them, particularly in social (rather than biological) contexts (see e.g. Probst, 1999; Gotts et al., 2003b; Vega-Redondo, 2003, pg. 372). These concerns led to the development of more general frameworks which would encompass as particular cases not only the RD but also a wider range of dynamic processes, and could be applied not only to 2-player symmetric games, but also to general games. Of special interest are the *multi-population* models with *regular* and *payoff monotonic* dynamics.

[5] Mutations can be superimposed as a separate component of the dynamic process (see e.g. Imhof et al. 2005).

- *Multi-population* models study *n*-player games, where each player is randomly drawn from a distinct (infinite) population. This setting allows modelling any finite game in normal form where players in different positions are subjected to independent evolutionary pressures.
- *Regularity* ensures that the proportional rates of change of strategies are well defined and are continuously differentiable.
- Finally, *payoff monotonicity* is a mild condition which imposes that for any given pair of strategies in any particular population, their proportional rates of change are ordered in the same way as their respective average payoffs (Vega-Redondo, 2003, pg. 377).

It turns out that most of the analytical results linking the concepts of ESS, NE, and the dynamics of the basic RD can be carried over to this general framework (once the appropriate generalisations for these concepts have been defined; see e.g. Weibull (1995, chapter 5) and Vega-Redondo (2003, chapter 10)). This type of general framework[6] represents a remarkable step forward in generality and, consequently, the applicability of the analytical results obtained with them is greatly increased. However, these general models still make two assumptions that somewhat limit their applicability to social contexts (Probst, 1999): regularity and infinite populations.

As pointed out by Probst (1999), the assumption of regularity rules out many adaptation mechanisms that are considered of much interest in modelling social systems (e.g. best-response dynamics). This assumption, which is rarely made in learning game theory (LGT), is one of the main differences between EGT models and LGT models, in terms of the mathematical properties of the induced formal systems.

The assumption of infinite populations effectively averages out the stochasticity of the system, so the obtained deterministic dynamics can be formalised as a system of differential equations. This assumption has greater implications than one may initially suspect. As Traulsen et al. (2006) point out, “the finiteness of

[6] There are various similar versions (see Weibull, 1995).

populations may indeed lead to fundamental changes in the conventional picture emerging from deterministic replicator dynamics in infinite populations". To be more precise, any model with finite populations can be formalised as a Markov process, and the system of differential equations is the approximation of the Markov process in the limit as the population tends to infinity. Also, one is often interested in studying the behaviour of the system in the long run, which involves calculating the limit of the dynamics as time goes to infinity. The problem in doing this is that results can be dramatically different depending on the order in which one takes these two limits. This will be clearly illustrated in a somewhat different context in chapter 4. Fortunately, our theoretical knowledge of these issues has progressed immensely in the last few years. In particular, the seminal paper by Benaim and Weibull (2003) is a breakthrough in the field of stochastic approximation in EGT. In any case, it is clear that "care is therefore needed in the application of these approximations" (Beggs, 2002).

In summary, the study of the evolution of finite populations is significantly different from that of infinite populations (both in terms of the methods that are adequate for their analysis and on the results obtained with them); thus, it is not surprising that the analysis of finite evolutionary systems is nowadays a field of great scientific dynamism (see e.g. Nowak et al., 2004; Taylor et al., 2004; Imhof et al., 2005; Santos et al., 2006; Traulsen et al., 2006).

2.3.4. Stochastic finite systems

Once it has been acknowledged that stochasticity plays an important role in the analysis of finite evolutionary systems, the main challenge for current EGT seems to lie in understanding the impact of the various other assumptions made in traditional EGT on these finite stochastic systems.

A feature of the system that has been long known to play a crucial role is the mechanism by which individuals pair to play the game. The pairing algorithm does not necessarily have to be imposed by a fixed population structure, but may be actively conducted by the players themselves (Eshel and Cavalli-Sforza, 1982). Naturally, the impact of the standard assumption (random encounters) is investigated by considering other pairing mechanisms. One of the first studies to show the relevance of different population structures in finite systems was conducted by Nowak and May (1992; 1993), who used a spatial model (where local interactions occurred between individuals occupying neighbouring nodes on a square lattice) to show that stable population states for the prisoner's dilemma depend upon the specific form of the payoff matrix. For a review of several studies in the context of social dilemmas that consider populations where some pairs of agents are more likely to interact than others see Gotts et al. (2003b). Of particular interest is the field of study on tags (Holland, 1993). Tags are arbitrary social marks that, in principle, are not linked to any particular form of behaviour, but they do influence the way individuals interact: individuals with similar tags have a preference to interact with each other (see e.g. Riolo, 1997; Hales, 2000; Riolo et al., 2001; Edmonds and Hales, 2003). In chapter 6 we investigate various pairing mechanisms and, in particular, we analyse one which is formally equivalent to the use of tags. For a recent illustration of the latest developments in the field of structured populations in finite systems, see Santos et al. (2006), who study social dilemma games played in (fixed) networks with various degrees of heterogeneity in the degree distributions. The most recent literature in this field is mainly focused on studying the emergence of cooperation in *spatially* structured populations (see e.g. Hauert and Doebeli, 2004; Doebeli and Hauert, 2005; Németh and Takács, 2007). For a recent illustration of the fact that allowing players to selectively choose their partners can have dramatic effects on the emergence of cooperation in finite systems see e.g. Joyce et al. (2006).

In chapter 6 we also investigate various selection mechanisms (i.e. algorithms that determine how the population composition varies as a function of the payoffs obtained by each individual). This is another area of research where a substantial amount of work has been conducted in the last few years. In a recent paper, Traulsen et al. (2006) develop a framework within which one can explore various intensities of selection, i.e. different ways in which payoffs relate to fitness (where fitness is the function that determines the potential to reproduce). This selection framework makes use of the Fermi distribution function from statistical mechanics to control the balance between selection and random drift in finite populations. Using this function, Traulsen et al. (2006) explore different intensities of selection –ranging from neutral, random drift, up to the extreme limit of cultural imitation dynamics– in the three 2-player 2-strategy social dilemma games (these are explained in detail in section 3.1). Traulsen et al. (2006) are able to calculate the fixation probabilities of different strategies, and they also use stochastic approximation theory to relate their results on finite populations to those obtained with infinite populations.

An assumption that –to our knowledge– has not been investigated in depth in evolutionary stochastic finite systems is the one relating to the properties of the set of strategies that players are allowed to select. In chapter 6 of this thesis we show that this assumption may have wider implications than one may initially suspect.

There are many other ways in which several authors have addressed some of the limitations of EGT outlined above. Unfortunately (but probably inevitably), the study of the implications of various assumptions made in mainstream EGT is being undertaken in a somewhat disorganised fashion. This inconvenience is probably a consequence of the dynamism of this field, and it will hopefully be corrected in time through the creation of general frameworks that facilitate rigorous and transparent comparisons between different models and the results obtained with them. Chapter 6 of this thesis is meant to be a step in this direction, by providing a single coherent framework within which results obtained from different stochastic finite models can be contrasted and compared.

2.4. Learning game theory

Like evolutionary game theory, learning game theory (LGT) abandons the demanding assumptions of classical game theory on players' rationality and beliefs. However, unlike evolutionary game theory –where players are often assumed to be pre-programmed to play a fixed strategy–, LGT assumes that players are able to learn over time about the game and the behaviour of others (through e.g. reinforcement, imitation or belief updating), and this learning process is *explicitly* modelled (Vega-Redondo, 2003, pg. 398). This distinction means that the level at which dynamic processes are defined in EGT and LGT is fundamentally different (Fudenberg and Levine, 1998). Models in EGT are aggregate in the sense that they describe the aggregate behaviour of a population of players through various generations; the population is subject to evolutionary pressures (and therefore the *population* adapts), but the individual components of the population have a predefined fixed behaviour. On the other hand, models in LGT comprise players who *individually* adapt through learning, and it is this learning process that is formally described. Models in LGT explicitly represent the learning processes that each individual player carries out, and the dynamics that are generated at the aggregate level (which are most often stochastic and non-regular) emerge out of the strategic interactions among the players.

Another fundamental difference between LGT and EGT relates to the relationship between the number of players in the game and the number of players in the population. Models in LGT tend to focus on one very small population of *n* players (most often *n* = 2), who play an *n*-player game (all individuals in the population play the game at once). This is in stark contrast with EGT models, where individuals within a large (usually infinite) population are drawn to play a 2-player game. As explained in section 2.3.1, this distinction can have very important implications.

Despite these differences, theoretical work linking results from EGT and LGT seems to indicate that we may be close to a point where the integration of the two approaches is within reach (Weibull, 1998). This is a question that is further discussed in section 7.4.

Interestingly, there seem to be two fundamentally different motivations to study learning models in the LGT literature. One is mainly concerned with identifying learning algorithms that will lead to NE or, ideally, to refinements of NE. The following quote by Vega-Redondo nicely summarises this motivation: "In particular, our concern is to identify different classes of games in which the corresponding learning processes bring about long-run convergence to some Nash equilibrium. As we shall see, many of the proposed models *fare reasonably well* for certain games but induce quite *unsatisfactory performance* for some others." [our emphasis] (Vega-Redondo, 2003, pg. 398).

This thesis follows another motivation: we are mainly concerned with identifying the strategic implications of decision-making algorithms that have received support from cognitive science research. Work following this second rationale has sometimes been labelled “cognitive game theory” (CogGT) in the literature (e.g. Flache and Macy, 2002). Nowadays, an increasing number of researches use CogGT to investigate animal –often human– behaviour in strategic contexts using models that seem more plausible than those deriving from classical game theory. Thus, CogGT models are often used to identify learning mechanisms that will lead to patterns of behaviour observed in real-world interactions (and these patterns often do not correspond to NE). The following summarises some features that characterise the way players are modelled in CogGT (Flache and Macy, 2002; Macy and Flache, 2002), in contrast with classical game theory:

- Players base their decisions on experience of past events as opposed to logical deductions about the future. This inductive approach requires fewer assumptions about other players and may be more adequate to model animal (including human) behaviour. Since inferences about other players' strategies –or about future payoffs– is made in the light of the history of the game, they can only lead to probable −rather than necessarily true− conclusions (even if the evidence used is accurate).
- Players have feedback on their actions; otherwise learning cannot occur. Learning takes many forms, depending on the available feedback, the available knowledge, and the way these are used to modify behaviour.
- The fact that players learn from experience means that they often cannot undertake an optimal behaviour (since inferences about other players' behaviour cannot be guaranteed to be true). An optimal approach requires knowledge that sometimes has to be inferred from experience. In the process of acquiring the necessary knowledge, suboptimal behaviour can occur as a result of exploring different actions or having drawn imperfect conclusions from experience. When modelling players who learn from experience, it often seems reasonable to assume that they satisfice rather than optimise. The concept of ‘satisficing’ was introduced by Simon (1957) to indicate that agents often seek for a solution to a problem until they have found one which is ‘good enough’, rather than persisting in the hope of finding an optimal solution (which could be nonexistent,

incalculable, or unidentifiable). The 'good enough' solution is usually defined by setting a certain aspiration threshold.

The distinction between the two different motivations outlined above becomes clear when one considers social dilemmas. In most single-stage social dilemma games, the cooperative strategy is dominated (i.e. it cannot lead to NE); however empirical studies have generally found that, while it is not easy to establish cooperation, levels of cooperation tend to be higher than would be expected if the assumptions made in CGT held true. Thus, when studying social dilemmas, researchers in LGT following the "NE motivation" would presumably consider models leading to cooperative solutions generally unsatisfactory. In stark contrast, in the context of social dilemmas, CogGT has been mainly concerned with identifying a set of model-independent learning principles that are necessary and sufficient to generate cooperative solutions (Flache and Macy, 2002). Interestingly –if unsurprisingly–, it seems that researchers more inclined towards CogGT tend to use computer simulation (instead of mathematical analysis) relatively more than those researchers following the "NE motivation".

2.4.1. Different learning algorithms

As mentioned above, the process of learning can take many different forms, depending on the available knowledge, the available feedback, and the way these are used to modify behaviour. The assumptions made in these regards give rise to different models of learning. In most models of LGT, players use the history of the game to decide what action to take. In the simplest models (e.g. reinforcement learning) this link between acquired information and action is direct (e.g. in a stimulus-response fashion); in more sophisticated models players use the history of the game to form expectations about the other players' behaviour, and they then react optimally to these inferred expectations. Following Vega-Redondo (2003, chapter 11) we briefly present here some of the most studied learning models in ascending order of sophistication, according to the amount of information that players use and their computational capabilities.

Reinforcement learning

Reinforcement learning models will be discussed at length in section 4.1. Let us say for now that they are arguably the simplest family of learning algorithms investigated in LGT. Reinforcement learning is also one of the most widespread adaptation mechanisms in nature. Reinforcement learners use their experience to choose or avoid certain actions based on their immediate consequences. Actions that led to satisfactory outcomes (i.e. outcomes that met or exceeded aspirations) in the past tend to be repeated in the future, whereas choices that led to unsatisfactory experiences are avoided. In general, reinforcement learners do not use more information than the immediately received payoff, which is used to adjust the probability of the conducted action accordingly. The specific details of how this general principle is implemented in different models can lead to substantially different dynamics, as explained in section 4.1.

Static perceptions; better and best (myopic) response

In this more sophisticated family of learning models, each player is assumed to know not only the payoff she receives in each possible outcome of the game, but also the actions that every player selected at a certain time t. When making her decision for time (t + 1) every player assumes that every other player will keep her strategy unchanged (i.e. static perception of the environment); then, each individual player, working under such assumption and knowing the payoff structure of the game in what pertains to her own payoff, can identify the set of strategies that will lead to an improvement in her current payoff (if possible). In better-response models, one of these payoff-improving strategies is selected at random; in best-response models, only those strategies that give the highest payoff given the prevailing assumptions are considered for selection. In these models players assume that their environment is static and deterministic, and respond to it in a myopic fashion, i.e. ignoring the implications of current choices on future choices and payoffs. Vega-Redondo (Vega-Redondo, 2003, pp. 415-420) summarises several results for this type of learning algorithm.

Fictitious play

Fictitious play models were first proposed by Brown (1951). Fudenberg and Levine (1998) provide a recent and comprehensive account of this family of models. As in best (myopic) response models, players in fictitious play (FP) models are assumed to have a certain model of the situation and decide optimally on the basis of it. The higher level of sophistication introduced in FP models concerns the (still stationary) model of the environment that players hold. FP players assume that the mixed strategy played by every other player at a certain time is equal to the frequency with which they have selected each of their available actions up until that moment. Thus, instead of considering the actions taken by every other player only in the immediately preceding time-step (as in the models explained in the previous section), they implicitly take into account the full history of the game. After forming her beliefs about every other player's strategy, a FP player (myopically) responds optimally to them.

In 2-player games, the belief sequence induced by FP is known to converge to a profile that defines a Nash equilibrium. This result, however, may be somewhat misleading, as it does not imply that players will play the strategy profile induced by such a sequence of beliefs in an *uncorrelated* fashion (Fudenberg and Kreps, 1993), randomising their decisions *independently* from each other as the definition of a Nash equilibrium requires. As an example, imagine that the belief sequence in a 2x2 game converges to a strategy profile (i.e. an assignment of frequencies to all the strategies available to a player) where fictitious player 1 selects action A_1 with frequency 1/3 (and action B_1 with frequency 2/3) and fictitious player 2 selects action A_2 with frequency 1/3 (and action B_2 with frequency 2/3). The mathematical result mentioned above guarantees that there is a Nash equilibrium with the strategy profile FP converges on. This would seem to suggest that the pattern of play in fictitious play will be the same as the pattern of play induced by a Nash equilibrium, but this is not necessarily the case. Thus, in our example, the Nash equilibrium in mixed strategies would imply that any outcome has a positive probability of occurring (e.g. outcome [A_1, B_2] would occur with probability 2/9). On the contrary, by setting players' initial beliefs appropriately (which are determined by numerical weights, one for each of the other player's pure strategies) one can construct examples where player 1 selects action A_1 if and only if player 2 selects action A_2 (Fudenberg and Kreps, 1993). This, in particular, would imply that outcome [A_1, B_2] would never occur. Thus, the payoff obtained by each player in this latter case can be completely different from the expected payoff obtained if players selected action A_i or B_i in an uncorrelated fashion. Therefore, each component of the belief sequence in FP must be understood as a *marginal* distribution for each player separately; the *joint* distribution may be very different from that resulting from Nash equilibrium play.

Smooth fictitious play

The perverse correlation effects outlined in the previous section motivated a stochastic version of the original fictitious play named *smooth* fictitious play (SFP, Fudenberg and Kreps, 1993). As in the original fictitious play, players in SFP assume that the mixed strategy played by every other player at a certain time is equal to the frequency with which they have selected each of their available actions up until that moment. In SFP models, however, players are no longer assumed to respond to their beliefs about the other players' strategies in the knife-edge fashion implied by the best-response correspondence; instead they respond in a continuous, differentiable way. The step-like determinism of the best-response correspondence used in FP is replaced by a smooth-looking function that returns a probabilistic response to the other players' inferred strategies in SFP. In SFP (as in FP), the rate of adjustment of behaviour slows down at a rate that permits the use of stochastic approximation theory, and this has facilitated the derivation of several theoretical results. In particular, SFP players' strategies are guaranteed to converge to Nash equilibrium in 2x2 games (Fudenberg and Levine, 1998).

Rational learning

The most sophisticated model of learning in LGT was proposed by Kalai and Lehrer (1993a; 1993b). Players in this model are assumed to be fully aware of the strategic context they are embedded in. They are also assumed to have a set of subjective beliefs over the behavioural strategies of the other players. Informally, as put by Vega-Redondo (2003, pg. 434), the only assumption made about such beliefs is that players cannot be "utterly surprised" by the course of the play, i.e. players must assign a strictly positive probability to any belief that is coherent with the history of the game. Finally, players are assumed to respond optimally to their beliefs with the objective of maximising the flow of future payoffs discounted at a certain rate. A detailed explanation of the (very powerful) results obtained with this model seems to fall out of the scope of this brief account of learning models. We refer the interested reader to Vega-Redondo (2003, pp. 433-441), who provides a brilliant account of this part of the literature, and concludes that "some of the assumptions underlying the rational-learning literature […] should be interpreted with great care".

Let us conclude this section by pointing out a common weakness of most current models in LGT (including those developed in this thesis): they almost invariably assume that every player in the game follows the same decision-making algorithm. This seems to be the natural first step in exploring the implications of a decision-making algorithm; however, it is clear that in many of these models the observed dynamics are very dependent on the fact that the game is played among "cognitive clones". Confronting the investigated learning algorithm with other decision-making algorithms seems to be a promising second step in LGT studies.

2.4.2. Assumptions in the learning models developed in this thesis

Reinforcement learning

Chapter 4 is an in-depth analysis of the transient and asymptotic dynamics of the Bush-Mosteller reinforcement learning algorithm for 2-player 2-strategy games. The following summarises the main assumptions made in this model in terms of the nature of the payoffs, the information players require and the computational capabilities that they have.

- Payoffs: In this model, payoffs and aspiration thresholds are not interpreted as von Neumann-Morgenstern utilities (for which the distinction between positive and negative values is irrelevant), but as a set of variables measured on an interval scale that is used to calculate stimuli (this is explained in detail in section 4.2).
- Information: Each player is assumed to know the range of possible actions available to her, and the maximum absolute difference between any payoff she might receive and her aspiration threshold. Players do not use any information regarding the other players.

- Memory and computational capabilities: Players are assumed to know their own (potentially) mixed strategy at any given time. They need to be able to conduct arithmetic operations.

Case-based reasoning

Chapter 5 is an exploration of cased-based reasoning as a decision-making algorithm in strategic contexts. The following summarises the main assumptions made in this model in terms of the nature of the payoffs, the information players require and the computational capabilities that they have.

- Payoffs: In this model, payoffs can be interpreted as preferences measured on an ordinal scale.
- Information: Each player is assumed to know the range of possible actions available to her, and her own aspiration threshold. Players do not use any information regarding the other players.
- Memory and computational capabilities: For each possible state of the world they may perceive, players are assumed to store in memory the last payoff they received for each of the possible actions available to them. They need to be able to rank their preferences.

2.5. Non-strictly-deductive branches of game theory

This thesis aims to be an advancement in the field of *deductive* game theory. It is important to note that there are other branches of game theory which are not purely deductive; these non-strictly-deductive branches tend to use game theory as a framework to fit observed empirical data and understand the underlying mechanisms that may be producing the observed results. There is clearly a lot to gain from the interaction of deductive and non-deductive game theory. Traditionally, deductive game theory has developed almost entirely from introspection and theoretical concerns. Unless this is corrected in the coming years, deductive game theory may suffer the danger of becoming practically irrelevant or, in less dramatic terms, not fulfilling all its potential as a useful tool to analyse real-world social interactions. On the other hand, if the objective is to find a model that fits empirical data to a satisfactory extent, it is crucial to understand the behaviour of different models in detail; if one is not content with fitting only, but some level of understanding is also pursued, then it becomes fundamental to know the implications of various cognitive mechanisms (i.e. assumptions) for the development of the game. Thus, it seems very clear that empirical studies have also a lot to gain from theoretical analyses. These issues will be discussed in chapter 7, but let us say for now that the work reported in this thesis has tried to be relevant by (a) studying the strategic implications of decision-making algorithms that have received empirical support from the cognitive sciences and (b) building frameworks to clearly identify the factors (i.e. types of assumption) that may have the greatest impact in the outcome of a social interaction (i.e. a game).

There are a number of learning models that have been proposed to explain experimental data (see chapter 6 in Camerer, 2003), and many of them have been investigated in purely theoretical terms. The transition from theoretical learning models to non-strictly deductive branches of game theory is very smooth. Here we mention two: psychological game theory and behavioural game theory. Psychological game theory is a term coined by Colman (2003).

> "Psychological game theory […] overlaps behavioral game theory but focuses specifically on non-standard reasoning processes rather than other revisions of orthodox game theory such as payoff transformations. Psychological game theory seeks to modify the orthodox theory by introducing formal principles of reasoning that may help to explain empirical observations and widely shared intuitions that are left unexplained by the othodox theory" (Colman, 2003).

Overlapping psychological game theory, behavioural game theory is completely driven by empirical (especially experimental) data, and models are assessed according to how well they are fitted to data. While models in cognitive game theory are designed to help us *reflect on* a certain process, behavioural game theory builds on models which are usually designed to *represent* the actual process.

> "Behavioral game theory is about what players *actually* do. It expands analytical theory by adding emotion, mistakes, limited foresight, doubts

> about how smart others are, and learning to analytical game theory. Behavioral game theory is one branch of behavioral economics, an approach to economics which uses psychological regularity to suggest ways to weaken rationality assumptions and extend theory." (Camerer, 2003, p.3)

Let us finish the chapter by stating that learning models have been reported to outperform classical game-theoretic predictions on experimental data (see Macy, 1995; Roth and Erev, 1995; Erev and Roth, 1998; Camerer, 2003, chapter 6). The empirical support of learning models in game theory will be expanded for reinforcement learning and case-based reasoning in the following chapters.

3. Scope and Method

This thesis provides some general results for *n*-player games; however, most of the research has been focused on 2-player 2-strategy (2x2) games. In several cases, it has been convenient to illustrate the obtained findings using specific types of 2x2 games, and for that purpose I have often selected 2x2 social dilemma games[7]. The first section of this chapter explains what social dilemmas are and how they can be formalised as 2x2 games; it also gives a brief account of some of the most relevant results obtained within each of the main branches of deductive game theory on the most famous 2x2 social dilemma, i.e. the Prisoner's Dilemma, and of how these results relate to empirical findings. The second section of this chapter outlines the range of formal methods that have been used to analyse the models developed in this thesis.

3.1. Social dilemmas

Social dilemmas are social interactions where everyone enjoys the benefits of collective action, but any individual would gain even more without contributing to the common good (provided that the others do not follow her defection). Social dilemmas are by no means exclusive to human interactions: in many social contexts, regardless of the nature of their component units, we find that individual interests lead to collectively undesirable outcomes for which there is a feasible alternative where every individual would be better off. The problem of how to promote cooperation in these situations without having to resort to central authority has been fascinating scientists from a broad range of disciplines for decades. The value of understanding such a question is clear: in the social and biological sciences, the emergence of cooperation is at the heart of subjects as diverse as the first appearance of life, the ecological functioning of countless environmental interactions, the efficient use of natural resources, the development of modern societies, and the sustainable stewardship of our planet. From an engineering point of view, the problem of understanding how cooperation can emerge and be promoted is crucial for the design of efficient decentralized systems where collective action can lead to a common benefit despite the fact that

[7] In chapter 5 I also investigate an *n*-player social dilemma.

individual units may (purposely or not) undermine the collective good for their own advantage.

At the most elementary level, social dilemmas can be formalised as two-person games where each player can either cooperate or defect. For each player *i*, the payoff when they both cooperate (R_i, for *Reward*) is greater than the payoff obtained when they both defect (P_i, for *Punishment*); when one cooperates and the other defects, the cooperator obtains S_i (*Sucker*), whereas the defector receives T_i (*Temptation*). Assuming no two payoffs are equal, the essence of a social dilemma is captured by the fact that both players prefer any outcome in which the opponent cooperates to any outcome in which the opponent defects ($\min(T_i, R_i) > \max(P_i, S_i)$), but they both can find reasons to defect. In particular, the temptation to cheat (if $T_i > R_i$) or the fear of being cheated (if $S_i < P_i$) can put cooperation at risk. There are three well-known social dilemma games: Chicken, Stag Hunt, and the Prisoner's Dilemma. In Chicken the problem is greed but not fear ($T_i > R_i > S_i > P_i$; $i = 1, 2$); in Stag Hunt, the problem is fear but not greed ($R_i > T_i > P_i > S_i$; $i = 1, 2$); and finally, both problems coincide in the paradigmatic Prisoner's Dilemma ($T_i > R_i > P_i > S_i$; $i = 1, 2$).

Social dilemmas have been studied from different perspectives, including empirical approaches (both experimental and field studies), discursive theoretical work, game theory, and computer simulation. Within the domains of these four approaches much of the work has been devoted to the study of the Prisoner's Dilemma (PD) or variations of it, often leading to conflicting conclusions (particularly relevant is the conflict between empirical work and classical game theory).

The most widespread results about the PD come from classical game theory. 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). The situation is very different when the game is played repeatedly. In the (finite or infinitely) repeated game, the range of possible strategies and outcomes is much wider and defecting in every round is no longer a dominant strategy. In fact, in the repeated PD, there is not necessarily one best strategy irrespective of the opponent's strategy. As an example, Kreps et al. (1982) showed that a cooperative outcome can be sustained in the finitely repeated PD if a rational player believes that there is at least a small probability that the other player is playing "Tit for Tat" (TFT)[8].

Since assuming players are instrumentally rational is not enough to narrow the set of solutions of the repeated PD sufficiently, common knowledge of rationality is brought into play. Assuming common knowledge of rationality it can be proved using backwards induction that a series of bilateral defections is the only possible outcome of the finitely repeated PD (Luce and Raiffa, 1957)[9]. 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, a very wide range of possible outcomes where the two players are responding optimally to each other's strategy still exists, even when assuming that the two players have detailed pre-planned strategies and these are common knowledge. Specifically, the "Folk Theorem" states that any individually-rational outcome[10] can be a Nash equilibrium in the infinitely-repeated PD if the discount rate of future payoffs is sufficiently close to one. In this case, orthodox game theory has little to say about the dynamics leading a set of players to one among many possible equilibria.

When classical game theoretical solutions of the PD and related games have been empirically tested, disparate anomalies have been found (see, for example, work reviewed by Colman (1995) in chapters 7 and 9, Roth (1995), Ledyard (1995), and Camerer (2003)). Generally, empirical studies have found that there is a wide variety of factors in addition to economic payoffs that affect our behaviour, and also that, while it is not easy to establish cooperation, levels of cooperation tend to

[8] This is the strategy consisting of starting by cooperating, and thereafter doing what the other player did on the previous move.

[9] For a detailed analysis of the finitely repeated Prisoner's Dilemma, see Raub (1988).

[10] An outcome giving each player at least the largest payoff that they can guarantee receiving regardless of the opponents' moves.

be higher than those predicted by classical game theory (see e.g. Dawes and Thaler, 1988). The explanation of the clash between classical game theory and empirical evidence is, of course, that the assumptions required to undertake a game theoretical analysis do not hold: economic payoffs do not readily correspond to preferences (e.g. considerations of fairness frequently influence behaviour); actual preferences are sometimes neither consistent nor static nor context-independent; players' cognitive capabilities are indeed limited, and players' assumptions of others' preferences and rationality assumed by game theory are therefore often wrong.

Research on the PD within evolutionary game theory was boosted by the computer simulations and empirical studies undertaken by Axelrod (1984). Axelrod's work represents a key event in the history of research on the PD. By inviting entries to two repeated PD computer tournaments, Axelrod studied the success of different strategies when pitted against themselves, all the others, and the random strategy. The strategy TFT won both tournaments and an extension of the second one. The extension, called ecological analysis, consisted of calculating the results of successive hypothetical tournaments, in each of which the initial proportion of the population using a strategy was determined by its success in the preceding tournament. Axelrod explains that TFT's success is due to four properties: TFT is nice (it starts by cooperating), provocable (it retaliates if its opponent defects), forgiving (it returns to play cooperatively if the opponent does so), and clear (it is easy for potentially exploitative strategies to understand that TFT is not exploitable). TFT's success is even more striking when one realises that it can never get a higher payoff than its opponent. Though severely criticised by some game theorists for drawing excessively on computer simulation and being partially flawed, Axelrod's work is widely accepted to have greatly stimulated analytical work within the domain of evolutionary game theory and further research on the PD using computer simulation. Findings on the repeated PD from evolutionary game theory are summarised by Bendor and Swistak (1995; 1998) and Gotts et al. (2003b); in particular, Gotts et al. (2003b) conclude that the assumptions about the dynamics of competition between strategies in mainstream EGT make the analytical results much less plausible as good approximations in social than in biological contexts. Gotts et al. (2003b) have also extensively reviewed work on social dilemmas using computer simulation.

As explained in the previous chapter, there are many different models in the branch of learning game theory, and their predictions for social dilemma games are far from uniform. In very general terms, models that have been designed to converge to Nash equilibria predict uncooperative solutions (see e.g. Fudenberg and Levine, 1998), while models including players who satisfice predict cooperative outcomes for certain ranges of aspiration thresholds (e.g. Karandikar et al., 1998; Bendor et al., 2001a, 2001b). There are also learning models where players do not satisfice and which lead to cooperative solutions; an interesting example is given by Erev and Roth (2001). Erev and Roth (2001) point out that the performance of reinforcement learning models in explaining human behaviour in games that facilitate reciprocation (i.e. games where players can coordinate and benefit from mutual cooperation, like the Prisoner's Dilemma) had traditionally been remarkably less successful than in explaining other types of games (e.g. zero-sum games and games with unique mixed strategy equilibria, see McAllister, 1991; Mookherjee and Sopher, 1994; Roth and Erev, 1995; Mookherjee and Sopher, 1997; Chen and Tang, 1998; Erev and Roth, 1998; Erev et al., 1999). As mentioned above, many people do learn to cooperate in the repeated Prisoner's Dilemma, whilst most simple models of reinforcement learning used in experimental game theory predicted uncooperative outcomes. Interestingly, Erev and Roth (2001) show that such a result does not reflect a limitation of the reinforcement learning approach but derives from the fact that previous models used to fit experimental data assumed that players can only learn over immediate actions (i.e. stage-game strategies) but not over a strategy set including repeated-game strategies (like e.g. tit-for-tat).

3.2. Method

In the following chapters we characterise the dynamics of various stochastic systems using a range of different techniques. The typical system investigated in this thesis contains a (potentially variable) finite number of players who interact to get certain payoffs, and are subject to stochasticity (either in their individual behaviour or in the dynamics of the population they belong to). In these systems, each of the players can adapt its behaviour (i.e. learn), or the population of players as a whole adapts through an evolutionary process. The payoff obtained by each of these players depends on the actions undertaken by other players; this feature is what makes game theory a useful framework to study the system.

This thesis makes extensive use of two distinct approaches to analyse the dynamics of these systems: computer simulation and mathematical analysis. As in Gotts et al. (2003a), it will be shown by example that mathematical analysis and simulation studies should not be regarded as alternative and even opposed approaches to the formal study of social systems, but as complementary. They are both extremely useful tools to analyse formal models, and they are complementary in the sense that they can provide fundamentally different insights on the same model (and also on one same question using different models, as argued by Gotts (2003b)). Chapter 4 will clearly illustrate the fact that the level of understanding gained by using these two techniques together could not be obtained using either of them on their own. Furthermore, each technique can produce both problems and hints for solutions for the other. The following explains how these two techniques have been used in this thesis.

3.2.1. Computer simulation

Simulations can usually provide an explicit and fully accurate representation of the original system and its stochastic dynamics. In this way, simulations allow us to explore the properties of formal models that are intractable using mathematical analysis, and they can also provide fundamentally new insights even when such analyses are possible.

The specific modelling technique used in this thesis is called agent-based modelling (ABM). ABM is a modelling paradigm with the defining characteristic that entities within the target system to be modelled –and the interactions between them– are explicitly and individually represented in the model (Edmonds, 2000). Because of this, ABM is especially appropriate to simulate game theoretical models, where the description of the system in terms of the behavioural and adaptive rules of the individual players is usually very simple. Clearly, running a stochastic agent-based model in a computer provides a formal proof that a particular micro-specification is *sufficient* to generate the pattern of behaviour that is observed during the simulation. However, one is usually interested not only in how the system *can* behave, but also in determining how the system behaves *in general*, which involves finding the probability distribution of different patterns. For this, it becomes necessary to run a large number of simulations with different random seeds and appropriately chosen initial conditions (see e.g. section 6.5.1). Most often, simulations cannot provide general closed-form results about how the system behaves, or about how it responds to changes in the parameter space. Thus, there is great value in complementing simulation with mathematical analysis.

In the work reported in this thesis simulation is often used as a starting point. There are two reasons for this. First, the very nature of the systems analysed here (see beginning of section 3.2) means that they can be easily described (and implemented) within an ABM framework. Secondly, the models developed here have not been designed to be mathematically tractable, but to study phenomena that we considered particularly interesting; thus, at least at first, they often seem to be mathematically intractable. Mathematical work is then used to analyse the patterns observed in the initial simulations, and this analysis sometimes leads to the production of simpler models that exhibit similar behavioural patterns and which are amenable to more detailed mathematical analysis. An example of this interaction between simulation and mathematical analysis is the development of deterministic approximations (i.e. simpler models) of the stochastic dynamics of a more complex system (e.g. see chapter 4). Simulation and mathematical analyses are therefore used complementarily: with simulation allowing us to explore intractable models, to extract the key features of such models, and to build new simpler models that still keep such key features; and mathematical work illuminating the behaviour of the initial models, and providing in-depth analyses of the simpler models. In many cases simulations have also suggested promising ways of pursuing new theoretical results.

As mentioned in the introduction, a great effort has been made in this thesis to make sure that every computational experiment conducted here can be easily inspected, rerun, scrutinised, reimplemented, and modified by independent researchers. Given the amount of care put on this task, I place as much confidence on the results obtained using computer simulation as I do on the mathematical derivations.

3.2.2. Mathematical analysis

The original systems investigated in this thesis can all be meaningfully formalised as Markov processes. However, the (sometimes infinite) number of states and the nature of the transitions between different states often mean that traditional Markov analysis cannot be readily applied. In the presence of these difficulties, there are two approaches that have been followed to characterise these systems using mathematical analysis: (a) partial analysis of the original Markov process, and (b) in-depth analysis of an approximation to the original Markov process.

The partial analysis often starts by finding out whether the Markov process is ergodic. If the process is ergodic, this means that the probability of finding the system in each of its states in the long run is unique (i.e. initial conditions are immaterial). This probability is also the long-run fraction of the time that the system spends in each of its states. Although calculating such probabilities may be unfeasible, one can always estimate them using computer simulation (see e.g. section 6.5.1). If the process is not ergodic, one can try to identify its various transient and recurring classes (see e.g. sections 4.7 and 5.4). This task may involve using very specific techniques which may be adequate only for certain types of Markov processes. A particular feature of Markov processes that often determines which techniques may be most appropriate for their analysis is how (if at all) the speed of change (e.g. the rate of learning) itself varies with time. As an example, it will be shown in chapter 4 that when the magnitude of change remains constant in time (e.g. in models where learning does not fade away in time), results from the theory of distance diminishing models (Norman, 1968, 1972) can be particularly useful. Another useful analysis that can be conducted on non-ergodic Markov chains with various absorbing states consists in identifying which of these absorbing states are robust to small perturbations (Foster and Young, 1990; Young, 1993; Ellison, 2000). This sort of analysis has been conducted in sections 4.8 and 5.7.3.

A complementary approach to the partial analysis of the original Markov process consists in studying a simpler approximation to it. In this thesis I have made extensive use of mean-field approximations. The use of mean-field (or expected-motion) approximations to understand the dynamics of complex stochastic models is common in the game theoretical literature (see e.g. Fudenberg and Levine, 1998; Vega-Redondo, 2003). Note, however, that these are approximations whose validity may be constrained to specific conditions. As a matter of fact, there is a whole field in mathematics, namely stochastic approximation theory (Benveniste et al., 1990; Kushner and Yin, 1997), devoted to analysing under what conditions the *expected* and the *actual* motion of a system should become arbitrarily close in the long run. This is generally true for processes whose motion slows down at an appropriate rate (as explained by e.g. Hopkins and Posch (2005) when studying the Erev-Roth reinforcement model), but not necessarily so in other cases.

In any case, mean-field approximations can be very useful even when it is known that they cannot be used to characterise the dynamics of the system in the long-run. As an example, in chapter 4 we use the expected motion of the system to get insights about what areas of the state space may be particularly stable (or unstable), to identify their basins of attraction, to clarify the crucial assumptions of the model, to assess its sensitivity to various parameters, and to characterise and graphically illustrate the *transient* dynamics of the model. We also show that the expected-motion approximation, while valid over bounded time intervals, deteriorates as the time horizon increases. In fact, the approximation becomes very misleading when studying the *asymptotic* behaviour of the model.

It is also worth mentioning that mean-field approximations are often used in the literature not only to average stochasticity out, but also to average out heterogeneity among players (e.g. see the studies conducted by Galán and Izquierdo (2005), Edwards et al. (2003), Castellano, Marsili, and Vespignani (2000), and Huet et al (2007)). Such approximations provide simpler, more abstract models which are often amenable to mathematical analysis and graphical representation. However, as pointed out above, even though they are usually useful, one should not forget that the insights provided by these mathematical abstractions could be misleading.

To conclude, let us mention that a range of other mathematical techniques (e.g. Brouwer’s fixed-point theorem in section 4.9, and graph theory in section 5.7.3) have been used to analyse various properties of the models developed in this thesis.

4. Dynamics of the Bush-Mosteller Reinforcement Learning Algorithm in 2x2 Games[♣]

4.1. Introduction

Reinforcement learners interact with their environment and use their experience to choose or avoid certain actions based on the observed consequences. Actions that led to satisfactory outcomes (i.e. outcomes that met or exceeded aspirations) in the past tend to be repeated in the future, whereas choices that led to unsatisfactory experiences are avoided. The empirical study of reinforcement learning dates back to Thorndike's animal experiments on instrumental learning at the end of the 19th century (Thorndike, 1898). The results of these experiments were formalised in the well known 'Law of Effect', which is nowadays one of the most robust properties of learning in the experimental psychology literature:

> *Of several responses made to the same situation those which are accompanied or closely followed by satisfaction to the animal will, other things being equal, be more firmly connected with the situation, so that, when it recurs, they will be more likely to recur; those which are accompanied or closely followed by discomfort to the animal will, other things being equal, have their connections to the situation weakened, so that, when it recurs, they will be less likely to occur. The greater the satisfaction or discomfort, the greater the strengthening or weakening of the bond.*
>
> (Thorndike, 1911, p. 244)

Nowadays there is little doubt that reinforcement learning is an important aspect of much learning in most animal species, including many phylogenetically very distant from vertebrates (e.g. earthworms (Maier and Schneirla, 1964) and fruit flies (Wustmann et al., 1996)).

In strategic contexts, empirical evidence for reinforcement learning is strongest in animals with limited reasoning abilities or in human subjects who have no

[♣] Some parts of the material presented in this chapter are in press in Izquierdo, L.R., Izquierdo, S.S., Gotts, N.M. and Polhill, J.G. (2007), "Transient and asymptotic dynamics of reinforcement learning in games", *Games and Economic Behavior* , and others have been accepted for publication in the *Journal of Artificial Societies and Social Simulation*.

information beyond the payoff they receive and specifically may be unaware of the strategic nature of the situation (Mookherjee and Sopher, 1994; Roth and Erev, 1995; Bendor et al., 2001a; Camerer, 2003; Duffy, 2006). In the context of experimental game theory with human subjects, several authors have used simple models of reinforcement learning successfully to explain and predict behaviour in a wide range of games (McAllister, 1991; Mookherjee and Sopher, 1994; Roth and Erev, 1995; Mookherjee and Sopher, 1997; Chen and Tang, 1998; Erev and Roth, 1998; Erev et al., 1999; Erev and Roth, 2001). Reinforcement models in the literature tend to differ in the following, somewhat interrelated, features:

- Whether learning slows down or not, *i.e.* whether the model accounts for the 'Power Law of Practice' (e.g. Erev and Roth (1998) vs. Börgers and Sarin (1997)).
- Whether the model allows for avoidance behaviour in addition to approach behaviour (e.g. Bendor et al. (2001b) vs. Erev and Roth (1998)). Approach behaviour is the tendency to repeat the associated choices after receiving a positive stimulus; avoidance behaviour is the tendency to avoid the associated actions after receiving a negative stimulus (one that does not satisfy the player). Models that allow for negative stimuli tend to define an aspiration level against which achieved payoffs are evaluated. This aspiration level may be fixed or vary endogenously (Bendor et al., 2001a, 2001b).
- Whether "forgetting" is considered, *i.e.* whether recent observations weigh more than distant ones (Erev and Roth, 1998; Rustichini, 1999; Beggs, 2005).
- Whether the model imposes inertia – a positive bias in favour of the most recently selected action (Bendor et al., 2001a, 2001b).

Laslier et al. (2001) present a more formal comparison of various reinforcement learning models. Each of the features above can have important implications for the behaviour of the particular model under consideration and for the mathematical methods that are adequate for its analysis. For example, when learning slows down, theoretical results from the theory of stochastic approximation (Benveniste et al., 1990; Kushner and Yin, 1997) and from the theory of urn models can often be applied (e.g. Ianni, 2001; Beggs, 2005; Hopkins and Posch, 2005), whereas if the learning rate is constant, results from the theory of distance diminishing models (Norman, 1968, 1972) tend to be more useful (e.g. Börgers and Sarin, 1997; Bendor et al., 2001b). Similarly, imposing inertia facilitates the analysis to a great extent, since it often ensures that a positive stimulus will be followed by an increase in the probability weight on the most recently selected action at some minimum geometric rate (Bendor et al., 2001b).

A popular model of reinforcement learning in the game theory literature is the Erev-Roth (ER) model (Roth and Erev, 1995; Erev and Roth, 1998). Understanding of the ER model (also called Cumulative Proportional Reinforcement model by Laslier et al. (2001) and Laslier and Walliser (2005)) and its relation with an adjusted version of the evolutionary replicator dynamics (Weibull, 1995) has been developed in papers by Laslier et al. (2001), Hopkins (2002), Laslier and Walliser (2005), Hopkins and Posch (2005) and Beggs (2005). An extension to the ER model covering both partial and full informational environments (in the latter, a player can observe the payoffs for actions not selected), as well as linear and exponential adjustment procedures, is analysed for single person decision problems by Rustichini (1999).

Arthur (1991) proposed a model differing from the ER model only in that the step size of the learning process in ER is stochastic whereas it is deterministic in Arthur's model – but step sizes are of the same order in both (see Hopkins and Posch (2005) for details). Theoretical results for Arthur's model in games and its relation with the ordinary evolutionary replicator dynamics are given by Posch (1997), Hopkins (2002), Hopkins and Posch (2005) and Beggs (2005): despite their similarity, the ER model and Arthur's model can have different asymptotic behaviour (Hopkins and Posch, 2005).

Another important set of reinforcement models are the aspiration-based models, which allow for negative stimuli (see Bendor et al. (2001a) for an overview). The implications of aspiration-based reinforcement learning in strategic contexts have been studied thoroughly by Karandikar et al. (1998) and Bendor et al. (2001b). This line of work tends to require very mild conditions on the way learning is conducted apart from the assumption of inertia. Assuming inertia greatly facilitates the mathematical analysis, enabling the derivation of sharp predictions for long-run outcomes in 2-player repeated games, even with evolving aspirations (see e.g. Karandikar et al. (1998), Palomino and Vega-Redondo (1999), and Bendor et al. (2001b)).

The model analysed here is a variant of Bush and Mosteller's (1955) linear stochastic model of reinforcement learning (henceforth BM model). The BM model is an aspiration-based reinforcement learning model, but does not impose inertia. In contrast to the ER model and Arthur's model, it allows for negative stimuli and learning does not fade with time. A special case of the BM model where all stimuli are positive was originally considered by Cross (1973), and analysed by Börgers and Sarin (1997), who also related it to the replicator dynamics. Börgers and Sarin (2000) studied an extension of the BM model where aspirations evolve simultaneously with choice probabilities in single person decision contexts. Here, we develop Börgers and Sarin's work by analysing the dynamics of the BM model in 2×2 games where aspiration levels are fixed, but not necessarily below the lowest payoff, so negative stimuli are possible. These dynamics have been explored by Hegselmann and Flache (2000), Macy and Flache (2002) and Flache and Macy (2002) in 2×2 social dilemmas using computer simulation. Here we formalize their analyses and extend their results to cover any 2×2 game.

In contrast to other reinforcement learning models in the literature, we show that, in general, the asymptotic behaviour of the BM model cannot be approximated using the continuous time limit version of its expected motion. Such an approximation may be valid over bounded time intervals but it can deteriorate as the time horizon increases. This important point –originally emphasized by Boylan (1992; 1995) in a somewhat different context– was already noted by Börgers and Sarin (1997) in the BM model for strictly positive stimuli, and has also been found in other models since then (Beggs, 2002). The asymptotic behaviour of the BM model is characterized in the present chapter using the theory of distance diminishing models (Norman, 1968, 1972). Börgers and Sarin (1997) also used this theory to analyse the case where aspirations are below the minimum payoff; here we extend their results for 2×2 games where aspiration levels can have any fixed value.

4.2. The BM model

The model we analyse here is an elaboration of a conventional Bush-Mosteller (Bush and Mosteller, 1955) stochastic learning model for binary choice. In this model, players decide what action to select stochastically: each player's strategy is defined by the probability of undertaking each of the two actions available to them. After every player has selected an action according to their probabilities, every player receives the corresponding payoff and revises her strategy. The revision of strategies takes place following a reinforcement learning approach: players increase their probability of undertaking a certain action if it led to payoffs above their aspiration level, and decrease this probability otherwise. When learning, players in the BM model use only information concerning their own past choices and payoffs, and ignore all the information regarding the payoffs and choices of their counterparts.

More precisely, let I = {1, 2} be the *set of players* in the game, and let Y_i be the *pure-strategy space* for each player $i \in I$. For convenience, and without loss of generality, later we will call the actions available to each of the players C (for Cooperate) and D (for Defect). Thus Y_i = {C, D}. Let u_i be the *payoff function* that gives player i's payoff for each profile $\boldsymbol{y} = (y_1, y_2)$ of pure strategies, where $y_i \in Y_i$ is a pure strategy for player i. As an example, u_i(C, D) denotes the payoff obtained by player i when player 1 cooperates and player 2 defects. Let $Y = \times_{i\in I} Y_i$ be the space of pure-strategy profiles, or possible outcomes of the game. We can represent any mixed strategy for player i as a *vector* $\boldsymbol{p_i}$ in the *unit simplex* Δ^1, where the jth coordinate $p_{i,j} \in$ R of the vector $\boldsymbol{p_i}$ is the probability assigned by $\boldsymbol{p_i}$ to player i's jth pure strategy. A *mixed-strategy profile* is a vector $\boldsymbol{p} = (\boldsymbol{p_1}, \boldsymbol{p_2})$, where each component $\boldsymbol{p_i} \in \Delta^1$ represents a mixed strategy for player $i \in I$.

In the BM model, strategy updating takes place in two steps. First, after outcome $\boldsymbol{y^n} = (y_1^n, y_2^n)$ in time-step n, each player i calculates her stimulus $s_i(\boldsymbol{y^n})$ for the action just chosen y_i^n according to the following formula:

$$s_i(\boldsymbol{y}) = \frac{u_i(\boldsymbol{y}) - A_i}{\mathbf{sup}_{\boldsymbol{k}\in Y} \ | u_i(\boldsymbol{k}) - A_i \ |}$$

where A_i is player i's aspiration level. Hence the stimulus is always a number in the interval [–1, 1]. Note that players are assumed to know $\mathbf{sup}_{\boldsymbol{k}\in Y} \mid u_i(\boldsymbol{k}) - A_i \mid$. Secondly, having calculated their stimulus $s_i(\boldsymbol{y^n})$ after the outcome $\boldsymbol{y^n}$, each player i updates her probability p_{i,y_i} of undertaking the selected action y_i as follows:

$$p_{i,y_i}^{n+1} = \begin{cases} p_{i,y_i}^{n} + l_i \cdot s_i(\boldsymbol{y}^n) \cdot (1 - p_{i,y_i}^{n}) & \text{if } s_i(\boldsymbol{y}^n) \geq 0 \\ p_{i,y_i}^{n} + l_i \cdot s_i(\boldsymbol{y}^n) \cdot p_{i,y_i}^{n} & \text{if } s_i(\boldsymbol{y}^n) < 0 \end{cases} \quad [4\text{-}1]$$

where p_{i,y_i}^{n} is player i's probability of undertaking action y_i in time-step n, and l_i is player i's learning rate ($0 < l_i < 1$). Thus, the higher the stimulus magnitude (or the learning rate), the larger the change in probability. The updated probability for the action not selected derives from the constraint that probabilities must add up to one.

A 2×2 BM model parameterization requires specifying both players' payoff function u_i, aspiration level (A_i), and learning rate (l_i). Unless otherwise stated, the analysis conducted here is valid for any 2×2 game but, for illustrative purposes, we focus on 2×2 symmetric social dilemma games where both players are parameterised in exactly the same way (homogeneous models). A certain parameterisation of such a homogeneous model will be specified using the template [*Temptation* , *Reward* , *Punishment* , *Sucker* | *A* | *l*][2].

The following notation will also be useful. A parameterized model will be denoted $\boldsymbol{S}$ (for System). Since the state of any particular system can be fully characterized by the strategy profile $\boldsymbol{p}$, $\boldsymbol{p}$ will also be named *state of the system*. Note, however, that there are only two independent variables in $\boldsymbol{p}$, so the state of the game can be determined using a two-dimensional vector [$p_{1,C}$, $p_{2,C}$], where $p_{i,C}$ is player i's probability to cooperate (the actual name of the action is irrelevant for the mathematical analysis). Let $\boldsymbol{P}_n(\boldsymbol{S})$ be the state of a system $\boldsymbol{S}$ in time-step n. Note that $\boldsymbol{P}_n(\boldsymbol{S})$ is a random variable and $\boldsymbol{p}$ is a particular value of that variable; the sequence of random variables $\{\boldsymbol{P}_n(\boldsymbol{S})\}_{n\geq 0}$ constitutes a discrete-time Markov process with potentially infinite transient states. In a slight abuse of notation we refer to such a process $\{\boldsymbol{P}_n(\boldsymbol{S})\}_{n\geq 0}$ as the BM process $\boldsymbol{P}_n$.

4.3. Attractors in the Dynamics of the System

Using computer simulation, Macy and Flache (2002) described two types of learning-theoretic equilibria that govern the dynamics of the BM model: self-reinforcing equilibria (SRE), and self-correcting equilibria (SCE). These are not static equilibria, but strategy profiles which act as attractors in the sense that, under certain conditions, the system will tend to approach them or linger around them. Here, we formalize these two concepts.

We define an SRE as an absorbing state of the system (*i.e.* a state $\boldsymbol{p}$ that cannot be abandoned) where both players receive a positive stimulus[11]. An SRE corresponds to a pair of pure strategies ($p_{i,j}$ is either 0 or 1) such that its certain associated outcome gives a strictly positive stimulus to both players (henceforth a *mutually satisfactory outcome*). For example, the strategy profile $[\ p_{1,\mathrm{C}}\ , p_{2,\mathrm{C}}\ ] = [\ 1\ ,\ 1\ ]$ is an SRE if both players' aspiration levels are below their respective $Reward_i$. Escape from an SRE is impossible since no player will change her strategy. More importantly, SREs act as attractors: near an SRE, there is a high chance that the system will move towards it, because there is a high probability that its associated mutually satisfactory outcome will occur, and this brings the system even closer to the SRE. The number of SREs in a system is the number of outcomes where both players obtain payoffs above their respective aspiration levels.

Flache and Macy (2002, p. 634) define SCEs in the following way: "The SCE obtains when the expected change of probabilities is zero and there is a positive probability of punishment as well as reward". In this context, punishment means negative stimulus while reward means positive stimulus; the expected change of probability for one player is defined as the sum of the possible changes in probability the player might experience weighted by the likelihood of such changes actually happening. As we show below, SCEs defined in this way are not necessarily attractors, but may be unstable saddle points where small

[11] The concept of SRE is extensively used by Macy and Flache but we have not found a clear definition in their papers (Flache and Macy, 2002; Macy and Flache, 2002). Sometimes their use of the word SRE seems to follow our definition (e.g. Macy and Flache, 2002, p. 7231), but often it seems to denote a mutually satisfactory outcome (e.g. Macy and Flache, 2002, p. 7231) or an infinite sequence of such outcomes (e.g. Macy and Flache, 2002, p. 7232).

perturbations can cause expected probabilities to move away from them. Figure 4-1 represents the expected movement after one time-step for different states of the system in a Stag Hunt game. The Expected Motion (**EM**) of a system ***S*** in state ***p*** for the following iteration is given by a function vector $\mathbf{EM}^{S}(\boldsymbol{p})$ whose components are, for each player, the expected change in the probabilities of undertaking each of the two possible actions. Mathematically,

$$\mathbf{EM}^{S}(\boldsymbol{p}) \equiv \mathbf{E}(\Delta \boldsymbol{P}_n(\boldsymbol{S}) \mid \boldsymbol{P}_n(\boldsymbol{S}) = \boldsymbol{p})$$

In the context of 2×2 social dilemma games, the two independent components of the equation above can be rewritten as follows:

$$\mathrm{EM}^{S}_{i,\mathrm{C}}(\boldsymbol{p}) =$$
$$\Pr\{\mathrm{CC}\} \cdot \Delta p_{i,\mathrm{C}}\big|_{\mathrm{CC}} + \Pr\{\mathrm{CD}\} \cdot \Delta p_{i,\mathrm{C}}\big|_{\mathrm{CD}} + \Pr\{\mathrm{DC}\} \cdot \Delta p_{i,\mathrm{C}}\big|_{\mathrm{DC}} + \Pr\{\mathrm{DD}\} \cdot \Delta p_{i,\mathrm{C}}\big|_{\mathrm{DD}}$$

where $\mathrm{EM}^{S}_{i,\mathrm{C}}(\boldsymbol{p})$ is the expected change in player *i*'s probability to cooperate, and {CC, CD, DC, DD} represent the four possible outcomes that may occur. Note that in general the expected change will not reflect the actual change in a simulation run, and to make this explicit we have included the trace of a simulation run starting in state [$p_{1,\mathrm{C}}$, $p_{2,\mathrm{C}}$] = [0.5 , 0.5] in Figure 4-1. The expected change – represented by the arrows in Figure 4-1 – is calculated considering the four possible changes that could occur (see equation above), whereas the actual change in a simulation run – represented by the numbered balls in Figure 4-1 – is only *one* of the four possible changes (*e.g.* $\Delta p_{i,\mathrm{C}}\big|_{\mathrm{CC}}$, if both agents happen to cooperate). The source code used to create every figure in this chapter is available in the Supporting Material.

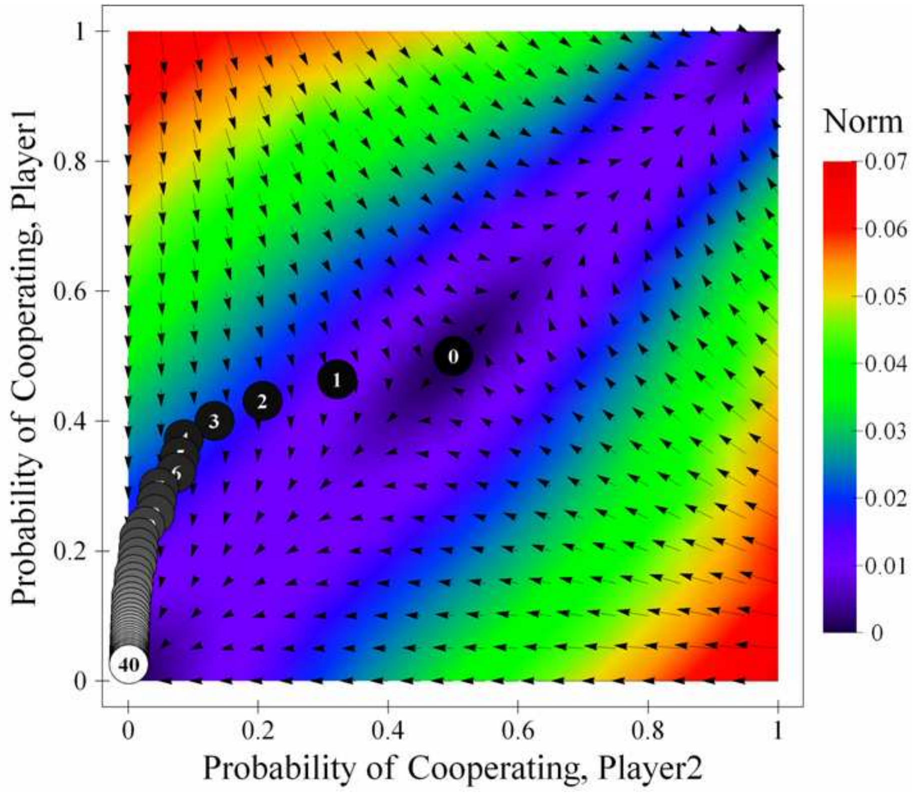


Figure 4-1. Expected motion of the system in a Stag Hunt game parameterised as [3 , 4 , 1 , 0 | 0.5 | 0.5]^2, together with a sample simulation run (40 iterations). The arrows represent the expected motion for various states of the system; the numbered balls show the state of the system after the indicated number of iterations in the sample run. The background is coloured using the norm of the expected motion. For any other learning rate the size of the arrows would vary but their direction would be preserved.

The state [$p_{1,C}$, $p_{2,C}$] = [0.5 , 0.5] in Figure 4-1 is an example of a strategy profile that satisfies Flache and Macy's requirements for SCE, but where small deviations tend to lead the system away from it (saddle point). To avoid such undesirable situations where an SCE is not self-correcting, we redefine the concept of SCE in a more restrictive way: an SCE of a system ***S*** is an asymptotically stable critical point (Mohler, 1991) of differential equation [4-2] (the continuous time limit approximation of the system's expected motion).

$$\dot{\boldsymbol{f}} = \mathbf{EM}^{S}(\boldsymbol{f}) \qquad [4\text{-}2]$$

Roughly speaking this means that all trajectories in the phase plane of Eq. [4-2] that at some instant are sufficiently close to the SCE will approach the SCE as the parameter *t* (time) approaches infinity and remain close to it at all future times. Note that, with this definition, there could be a state of the system that is an SRE and an SCE at the same time (this is not possible using Flache and Macy's definitions of SRE and SCE).

Figure 4-2 shows several trajectories for the differential equation corresponding to the Stag Hunt game used in Figure 4-1. It can be clearly seen that state [$p_{1,C}$, $p_{2,C}$] = [0.5 , 0.5] is not an SCE according to our definition, since there are trajectories that get arbitrarily close to it, but then escape from its neighbourhood.

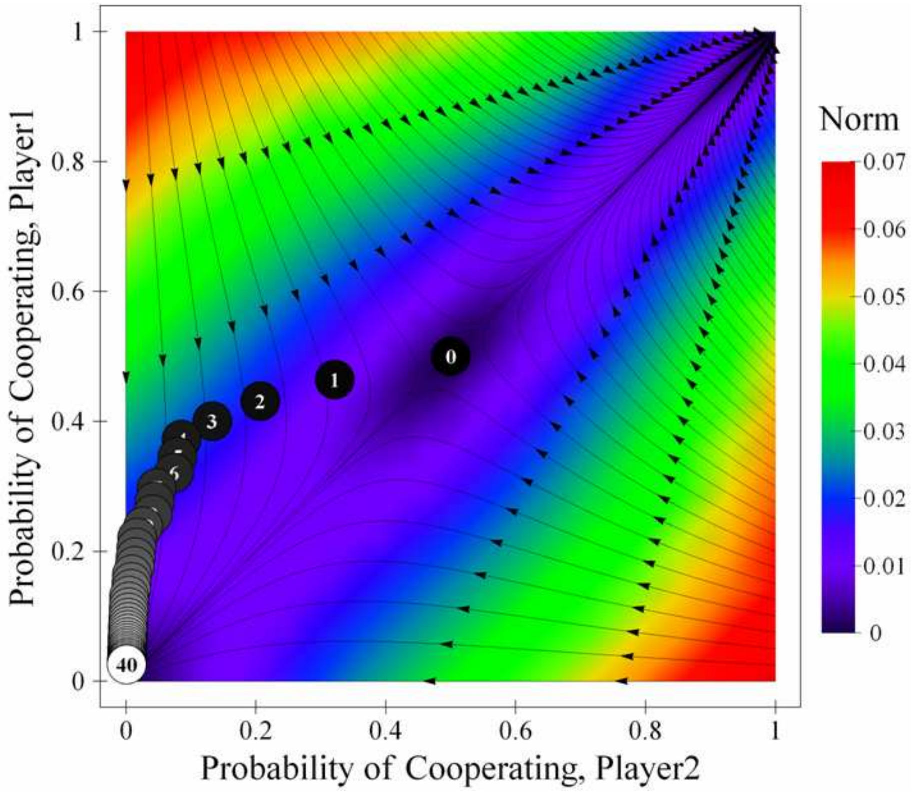


Figure 4-2. Trajectories in the phase plane of the differential equation corresponding to a Stag Hunt game parameterised as [3 , 4 , 1 , 0 | 0.5 | 0.5]^2, together with a sample simulation run (40 iterations). The background is coloured using the norm of the expected motion.

Figure 4-3 shows some trajectories of the differential equation corresponding to the Prisoner's Dilemma parameterised as [4 , 3 , 1 , 0 | 2 | l]^2. This system exhibits a unique SCE at [$p_{1,C}$, $p_{2,C}$] = [0.37 , 0.37] and a unique SRE at [$p_{1,C}$, $p_{2,C}$] = [1 , 1]. The two independent components of the function **EM(*p*)** for this system can be written as follows:

$$[\mathrm{EM}^{S}_{1,\mathrm{C}}(\boldsymbol{p}),\mathrm{EM}^{S}_{2,\mathrm{C}}(\boldsymbol{p})] = l\left[p_{1,\mathrm{C}}p_{2,\mathrm{C}} \quad p_{1,\mathrm{C}}(1-p_{2,\mathrm{C}}) \quad (1-p_{1,\mathrm{C}})p_{2,\mathrm{C}} \quad (1-p_{1,\mathrm{C}})(1-p_{2,\mathrm{C}})\right]\cdot \begin{bmatrix} (1-p_{1,\mathrm{C}})/2 & (1-p_{2,\mathrm{C}})/2 \\ -p_{1,\mathrm{C}} & -p_{2,\mathrm{C}} \\ -p_{1,\mathrm{C}} & -p_{2,\mathrm{C}} \\ (1-p_{1,\mathrm{C}})/2 & (1-p_{2,\mathrm{C}})/2 \end{bmatrix}$$

And the associated differential equation is

$$\left[\frac{df_1}{dt},\frac{df_2}{dt}\right] = l\left[f_1 f_2 \quad f_1(1-f_2) \quad (1-f_1)f_2 \quad (1-f_1)(1-f_2)\right]\cdot \begin{bmatrix} (1-f_1)/2 & (1-f_2)/2 \\ -f_1 & -f_2 \\ -f_1 & -f_2 \\ (1-f_1)/2 & (1-f_2)/2 \end{bmatrix}$$

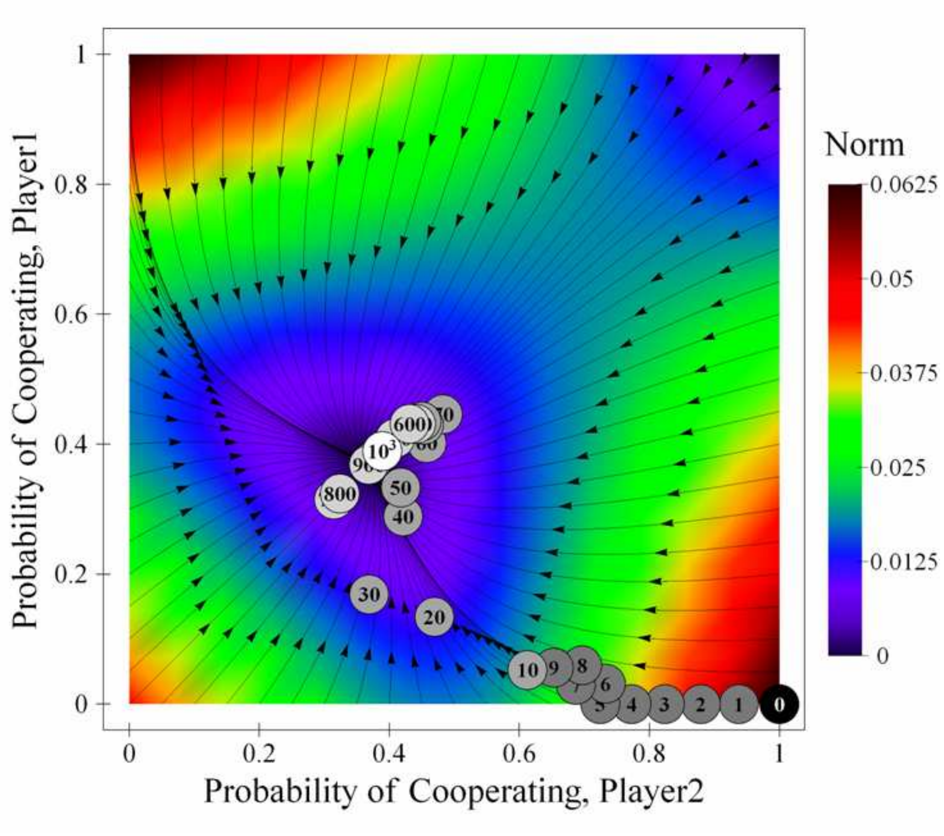


Figure 4-3. Trajectories in the phase plane of the differential equation corresponding to the Prisoner's Dilemma game parameterised as [4 , 3 , 1 , 0 | 2 | l]^2, together with a sample simulation run ($l = 2^{-4}$). This system has a SCE at [$p_{1,\mathrm{C}}$, $p_{2,\mathrm{C}}$] = [0.37 , 0.37]. The background is coloured using the norm of the expected motion.

Let $\boldsymbol{f_x}(t)$ denote the solution of the differential equation [4-2] for some initial state $\boldsymbol{x}$. As an example, Figure 4-4 shows $\boldsymbol{f_x}(t)$ for the Prisoner's Dilemma game parameterised as $[\ 4\ ,\ 3\ ,\ 1\ ,\ 0\ |\ 2\ |\ l\ ]^2$ for different (and symmetric) initial conditions $[\ p_{1,\mathrm{C}}\ ,\ p_{2,\mathrm{C}}\ ] = [\ x_0\ ,\ x_0\ ]$. For this particular case and settings, the two independent components of $\boldsymbol{f_x}(t)$ corresponding to each player's probability to cooperate – denoted $f_{i,x}(t)$ – take the same value at any given t, so the representation in Figure 4-4 corresponds to both these independent components. Convergence to the SCE at [0.37 , 0.37] can be clearly observed for every initial condition $[\ x_0\ ,\ x_0\ ]$, except for $[\ x_0\ ,\ x_0\ ]$ = [1, 1], which is the SRE.

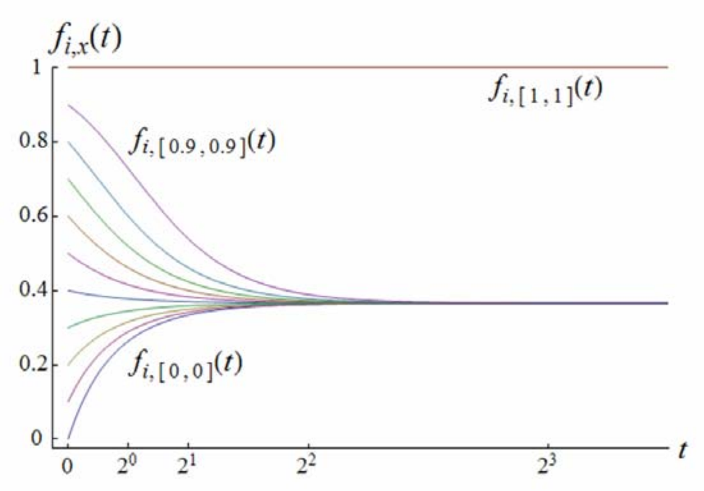


Figure 4-4. Solutions of differential equation [4-2] for the Prisoner's Dilemma game parameterised as $[\ 4\ ,\ 3\ ,\ 1\ ,\ 0\ |\ 2\ |\ l\ ]^2$ with different (and symmetric) initial conditions $[\ p_{1,\mathrm{C}}\ ,\ p_{2,\mathrm{C}}\ ] = [x_0\ ,\ x_0]$. This system has a unique SCE at $[\ p_{1,\mathrm{C}}\ ,\ p_{2,\mathrm{C}}\ ]$ = [0.37 , 0.37] and a unique SRE at $[\ p_{1,\mathrm{C}}\ ,\ p_{2,\mathrm{C}}\ ]$ = [1 , 1].

The expected motion at any point $\boldsymbol{p}$ in the phase plane is a vector tangent to the unique trajectory to which that point belongs. The use of expected motion (or mean-field) approximations to understand simulation models and to design interesting experiments has already proven to be very useful in the literature (e.g. Huet et al (2007); Galán and Izquierdo (2005); Edwards et al. (2003); Castellano, Marsili, and Vespignani (2000)). Note, however, that such approaches are approximations whose validity may be constrained to specific conditions: as we can see in Figure 4-3, simulation runs and trajectories will not coincide in general. A crucial question to characterize the dynamics of learning models, and one to which stochastic approximation theory (Benveniste et al., 1990; Kushner and Yin, 1997) is devoted, is whether the *expected* and *actual* motion of the system should become arbitrarily close in the long run. This is generally true for processes whose motion slows down at an appropriate rate (as explained by Hopkins and Posch (2005) when studying the ER model), but not necessarily so in other cases. We show in the next sections that the BM model's *asymptotic* behaviour can be dramatically different from that suggested by its associated ODE, which is, however, very relevant for characterizing the *transient* dynamics of the system, particularly with small learning rates. From now on we will use our definitions of SRE and SCE.

4.4. Attractiveness of SREs

Macy & Flache's experiments (Flache and Macy, 2002; Macy and Flache, 2002) with the BM model showed a puzzling phenomenon. A significant part of their analysis consisted in studying, in a Prisoner's Dilemma in which mutual cooperation was mutually satisfactory (i.e. $A_i < Reward_i = u_i$(C, C)), the proportion of simulation runs that "locked" into mutual cooperation. Such "lock-in rates" were reported to be as high as 1 in some experiments. However, starting from an initial state which is not an SRE, the BM model specifications guarantee that after any finite number of iterations any outcome has a positive probability of occurring (i.e. strictly speaking, lock-in is impossible)[12]. To investigate this apparent contradiction we conducted some qualitative analyses that we present here to familiarise the reader with the complex dynamics of this model. Our first qualitative analysis consisted in studying the expected dynamics of the model. Figure 4-5 illustrates the expected motion of a system extensively studied by Macy & Flache: the Prisoner's Dilemma game parameterised as [4 , 3 , 1 , 0 | 2 | 0.5]^2. As we saw before, this system features a unique SCE at [$p_{1,C}$, $p_{2,C}$] = [0.37 , 0.37] and a unique SRE at [$p_{1,C}$, $p_{2,C}$] = [1 , 1]. Figure 4-5 also includes the trace of a sample simulation run. Note that the only difference between the

[12] The specification of the model is such that probabilities cannot reach the extreme values of 0 or 1 starting from any other intermediate value. Therefore if we find a simulation run that has actually ended up in an SRE starting from any other state, we know for sure that such simulation run did not follow the specifications of the model (e.g. perhaps because of floating-point errors). For a detailed analysis of the effects of floating point errors in computer simulations, with applications to this model in particular, see Izquierdo and Polhill (2006), Polhill and Izquierdo (2005), Polhill et al. (2006), Polhill et al. (2005).

parameterisation of the system shown in Figure 4-3 and that shown in Figure 4-5 is the value of the learning rate.

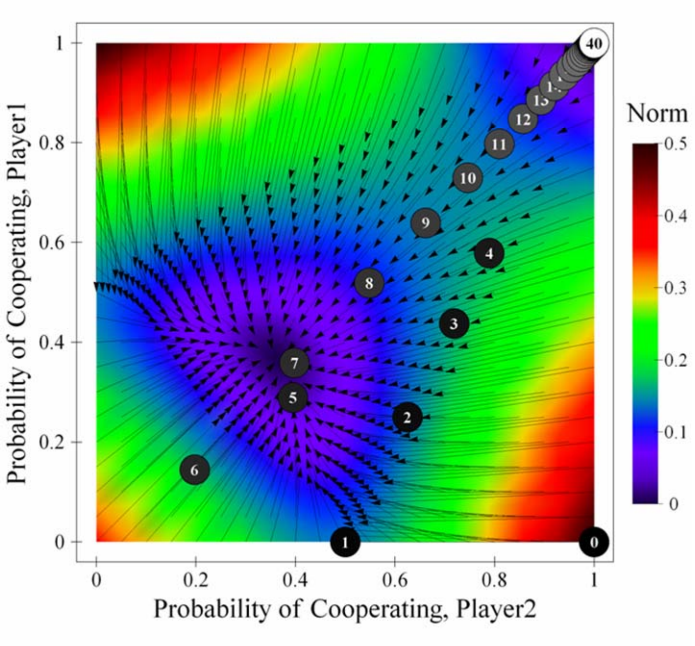


Figure 4-5. Expected motion of the system in a Prisoner's Dilemma game parameterised as [4 , 3 , 1 , 0 | 2 | 0.5]^2, with a sample simulation run.

Figure 4-5 shows that the expected movement from any state is towards the SCE, except for the only SRE, which is an absorbing state. In particular, near the SRE, where both probabilities are high but different from 1, the distribution of possible movements is very peculiar: there is a very high chance that both agents will cooperate and consequently move a small distance towards the SRE, but there is also a positive chance, tiny as it may be, that one of the agents will defect, causing both agents to jump away from the SRE towards the SCE. The improbable, yet possible, leap away from the SRE is of such magnitude that the resulting expected movement is biased towards the SCE despite the unlikelihood of such an event actually occurring. The dynamics of the system can be further explored analysing the most likely movement from any given state, which is represented in Figure 4-6.

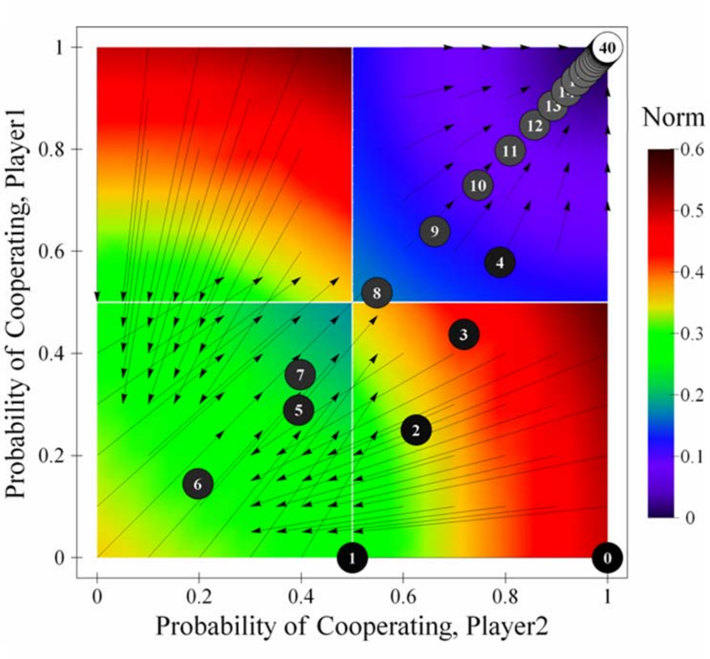


Figure 4-6 Figure showing the most likely movements at some states of the system in a Prisoner's Dilemma game parameterised as [4 , 3 , 1 , 0 | 2 | 0.5]^2, with a sample simulation run. The background is coloured using the norm of the most likely movement.

Figure 4-6 differs significantly from Figure 4-5; it shows that the most likely movement in the upper-right quadrant of the state space is towards the SRE. Thus the walk towards the SRE is characterized by a fascinating puzzle: on the one hand, the most likely movement leads the system towards the SRE, which is even more likely to be approached the closer we get to it; on the other hand, the SRE cannot be reached in any finite number of steps and the expected movement as defined above is to walk away from it (see Figure 4-5).

It is also interesting to note in this game that, starting from any mixed (interior) state, both players have a positive probability of selecting action D in any future time-step, but there is also a positive probability that both players will engage in an infinite chain of the mutually satisfactory event CC forever, i.e., that neither player will ever take action D from then onwards. This latter probability can be calculated using a result derived by Professor Jörgen W. Weibull (see Appendix A). The probability of starting an infinite chain of CC events depends largely on the value of the learning rate *l*. Figure 4-7 shows the probability of starting an infinite chain of the mutually satisfactory outcome CC in a Prisoner's Dilemma game parameterised as [4 , 3 , 1 , 0 | 2 | *l*]^2, for different learning rates *l*, and different initial probabilities to cooperate x_0 (the same probability for both players). For some values, the probability of immediately starting an infinite chain of mutual cooperation can be surprisingly high (e.g. for l = 0.5 and initial conditions [x_0 , x_0] = [0.9 , 0.9] such probability is approximately 44%).

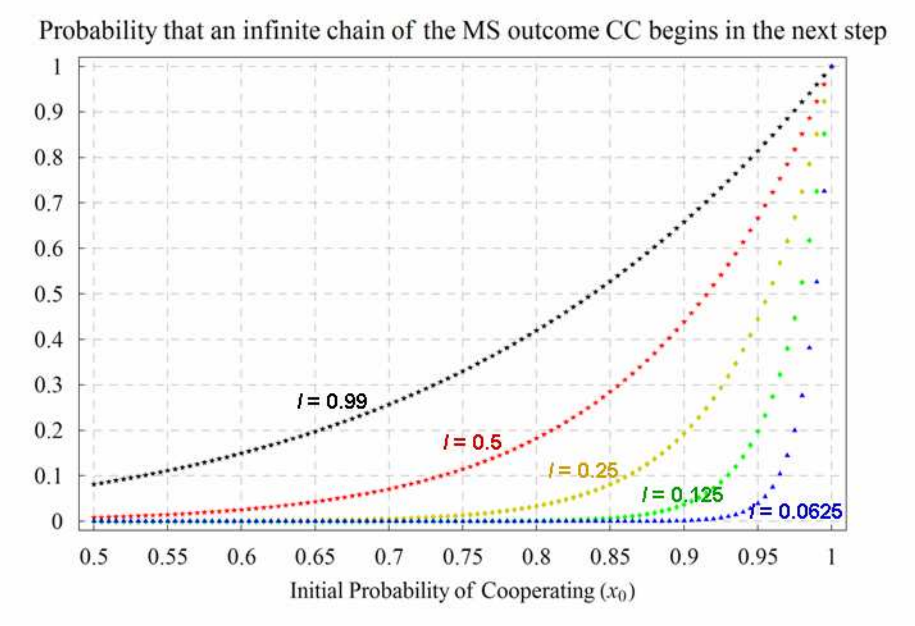


Figure 4-7. Probability of starting an infinite chain of the Mutually Satisfactory (MS) outcome CC in a Prisoner's Dilemma game parameterised as [4 , 3 , 1 , 0 | 2 | l]^2. The 5 different (coloured) series correspond to different learning rates l. The variable x_0, represented in the horizontal axis, is the initial probability of cooperating for both players.

In summary, assuming that aspirations are different from payoffs, a BM process that starts in an initial state different from an SRE will never reach an SRE in finite time, and there is always a positive probability that the process leaves the proximity of an SRE. However, if there is some SRE, there is also a positive probability that the system will approach it indefinitely (i.e. forever) through an infinite chain of the mutually satisfactory outcome associated to the SRE.

4.5. Three Dynamic Regimes

In the general case, the dynamics of the BM model may exhibit three different regimes: medium run, long run, and ultralong run. This terminology is borrowed from Binmore and Samuelson (1993) and Binmore et al. (1995, p. 10), who reserve the term short run for the initial conditions. The medium run is '*the time intermediate between the short run* [i.e. initial conditions] *and the long run, during which the adjustment to equilibrium is occurring*'. The long run is '*the time span* *needed for the system to reach the vicinity of the first equilibrium in whose neighborhood it will linger for some time*'. Finally, the ultralong run is '*a period of time long enough for the asymptotic distribution to be a good description of the behavior of the system*'.

Binmore et al.'s terminology is particularly useful for our analysis because it is often the case in the BM model that the transient dynamics of the system are dramatically different from its asymptotic behaviour. Whether the three different regimes (i.e. medium, long, and ultralong run) are clearly distinguishable strongly depends on the players' learning rates. For high learning rates the system quickly approaches its asymptotic behaviour and the distinction between the different regimes is not particularly useful. For small learning rates, however, the three different regimes can be clearly observed.

In brief, it is shown in the following section that with sufficiently small learning rates l_i and number of iterations n not too large ($n \cdot l_i$ bounded), the medium run dynamics of the system are best characterised by the trajectories in the phase plane of eq. [4-2]. Under these conditions, SCEs constitute the '*the first equilibrium in whose neighborhood it* [the system] *will linger for some time*' and, as such, they usefully characterize the long run dynamics of the system. After a potentially very lengthy long-run regime in the neighborhood of an SCE, the system will eventually reach its ultralong run behaviour, which in most BM systems consists in approaching an SRE asymptotically (see formal analysis below).

For an illustration of the different regimes, consider once again the Prisoner's Dilemma game parameterised as $[\ 4\ ,\ 3\ ,\ 1\ ,\ 0\ |\ 2\ |\ l\ ]^2$. It is shown below that this system asymptotically converges to its unique SRE with probability 1 regardless of the value of l. The evolution of the probability to cooperate with initial state $[p_{1,C}\ ,\ p_{2,C}] = [\ 0.5\ ,\ 0.5\ ]$ (with these settings the probability is identical for both players) is represented in the rows of Figure 4-8 for different learning rates l.

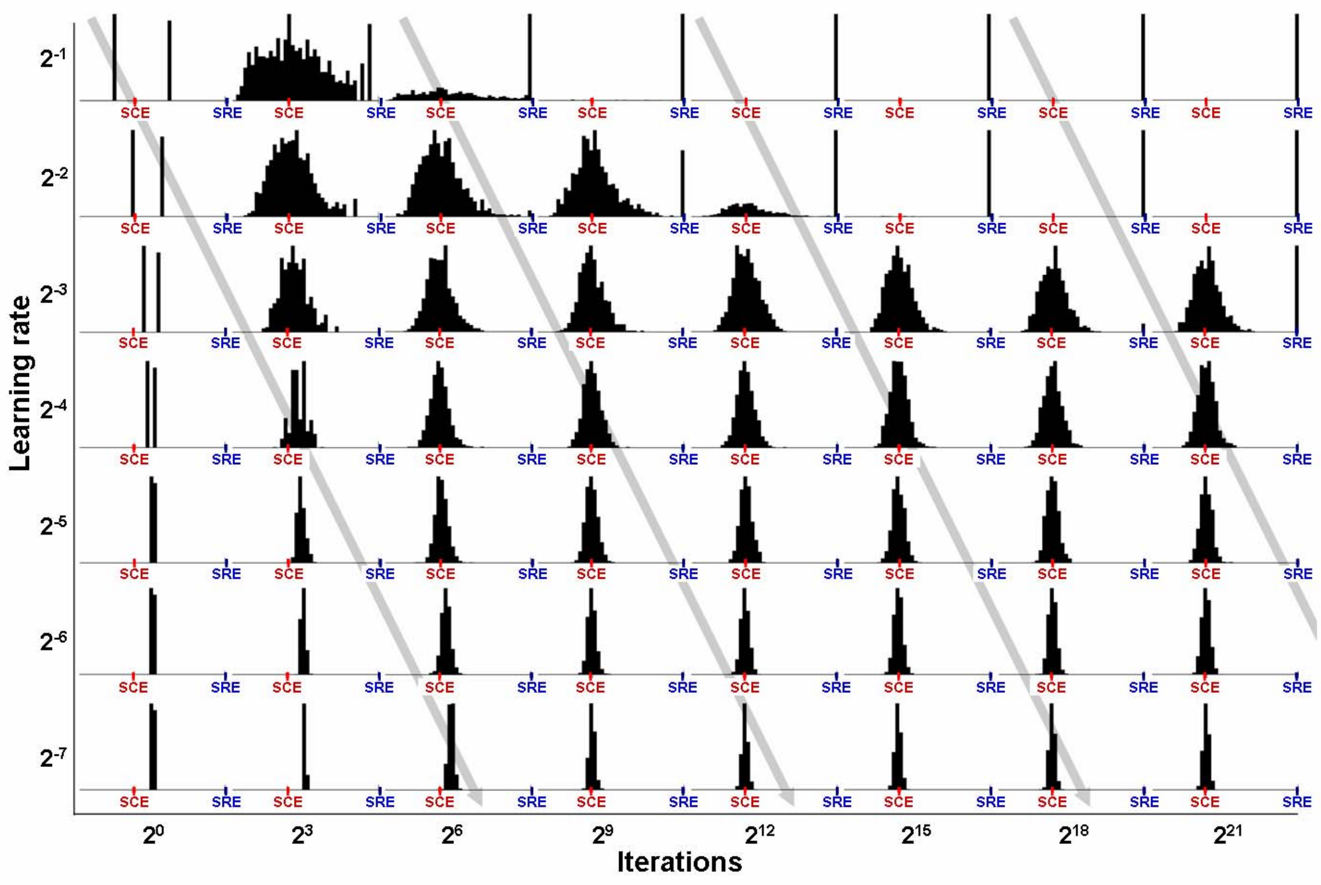


Figure 4-8. Histograms representing the probability to cooperate for one player (both players' probabilities are identical) after *n* iterations, for different learning rates $l_i = l$, with $A_i = 2$, in a symmetric Prisoner's Dilemma with payoffs [4 , 3 , 1 , 0]. Each histogram has been calculated over 1,000 simulation runs. The initial probability for both players is 0.5. The significance of the gray arrows will be explained later in the text.

For l = 0.5 (see top row in Figure 4-8), after only 2^9 = 512 iterations, the probability that both players will be almost certain to cooperate is very close to 1, and it remains so thereafter. For $l = 2^{-4}$ and lower learning rates, however, the distribution is still clustered around the SCE even after 2^{21} = 2097152 iterations. With low learning rates, the chain of events that is required to escape from the neighbourhood of the SCE is extremely unlikely, and therefore this long run regime seems to persist indefinitely. However, given sufficient time, such a chain of coordinated moves will occur, and the system will eventually reach its ultralong run regime, i.e. almost-certain mutual cooperation. The following sections are devoted to the formal analysis of the transient and asymptotic dynamics of the BM model. The proofs of every proposition in this chapter are included in Appendix A.

4.6. Transient Dynamics

As mentioned above, when learning takes place by large steps the system quickly approaches its asymptotic behaviour, and no clear (transient) patterns are observed before it does so (see top row in Figure 4-8). With small learning rates, however, the two transient regimes, which may be significantly different from the asymptotic regime, are clearly distinguishable. This section shows that SCEs are powerful attractors of the *actual* dynamics of the system when learning occurs by small steps. Specifically, it is demonstrated that the BM process $\boldsymbol{P}_n$ follows the trajectories of its associated ODE with probability approaching 1 as learning rates decrease and n is kept within certain limits.

Consider a family of BM systems $\boldsymbol{S}^l$ whose members, indexed in $l = l_1$, only differ in both players' learning rates, and such that l_1/l_2 is a fixed constant for every model in the family. Let $\boldsymbol{P}_n^l = \boldsymbol{P}_n(\boldsymbol{S}^l)$ be the family of stochastic processes associated with such a family of systems $\boldsymbol{S}^l$. As an example, note that Figure 4-8 shows simulation runs of seven stochastic processes $(\boldsymbol{P}_n(\boldsymbol{F}^{0.5}), \boldsymbol{P}_n(\boldsymbol{F}^{0.25}) \ldots)$ belonging to one particular family $\boldsymbol{F}^l$. Consider the ODE given by eq. [4-3] below, and let $\boldsymbol{f}_{\boldsymbol{x}}(t)$ be the trajectory of this ODE with initial state $\boldsymbol{x}$.

$$\dot{\boldsymbol{f}} = \frac{1}{l}\mathbf{EM}^{S^l}(\boldsymbol{f}) \qquad [4\text{-}3]$$

The ODE in eq. [4-3] is common to every member of a given family, and its solution trajectories $\boldsymbol{f_x}(t)$ only differ from those given by eq. [4-2] (which determines a different ODE for each member) in the time scale, *i.e.* the representation of the trajectories of ODEs [4-2] and [4-3] in the phase plane is identical: the learning rate determines how quickly the path is walked, but the path is the same for every model of a family. Similarly, SCEs and SREs are common to every model in a family. The following proposition characterizes the medium-run (statements (i) and (ii)) and the long-run (statement (iii)) dynamics of the BM model when l is small. No conditions are imposed on players' aspirations.

Proposition 4-1: Consider the family of stochastic processes $\{\boldsymbol{P}_n^{l,x}\}_{n\geq 0}$ with initial state $\boldsymbol{P}_0^l = \boldsymbol{x}$ for every l. Let K be an arbitrary constant. For learning by small steps ($l \rightarrow 0$) and transient behaviour ($n{\cdot}l \leq K < \infty$), we have:

i. For fixed $\varepsilon > 0$ and l sufficiently small,

$$\Pr\{\max_{n\leq(K/l)} \left\|\boldsymbol{P}_n^{l,x} - \boldsymbol{f}_x(n\cdot l)\right\| > \varepsilon\} \leq C(l,K)$$

where, for fixed $K < \infty$, $C(l, K) \rightarrow 0$ as $l \rightarrow 0$. Thus, for transient behaviour and learning by small steps, we have uniform convergence in probability of $\boldsymbol{P}_n^{l,x}$ to the trajectory $\boldsymbol{f_x}$ of the ODE in [4-3].

ii. The distribution of the variable $\dfrac{\boldsymbol{P}_n^{l,x} - \boldsymbol{f}_x(n\cdot l)}{\sqrt{l}}$ converges to a normal distribution with mean 0 and variance independent of l as $l \rightarrow 0$ and $n{\cdot}l \rightarrow K < \infty$.

iii. Let $\mathrm{L}_{\boldsymbol{x}}$ be the limit set of the trajectory $\boldsymbol{f_x}(t)$. For $n = 0, 1\ldots N < \infty$, and for any $\delta > 0$, the proportion of values of $\boldsymbol{P}_n^{l,x}$ within a neighborhood $\mathrm{B}_\delta(\mathrm{L}_{\boldsymbol{x}})$ of $\mathrm{L}_{\boldsymbol{x}}$ goes to 1 (in probability) as $l \rightarrow 0$ and $N{\cdot}l \rightarrow \infty$.

To see an application of Proposition 4-1, consider the particular family $\boldsymbol{F}^l$ (Figure 4-8). Statement (i) says that when n is not too large ($n{\cdot}l$ bounded), with probability increasingly close to 1 as l decreases, the process $\boldsymbol{P}_n^x(\boldsymbol{F}^l)$ with initial state $\boldsymbol{P}_0(\boldsymbol{F}^l) = \boldsymbol{x}$ follows the trajectory $\boldsymbol{f_x}(n{\cdot}l)$ of the ODE in [4-3] within a distance never greater than some arbitrary, a priori fixed, $\varepsilon > 0$. (This proves the conjecture put forward by Börgers and Sarin (1997) in remark 2.) The trajectories corresponding to $\boldsymbol{P}_n(\boldsymbol{F}^l)$ are displayed in Figure 4-3, and the convergence of the processes to the appropriate point in the trajectory $\boldsymbol{f}_{\boldsymbol{x}}(n{\cdot}l)$ as $l \to 0$ can be appreciated following the gray arrows (which join histograms for which $n{\cdot}l$ is constant) in Figure 4-8. Figure 4-9 illustrates this convergence in the phase plane. The grey arrows in Figure 4-8 also illustrate statement (ii): the distribution of $\boldsymbol{P}_n^x(\boldsymbol{F}^l)$ approaches normality with decreasing variance as $l \to 0$, keeping $n{\cdot}l$ constant.

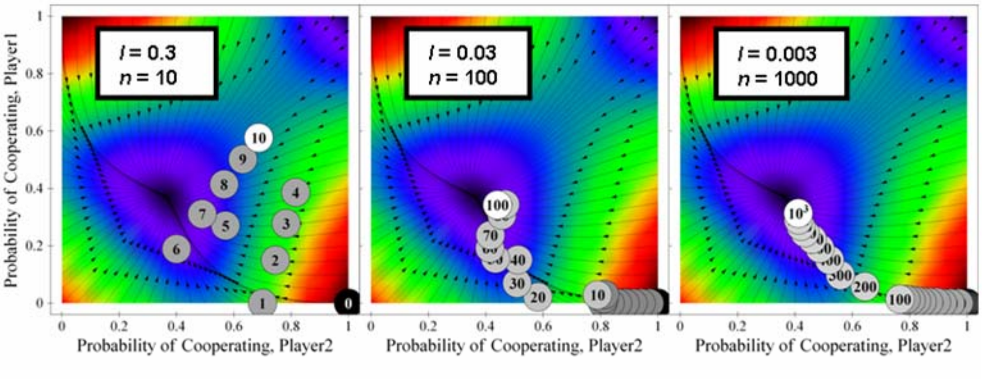


Figure 4-9. Three sample runs of a system parameterised as [4 , 3 , 1 , 0 | 2 | l]^2 for different values of n and l. The product $n{\cdot}l$ is the same for the three simulations; therefore, for low values of l, the state of the system at the end of the simulations tends to concentrate around the same point.

The fact that the trajectory $\boldsymbol{f}_{\boldsymbol{x}}$ is a good approximation for the medium-run dynamics of the system for slow learning shows the importance of SCEs as attractors of the actual dynamics of the system. To illustrate this, consider family $\boldsymbol{F}^l$ again. It can be shown using the square of the Euclidean distance to the SCE as a Liapunov function that every trajectory starting in any state different from the SRE [$p_{1,C}$, $p_{2,C}$] = [1 , 1] will end up in the SCE [$p_{1,C}$, $p_{2,C}$] = [0.37 , 0.37] – *i.e.* the limit set $\mathrm{L}_{\boldsymbol{x}}$ is formed exclusively by the SCE for any $\boldsymbol{x} \neq$ SRE (see Figure 4-3). This means that starting from any initial state $\boldsymbol{x} \neq$ SRE, if K is sufficiently large and $n < K/l$ (*i.e.* if in Figure 4-8 we consider the region to the left of a grey arrow that is sufficiently to the right), the distribution of $\boldsymbol{P}_n^x(\boldsymbol{F}^l)$ will be tightly clustered around the SCE [0.37 , 0.37] and will approach normality as n increases. Furthermore, statement (iii) says that, for any $\boldsymbol{x} \neq$ SRE, any $\delta > 0$, and $n = 0, 1 \ldots N < \infty$, the proportion of values of $\boldsymbol{P}_n^x(\boldsymbol{F}^l)$ within a neighbourhood B_δ(SCE) of the SCE goes to 1 (in probability) as $l \to 0$ and $N{\cdot}l \to \infty$. This is the long run. Remember, however, that given any l, $\boldsymbol{P}_n^x(\boldsymbol{F}^l)$ will eventually converge to the unique SRE [1, 1] in the ultralong run ($n \rightarrow \infty$). This is proved in the following section.

4.7. Asymptotic Behaviour

This section presents theoretical results on the asymptotic (i.e. ultralong run) behaviour of the BM system. Note that with low learning rates the system may take an extraordinarily long time to reach its ultralong-run behaviour (e.g. see bottom row in Figure 4-8).

Proposition 4-2: In any 2×2 game, assuming players' aspirations are different from their respective payoffs ($u_i(\boldsymbol{d}) \neq A_i$ for all i and $\boldsymbol{d}$) and below their respective *maximin*[13], the BM process $\boldsymbol{P}_n$ converges to an SRE with probability 1 (the set formed by all SREs is asymptotically reached with probability 1). If the initial state is completely mixed, then every SRE can be asymptotically reached with positive probability.

Proposition 4-3: In any 2×2 game, assuming players' aspirations are different from their respective payoffs and above their respective *maximin*:

i. If there is any SRE then the BM process $\boldsymbol{P}_n$ converges to an SRE with probability 1 (the set formed by all SREs is asymptotically reached with probability 1). If the initial state is completely mixed, then every SRE can be asymptotically reached with positive probability.
ii. If there is no SRE then the BM process $\boldsymbol{P}_n$ is ergodic[14] with no absorbing state.

[13] Maximin is the largest possible payoff players can guarantee themselves in a single-stage game using pure strategies.

[14] Following Norman (1968, p. 67), by 'ergodic' we mean that the sequence of stochastic kernels defined by the n-step transition probabilities of the Markov process associated with the BM system converges uniformly to a unique limiting kernel independent of the initial state. Intuitively, this means that the asymptotic probability distribution over the states of the system (*i.e.* the distribution of $\boldsymbol{P}_n$ when $n \rightarrow \infty$) is unique and does not depend on the initial state.

Corollary to Proposition 4-3: Consider any of the three 2×2 social dilemma games: Prisoner's Dilemma, Chicken, and Stag Hunt (see section 3.1). Assuming players' aspirations are different from their respective payoffs and above their respective *maximin*:

i. The BM process $\boldsymbol{P}_n$ is ergodic.
ii. There is an SRE if and only if mutual cooperation is satisfactory for both players. In that case, the process converges to the unique SRE (*i.e.* certain mutual cooperation) with probability 1.

Since most BM systems end up converging to an SRE in the ultralong run, but their transient dynamics with slow learning are governed by their associated ODE, mathematical results that relate SREs with the solutions of the ODE can be particularly useful. The following proposition shows that the Nash equilibrium concept is key to determining the stability of SREs under the associated ODE.

Proposition 4-4: Consider the BM process $\boldsymbol{P}_n$ and its associated ODE (eq. [4-2] or [4-3]) in any 2×2 game:

i. All SREs whose associated outcome is not a Nash equilibrium are unstable.
ii. All SREs whose associated outcome is a strict Nash equilibrium where at least one unilateral deviation leads to a satisfactory outcome for the non-deviating player are asymptotically stable (*i.e.* they are SCEs too).

Thus, our analysis adds to the growing body of work in learning game theory that supports the general principle that to assess the stability of *outcomes* in games, it is important to consider not only how unilateral deviations affect the deviator, but also how they affect the non-deviators. Outcomes where unilateral deviations hurt the deviator (strict Nash) but not the non-deviators (protected[15]) tend to be the most stable. In the particular case of reinforcement learning with fixed aspirations, an additional necessary condition for the stability of an outcome is, of course, that every player finds the outcome satisfactory. Remark: Proposition 4-4 can be

[15] An outcome is protected if unilateral deviations by any player do not hurt any of the other players (Bendor et al., 2001b).

strengthened for the special case where all stimuli are positive (Phansalkar et al., 1994; Sastry et al., 1994).

4.8. Trembling hands process

To study the robustness of the previous asymptotic results we consider an extension of the BM model where players suffer from 'trembling hands' (Selten 1975): after having decided which action to undertake, each player *i* may select the wrong action with some probability $\varepsilon_i > 0$ in each iteration. This noisy feature generates a new stochastic process, namely the *noisy process* $\boldsymbol{N}_n$, which can also be fully characterized by a 2-dimensional vector ***prop*** = [$prop_1$, $prop_2$] of *propensities* (rather than probabilities) to cooperate. Player *i*'s actual probability to cooperate is now $(1 - \varepsilon_i) \cdot prop_i + \varepsilon_i \cdot (1 - prop_i)$, and the profile of propensities ***prop*** evolves after any particular outcome following the rules given by eq. [4-1]. Theorem 2.2 in Norman (1968, p. 67) can be used to prove that this noisy process is ergodic in any 2×2 game[16]. Proposition 4-1 applies to this extension too.

The noisy process has no absorbing states (i.e. SREs) except in the trivial case where both players find one of their actions always satisfactory and the other action always unsatisfactory – thus, for example, in the Prisoner's Dilemma the inclusion of noise precludes the system from convergence to a single state. However, even though noisy processes have no SREs in general, the SREs of the associated unperturbed process (SREUPs, which correspond to mutually satisfactory outcomes) do still act as attractors whose attractive power depends on the magnitude of the noise: *ceteris paribus* the lower the noise the higher the long run chances of finding the system in the neighborhood of an SREUP (see Figure 4-10). This is so because in the proximity of an SREUP, if ε_i are low enough, the SREUP's associated mutually satisfactory outcome will probably occur, and this brings the system even closer to the SREUP. The dynamics of the noisy system will generally be governed also by the other type of attractor, the SCE (see Figure 4-10).

[16] We exclude here the meaningless case where the payoffs for some player are all the same and equal to her aspiration.

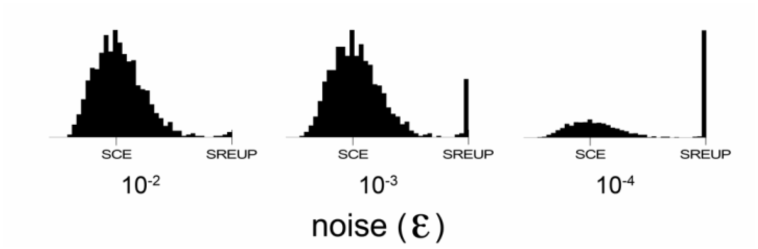


Figure 4-10. Histograms representing the propensity to cooperate for one player (both players' propensities are identical) after 1,000,000 iterations (when the distribution is stable) for different levels of noise ($\varepsilon_i = \varepsilon$) in a Prisoner's Dilemma game parameterised as $[\ 4\ ,\ 3\ ,\ 1\ ,\ 0 \mid 2 \mid 0.25\ ]^2$. Each histogram has been calculated over 1,000 simulation runs.

Figure 4-11 and Figure 4-12, which correspond to a Prisoner's Dilemma game parameterised as $[\ 4\ ,\ 3\ ,\ 1\ ,\ 0 \mid 2 \mid l\ ]^2$, show that the presence of noise can greatly damage the stability of the (unique) SREUP associated to the event CC. Note that the inclusion of noise implies that the probability of an infinite chain of the mutually satisfactory event CC becomes zero.

The systems represented on the left-hand side of Figure 4-11, corresponding to a learning rate $l = 0.5$, show a tendency to be quickly attracted to the state [1 , 1], but the presence of noise breaks the chains of mutually satisfactory CC events from time to time (see the series on the bottom-left corner); unilateral defections make the system escape from the area of the SREUP before going back towards it again and again. The systems represented on the right-hand side of Figure 4-11, corresponding to a lower learning rate ($l = 0.25$) than those on the left, show a tendency to be lingering around the SCE for longer. In these cases, when a unilateral defection breaks a chain of mutually satisfactory events CC and the system leaves the proximity of the state [1 , 1], it usually takes a large number of periods to go back into that area again.

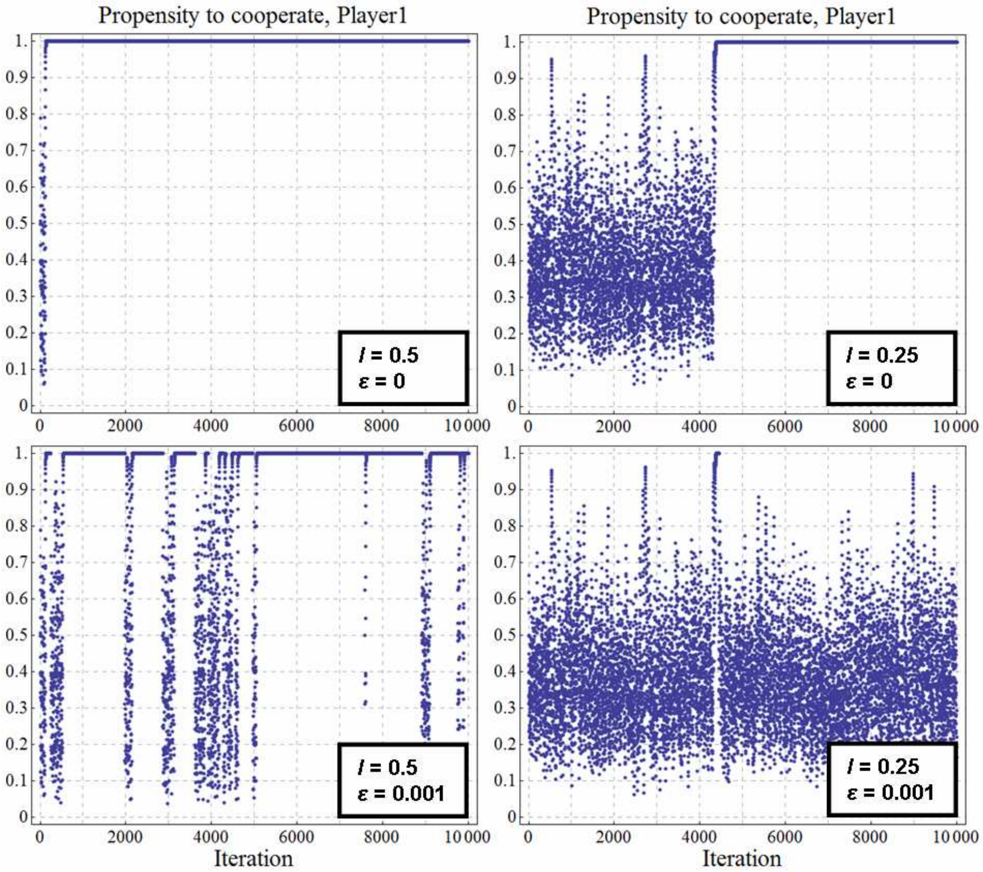


Figure 4-11. Representative time series of player 1's propensity to cooperate over time for the Prisoner's Dilemma game parameterised as [4 , 3 , 1 , 0 | 2 | 0.5 $]^2$ (left) and [4 , 3 , 1 , 0 | 2 | 0.25 $]^2$ (right), with initial conditions [x_0 , x_0] = [0.5 , 0.5], both without noise (top) and with noise level $\varepsilon_i = 10^{-3}$ (bottom).

Figure 4-12 shows that a greater level of noise implies higher destabilisation of the SREUP. This is so because, even in the proximity of the SREUP, the long chains of reinforced CC events needed to stabilise the SREUP become highly unlikely when there are high levels of noise, and unilateral defections (whose probability increases with noise in the proximity of the SREUP) break the stability of the SREUP.

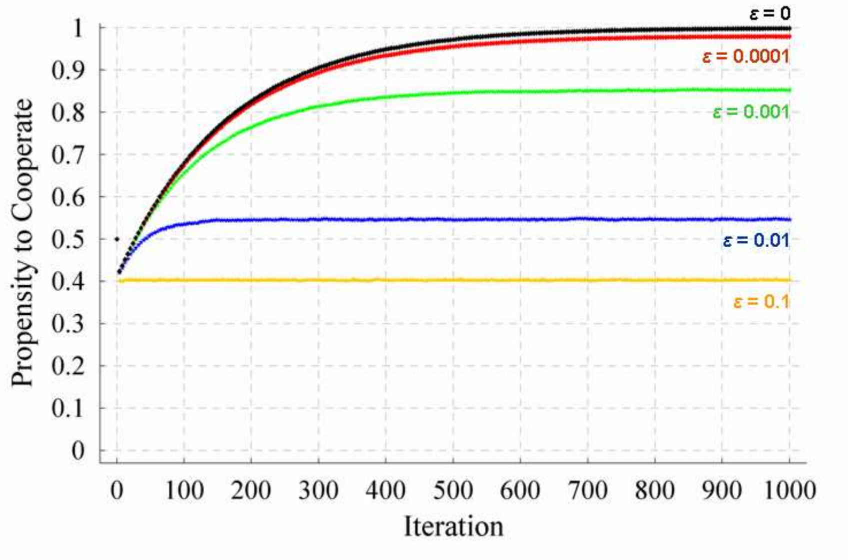


Figure 4-12. Evolution of the average probability / propensity to cooperate of one of the players in a Prisoner's Dilemma game parameterised as [4 , 3 , 1 , 0 | 2 | 0. 5]^2 with initial state [0.5 , 0.5], for different levels of noise ($\varepsilon_i = \varepsilon$). Each series has been calculated averaging over 100,000 simulation runs. The standard error of the represented averages is lower than $3{\cdot}10^{-3}$ in every case.

Stochastic stability

Importantly, not all the SREs of the unperturbed process are equally robust to noise. Consider, for instance, the system [4 , 3 , 1 , 0 | 0.5 | 0. 5]^2, which has two SREs: $[p_{1,C} , p_{2,C}]$ = [1 , 1] and $[p_{1,C} , p_{2,C}]$ = [0 , 0]. Using Proposition 4-2 we know that the set formed by the two SREs is asymptotically reached with probability 1; the probability of the process converging to one particular SRE depends on the initial state; and if the initial state is completely mixed, then the process may converge to either SRE. Simulations of this process show that, in almost every case, the system quickly approaches one of the SREs and then remains in its close vicinity. Looking at the line labelled "ε = 0" in Figure 4-13 we can see that this system with initial state [0.9 , 0.9] has a probability of converging to its SRE at [1 , 1] approximately equal to 0.7, and a probability of converging to its SRE at [0 , 0] approximately equal to 0.3.

However, the inclusion of (even tiny levels of) noise may alter the dynamics of the system dramatically. In general, for low enough levels of "trembling hands" noise we find an ultralong run (invariant) distribution concentrated on neighbourhoods of SREUPs. The lower the noise, the higher the concentration around SREUPs. If there are several SREUPs, the invariant distribution may concentrate on some of these SREUPs much more than on others. In the limit as the noise goes to zero, it is often the case that only some of the SREUPs remain points of concentration. These are called stochastically stable equilibria (Foster and Young, 1990; Young, 1993; Ellison, 2000) and will be discussed in detail in chapter 5. As an example, consider the simulation results shown in Figure 4-13, which clearly suggest that the SRE at [0 , 0] is the only stochastically stable equilibrium even though the unperturbed process converges to the other SRE more frequently with initial conditions [0.9 , 0.9]. Note that whether an equilibrium is stochastically stable or not is independent on the initial conditions.

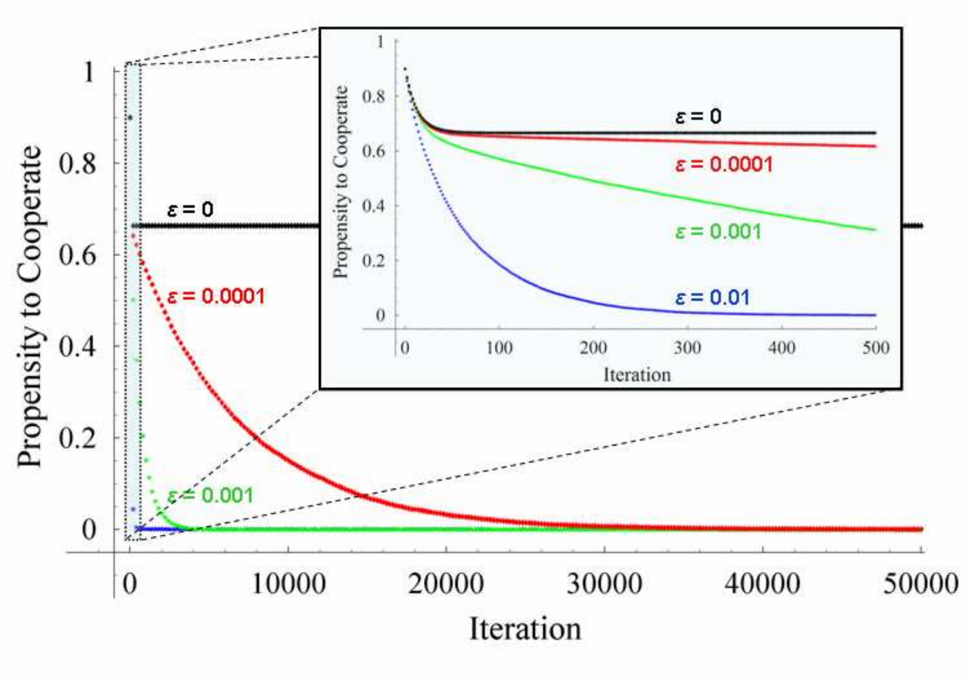


Figure 4-13. Evolution of the average probability / propensity to cooperate of one of the players in a Prisoner's Dilemma game parameterised as [4 , 3 , 1 , 0 | 0.5 | 0. 5 $]^2$ with initial state [0.9 , 0.9], for different levels of noise (ε_i = ε). Each series has been calculated averaging over 10,000 simulation runs. The inset graph is a magnification of the first 500 iterations. The standard error of the represented averages is lower than 0.01 in every case.

Intuitively, note that in the system shown in Figure 4-13, in the proximities of the SRE at [1 , 1], one single (possibly mistaken) defection is enough to lead the system away from it. On the other hand, near the SRE at [0 , 0] one single (possibly mistaken) cooperation will make the system approach this SRE at [0 , 0] even more closely. Only a coordinated mutual cooperation (which is highly unlikely near the SRE at [0 , 0]) will make the system move away from this SRE. This makes the SRE at [0 , 0] much more robust to occasional mistakes made by the players when selecting their strategies than the SRE at [1, 1], as illustrated in Figure 4-14 and Figure 4-15.

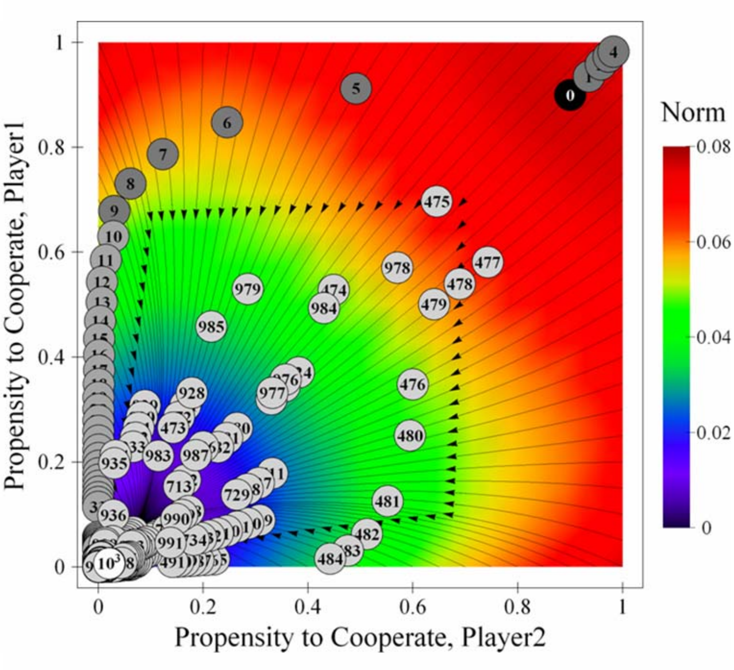


Figure 4-14. One representative run of the system parameterised as [4 , 3 , 1 , 0 | 0.5 | 0. 5]^2 with initial state [0.9 , 0.9], and noise $\varepsilon_i = \varepsilon = 0.1$. This figure shows the evolution of the system in the phase plane of propensities to cooperate, while figure 15 below shows the evolution of player 1's propensity to cooperate over time for the same simulation run.

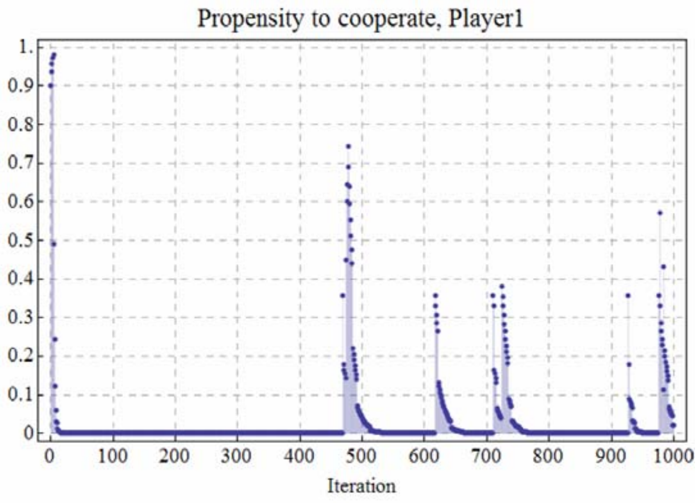


Figure 4-15. Time series of player 1's propensity to cooperate over time for the same simulation run displayed in Figure 4-14.

4.9. Extensions

The theoretical results on asymptotic behaviour presented in this chapter derive from the theory of distance diminishing models developed by Norman (1968; 1972), which can also be applied to 2-player games with any finite number of strategies without losing much generality. The results on transient behaviour when learning takes place by small steps (which derive from the theory of stochastic approximation (Benveniste et al., 1990; Kushner and Yin, 1997) and the theory of slow learning (Norman, 1972)) and Proposition 4-4 (which derives from Sastry et al. (1994)) can be easily extended to any finite game.

More immediately, every proposition in this chapter can be directly applied to finite populations from which two players are randomly[17] drawn repeatedly to play a 2×2 game. Indications on how to prove this are given in Appendix A. As an example, assume that there is a finite population of BM reinforcement learners with aspirations above their respective *maximin* and below their payoff for mutual cooperation, who meet randomly to play a 2×2 social dilemma game (Macy and Flache, 2002). Then, every player in the group will end up cooperating with probability 1 in the ultralong run. The more players in the group, the longer it takes the group to reach universal cooperation.

As for the general existence of SREs and SCEs in games with any finite number of players and strategies, note that both solution concepts require that the expected change in every player's strategy is zero – i.e. they are both critical points of the expected motion of the system. This is an important property since if any system converges to a state, that state must be a critical point of its expected motion. The following shows that every game has at least one such critical point for a very wide range of models. Consider the extensive set of models of normal-form games where every player's strategy is determined at any time-step by the probability of undertaking each of their possible actions. Assume that, after any given outcome $\boldsymbol{y}$ in time step n, every player i (i = 1, 2…m) updates her strategy $\boldsymbol{p}_i$ using an adaptation rule $\boldsymbol{p}_i^{n+1} = \boldsymbol{r}_i^{y}(\boldsymbol{p}^n)$, where $\boldsymbol{r}_i^{y}(\boldsymbol{p}^n)$ is continuous for every $\boldsymbol{y}$

[17] The important point here is that, at any time, every player must have a positive probability of being selected to play the game.

and every *i*. Let us call such adaptation rules continuous. Note that BM adaptation rules are continuous, and consider the following proposition.

Proposition 4-5: Assuming that players' adaptation rules after every possible outcome of the game are continuous, every finite normal-form game has at least one critical point (a strategy profile where the expected change of every player's strategy is zero).

4.10. Conclusions of this chapter

This chapter has focused on the study of games played by individuals who use one of the most widespread forms of learning in nature: reinforcement learning. This analysis (and related literature cited in section 4.1) has shown that the outcome of games played by reinforcement learners can be substantially different from the expected outcomes when the game is played among perfectly rational individuals with common knowledge of rationality. As an example, cooperation in the repeated Prisoner's Dilemma is not only feasible but also the unique asymptotic outcome in many cases. More generally, outcomes where players select dominated strategies can emerge through social interaction and persist through time.

This chapter in particular has characterised the dynamics of the Bush-Mosteller (Bush and Mosteller, 1955) aspiration-based reinforcement learning model in 2x2 games. These dynamics have been shown to depend mainly on three features:

- The speed of learning.
- The existence of self-reinforcing equilibria (SREs). SREs are states which are particularly relevant for the ultralong-run or asymptotic behaviour of the process.
- The existence of self-correcting equilibria (SCEs). SCEs are states which are particularly relevant for the transient behaviour of the process with low learning rates.

With high learning rates, the model approaches its asymptotic behaviour fairly quickly. If there are SREs, such asymptotic dynamics are concentrated on the SREs of the system. With low learning rates, two transient distinct regimes (medium-run and long-run) can usually be distinguished before the system approaches its asymptotic regime. Such transient dynamics are strongly linked to the solutions of the continuous time limit approximation of the system's expected motion.

An extension of the Bush-Mosteller model where players suffer from trembling hands has also been explored. It has been shown that the inclusion of small quantities of noise in the original Bush-Mosteller model can change its dynamics quite dramatically. Some states of the system that are asymptotically reached with high probability in the unperturbed model (i.e. some SREs) can effectively lose all their attractiveness when players make occasional mistakes in selecting their actions. A field for further research is the analytical identification of the asymptotic equilibria of the unperturbed process that are robust to small trembles (i.e. the set of stochastically stable equilibria).

This chapter has characterised not only the asymptotic behaviour of the Bush-Mosteller model of reinforcement learning, but also its transient dynamics. The study of the transient dynamics of learning algorithms has been neglected until recently due to the complexity of its formal analysis. Thus, most of the literature in learning game theory focuses on asymptotic equilibria. This may be insufficient since, as this chapter has illustrated, the transient dynamics of learning algorithms may be substantially different from their asymptotic behaviour. In broader terms, the importance of understanding the transient dynamics of formal models of social interactions is clear: social systems tend to exhibit an impressive ability to adapt and reorganize themselves structurally, meaning that most likely it is not asymptotic behaviour that we observe in the real world.

5. The Implications of Case-Based Reasoning in Strategic Contexts[♣]

Though analogy is often misleading, it is the least misleading thing we have.
−SAMUEL BUTLER

5.1. Introduction

Case-Based Reasoning (CBR) is a form of reasoning by analogy within a particular domain (Aamodt and Plaza, 1994; Nicolov, 1997). 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). Thus, 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 is that if a solution turned out to be satisfactory when applied to a certain problem it might work in a similar situation too.

Case-based reasoners do not employ abstract rules as the basis to make their decisions, but instead 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):

[♣] Some parts of the material presented in this chapter have been published in Izquierdo L.R., Gotts, N.M. and Polhill, J.G. (2004) "Case-based reasoning, social dilemmas, and a new equilibrium concept", *Journal of Artificial Societies and Social Simulation*, 7(3), and in Izquierdo, L.R. and Gotts, N.M. (2005) "The implications of case-based reasoning in strategic contexts", *Lecture Notes in Economics and Mathematical Systems* 564, pp. 163-174.

1. *Retrieve* the most similar case or cases. Generally a case in CBR is rich in information and quite complex. Thus, performing similarity judgements is often an integral part of CBR. Admittedly, the representation of cases used in this chapter is particularly simple and, consequently, similarity judgements are straightforward; this is so because the primary objective of this research is to study the strategic implications of processes of reasoning based on one *single* distinctive past experience (in contrast with rule-based systems), and issues relating case representation are not so crucial for our purposes. The simple representation of cases used here may mean that certain researchers find the reasoning processes investigated in this chapter too unsophisticated to be called CBR; 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, it is worth noting that the term CBR is used in this chapter –in a wider sense than Aamodt and Plaza's– to denote a process of reasoning based on one *single* distinctive past experience, selected for its similarity to the current situation.
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 easy to make 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 (1957) notions of satisficing, as noted by Gilboa and Schmeidler (2001).

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, 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 frequently 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 general rules.

Origins and use of case-based reasoning

CBR arose out of cognitive science research in the late 1970s (Schank and Abelson, 1977; Schank, 1982). Schank and Abelson (1977) proposed that the general knowledge that we gain from experience is encoded in episodic memory as "scripts" that allow us to set up expectations and inferences. New episodes are processed by using dynamic memory structures which contain the episodes that are most closely related to the new episode; this process is called "reminding". Schank (1982) develops the idea that, far from being an irrelevant artefact of memory, reminding is at the root of how we understand and how we learn. Reminding occurs during the normal course of understanding, or processing some new information, as a natural consequence of the processing of that information. He argues that "we understand in terms of what we already understood".

There are several psychological studies that provide support for the importance of CBR as problem-solving process in human reasoning, especially for novel or difficult tasks (see Ross (1989) for a summary). Klein and Calderwood (1988) studied over 400 decisions made by experienced decision makers performing a variety of tasks in operational environments and concluded that "processes involved in retrieving and comparing prior cases are far more important in naturalistic decision making than are the application of abstract principles, rules, or conscious deliberation between alternatives". Drawing on their empirical studies, they also developed a descriptive model of decision making in which the attempt is to satisfice rather than optimise.

More recently, Gayer et al. (2007) have empirically examined the *relative* importance of rule-based versus case-based reasoning in housing asking prices. They hypothesise on theoretical grounds that case-based reasoning has relatively more explanatory power in the rental apartment market, whilst rule-based reasoning is relatively more prevalent in the sales market, and they find empirical support for this hypothesis when tested with two databases (rentals and sales) of asking prices on apartments in the greater Tel-Aviv area. However, their interpretation of case-based reasoning is significantly different from that explained above. In their model, case-based reasoning is modelled using a similarity-weighted average that makes use of *all* cases available at the time of making a decision. In general terms, they conjecture that, in comparison to rule-based reasoning, case-based reasoning will be more prevalent in non-speculative markets than in speculative ones. They also state their belief that both modes of reasoning are likely to play a role in almost any decision-making process, and that a variety of factors may affect their relative importance.

It seems therefore that CBR is 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 Kahneman et al., 1982).

There are 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.

Within the domain of theoretical economics, a Case-Based Decision Theory (CBDT) has been proposed by Gilboa and Schmeidler (1995; 2001). CBDT is 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". As pointed out when describing the empirical study conducted by Gayer et al. (2007), 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. For the formalisation of an assessment rule based on such a similarity-weighted function see Gilboa et al. (2006). In any case, Gilboa and Schmeidler (1995; 2001) 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). They also highlight that CBDT naturally gives rise to the notions of satisficing decisions and aspiration levels.

Pazgal (1997) and Kim (1999) apply CBDT in strategic contexts. Pazgal (1997) analyses general games of mutual interest (i.e. games where there exists a unique pure strategy profile that gives the highest possible payoff to every player), and Kim (1999) focuses on symmetric 2x2 games of mutual interest to study the aspiration updating mechanism in greater depth[18]. The decision-making algorithm employed by players in these two studies bears very little resemblance to CBR as interpreted above: players in Pazgal's and Kim's models do not consider different cases or experiences, they choose the action that has given them the highest cumulative past payoff (relative to their current aspiration) throughout the whole history of the game, and their aspiration thresholds are updated using a weighted average of its previous value and an average function of received payoffs. This

[18] Kim (1999) studies 2x2 games with an outcome (i.e. a *pure strategy profile*) which every player strictly prefers, and refers to these as "common interest" games. Following Aumann and Sorin (1989), I use the term "common interest game" to denote the wider class of games where there is a unique *payoff profile* that strongly Pareto dominates all other payoff profiles (and this payoff profile may be achieved via several strategy profiles), and I use the more specific term "mutual interest game" to denote games where there exists a unique *pure strategy profile* that gives the highest possible payoff to every player.

decision-making algorithm (identified by the authors as a form of case-based maximisation) is significantly different from that consisting in maximising average payoffs (as nicely illustrated by Kim (1999)), but it is also fundamentally different from CBR as interpreted in this chapter. As a matter of fact, it seems to us that the essence of these two models is closer to reinforcement learning than to case-based reasoning, as also noted by Bendor et al. (2001a; 2001b).

To our knowledge, the implications of CBR interpreted as explained above in strategic contexts had never been formally explored up until now. In this chapter we develop and analyse a game theoretical model where individuals use a very simple form of CBR.

Structure of this chapter

In this chapter we use social dilemma games to illustrate the strategic implications of case-based reasoning. The following section is devoted to explaining why social dilemmas in particular are especially revealing to understand the differences between reasoning by cases and reasoning by rules. Section 5.3 presents a simple model that is used to shed light on the conditions under which CBR as individual decision mechanism may entail cooperation in social dilemmas. The results obtained with this model are presented and discussed in sections 5.4 and 5.5 respectively. Section 5.6 presents a generalisation of the model analysed in sections 5.4 and 5.5. In particular, players in the more general model may make occasional mistakes in their decisions. The dynamics of this second model are explained and discussed in 5.7. Finally, section 5.8 presents the conclusions of this chapter.

5.2. Case-based reasoning and social dilemmas

This chapter provides various results on the *asymptotic* dynamics of a rather general form of CBR for any finite normal-form game (see section 5.7). The *transient* dynamics of CBR models, however, strongly depend on the definition of the particular CBR algorithm employed by players and also on the specific game they play. Thus, to explore the whole dynamics of games played by agents who use a simple form of CBR, the scope of study has had to be limited to some extent. In particular, whenever it has been found that the specific parameterisation of the game has made a difference I have focused on analysing social dilemmas.

Social dilemmas offer a promising arena to distinguish the differences between reasoning by cases (or outcomes[19]) and reasoning by rules (or strategies). The following illustrates why this is the case using the Prisoner's Dilemma. 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[20], and b) rational players do not choose dominated strategies[21]. Similarly, even though defining rational outcomes cannot be done without controversy, it also seems sensible to agree that rational outcomes must be Pareto optimal[22]. Assuming only those necessary conditions for the rationality of strategies and outcomes, we can state that in the one-shot Prisoner's Dilemma and other social dilemmas, even though there is a clear causal link between strategies and outcomes, rational strategies (understood as those chosen by rational players) lead to outcomes that are not rational, whereas rational outcomes are generated by strategies that are not rational (i.e. those strategies that a rational player would never select).

In this chapter we explore two social dilemma games: a 2-player and an *n*-player version of the Prisoner's Dilemma (PD). Because of the players' decision making algorithms (explained in sections 5.3 and 5.7), the actual values of the payoffs are not relevant as long as they satisfy:

$$\textit{Temptation} > \textit{Reward} > \textit{Punishment} > \textit{Sucker}$$

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

[20] Recall that, for a player A, strategy S_A is (strictly) dominant if for each 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 (Gibbons, 1992, p. 5).

[21] For a player A, strategy S_A is (strictly) dominated by strategy S^*_A if for each 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).

[22] An outcome is Pareto optimal if there is no other outcome where at least one player is better off and no player is worse off.

In the n-player social dilemma every player gets a reward as long as there are no more than M defectors ($M < n$). The payoff that defectors get is always higher than the payoff obtained by those who cooperate (*Def-P* > *Coop-P*). However, every player is better off if they all cooperate than if they all defect (*Coop-P* + *Reward-P* > *Def-P*). Figure 5-1 shows the payoff matrix for a particular player:

	Fewer than *M* others defect	***M* others defect**	**More than *M* others defect**
Player cooperates	***Coop-P + Reward-P***	***Coop-P + Reward-P***	***Coop-P***
Player defects	***Def-P + Reward-P***	***Def-P***	***Def-P***

Figure 5-1. 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, 2001) after the influential paper written by Hardin (1968). Henceforth we will refer to this game as the TC game. When the maximum number of defectors M for which the reward is given is high, it represents a version of the "volunteer's dilemma" (Brenan and Lomasky, 1984; Diekmann, 1985): 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 M 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 that result in a deficient equilibrium"[23]. In any case, we have a "problematic social situation" (Diekmann, 1986; Raub and Voss, 1986), or social dilemma in a broader sense, which can be defined in game theory terms as a game with Pareto inefficient[24] Nash equilibria. The TC game differs from the two-player PD in three important ways:

1. In the TC game, for a small number of players, the state of "minimally effective cooperation" (exactly M defectors) is not negligible, so there is not a dominant strategy.

[23] An equilibrium is deficient if there exists another outcome which is preferred by every player.

[24] An outcome is Pareto inefficient if there is an alternative in which at least one player is better off and no player is worse off.

2. In the TC game, using pure strategies, there are two Nash equilibria: everyone defecting (universal defection[25]) and exactly *M* defectors (minimally effective cooperation).
3. In the two-player PD, universal cooperation is a Pareto optimal outcome since no player can be better off without making the other player worse off. However, in the TC game the only Pareto optimal outcome is the state of minimally effective cooperation.

5.3. The CBR model

In this section we present a simple CBR decision-making algorithm that players will use to decide whether to cooperate or not when confronted with one of the two social dilemma games described in the previous section. This model will be named "the CBR model". Individuals play repeatedly the game – once per time-step – and every time they do so, each player retains a case (representing the experience they lived in time-step *t*) which comprises:

1. The time-step *t* when the case occurred.
2. The *perceived* state of the world at the beginning of time-step *t*, characterised by the value of the following descriptors in the preceding *ml* (for *memory length*) time-steps:
    - Descriptor 1 (D1): the number of other defectors.
    - Descriptor 2 (D2): the decision that the player holding the case made.

   As an example, if $ml = 2$ then the perceived state of the world for the case-holder will be determined by the number of other defectors and the decision she made, both in time-step $t - 1$ and in time-step $t - 2$).
3. The decision the case-holder made in that situation, *i.e.* whether she cooperated or defected in time-step *t*, having observed the state of the world in that same time-step.
4. The payoff that the case-holder obtained after having decided in time-step *t*.

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

[25] Universal defection is a Nash equilibrium as long as $M < n$-1.

t	$df_{t-ml} \ldots df_{t-2} \quad df_{t-1}$	d_t	p_t
	$d_{t-ml} \ldots d_{t-2} \quad d_{t-1}$		

where

df_t is the number of defectors (excluding player A) in time-step t,

d_t is the decision made by player A in time-step t, and

p_t is the payoff obtained by player A in time-step t.

The number of cases that players can keep in memory is unlimited. It is also worth noting that no cases are available for any player until (ml + 1) time-steps have gone by in the simulation. Players 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 possible decisions (*i.e.* each of the two possible values of d_t). 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 player holding the case. 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 particular situation (*i.e.* for a given perceived state of the world) a player must face one of the following three possibilities:

1. The player cannot recall any previous situations that match the current perceived state of the world. In CBR terms, the Agent does not hold any appropriate cases for the current perceived state of the world. In this situation the player will decide at random.
2. The player does not remember a previous similar situation when she made a certain decision, but she does recall at least one similar situation when she made the other decision. In CBR terms, all the appropriate cases the player recalls have the same value for d_t. In this situation, the player will explore the non-applied decision if the payoff she obtained in the last previous similar situation was below her *Aspiration Threshold AT*; otherwise she will keep the same decision she previously applied in similar situations.
3. The player remembers at least one previous similar situation when she made each of the two possible decisions. In this situation, the player will focus on the most recent case for each of the two decisions and choose the decision

that provided her with the higher payoff[26]. In this way, players adapt their behaviour according to the most recent feedback they got in a similar situation.

This completes the specifications of "the CBR model". The UML activity diagram of the players' decision making algorithm is outlined in Figure 5-2. In the simulation experiments reported in this chapter, all the players share the same aspiration threshold *AT* and the same memory length *ml*. These are the two crucial parameters in the CBR model, determining when an outcome is satisfactory and when two situations are similar, respectively. The behaviour of a slightly more advanced socioeconomic Agent which also uses CBR in their decision-making algorithm but takes into account social approval is explored in Izquierdo *et al.* (2003).

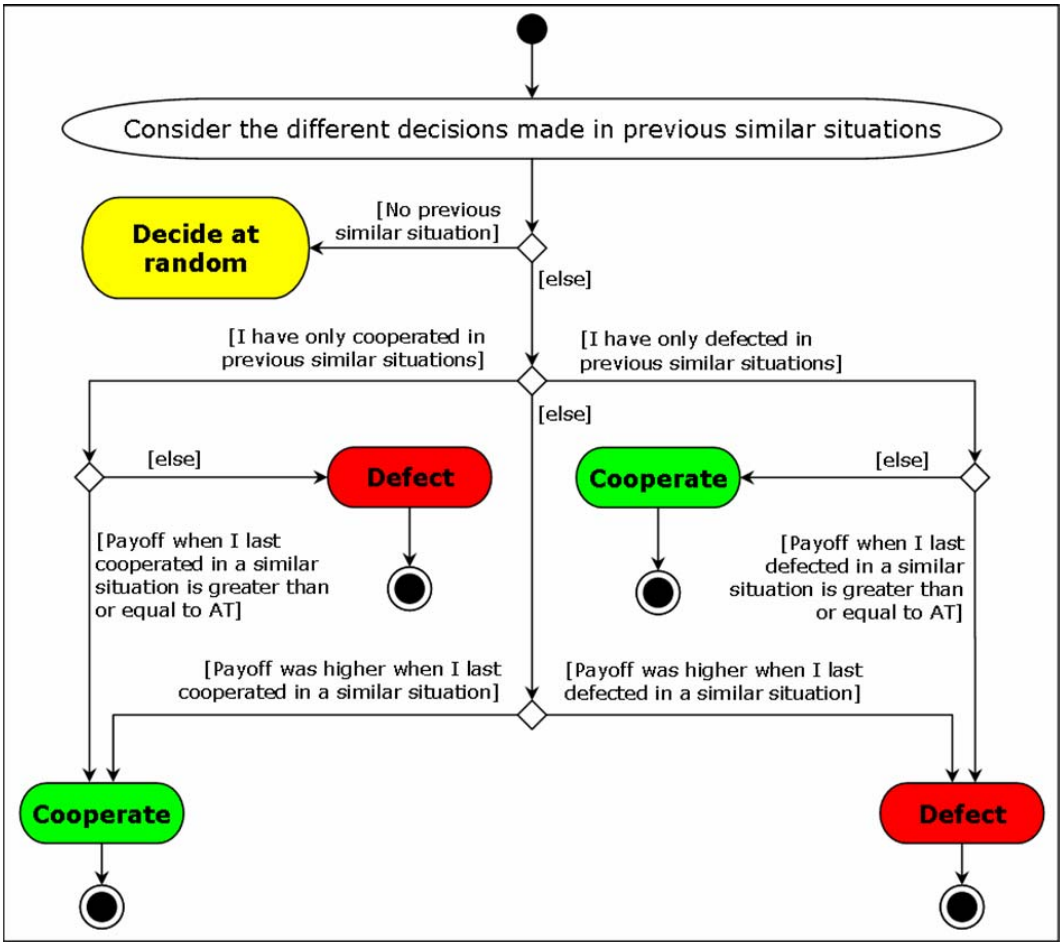

Consider the different decisions made in previous similar situations
[No previous similar situation]
Decide at random
[else]
[I have only cooperated in previous similar situations]
[I have only defected in previous similar situations]
[else]
[else]
Defect
Cooperate
[else]
[Payoff when I last cooperated in a similar situation is greater than or equal to AT]
[Payoff when I last defected in a similar situation is greater than or equal to AT]
[Payoff was higher when I last cooperated in a similar situation]
[Payoff was higher when I last defected in a similar situation]
Cooperate
Defect

Figure 5-2. UML activity diagram of the CBR decision making algorithm.

[26] A tie is impossible in either of the two games analysed in this chapter.

5.4. Results with the CBR model

The software used to conduct the experiments reported in this section was written in Objective-C using the Swarm libraries (http://www.swarm.org) and is available in the Supporting Material together with a user guide under the 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 CBR model is very sensitive to the decisions that players make 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 only as a range of possible outcomes, each with a certain probability of happening. The probability of each outcome can be either estimated by running the model several times with different random seeds or, under certain circumstances, exactly computed.

Players in the CBR 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 a player can perceive is finite, so is the number of random decisions the player can make. Therefore, simulations must end up in a cycle. To study how often players cooperate in the PD we define the 'cooperation rate' as the number of times bilateral cooperation is observed in a cycle divided by the length of the cycle. Similarly, we define the 'reward rate' in the TC game as the number of times the reward is given in a cycle divided by the length of the cycle.

5.4.1. Prisoner's Dilemma

Aspiration Thresholds

It is important to realise that when players play the PD, they share the same perception of the state of the world (defined by the last *ml* moves of the two Players) in the sense that any two situations that look the same to one player will also look the same to the other player and any two situations that look different to one player will also look different to the other player. Therefore, at any given time in the simulation our players 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.

The decision dynamics for a certain state of the world are summarised in Table 5-1. Consider for now the first four rows of the table ($T < AT$). These represent the case where the aspiration threshold AT (for both players) exceeds T. The first time any particular state of the world occurs, both players 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 players are cooperating or both players 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.

Aspiration Thresholds (AT)	1st visit (random)	2nd visit	3rd visit	4th visit and onwards	x	y
$T < AT$	CC	DD	CC	CC	1	-
	CD	DC	DD	DD	-	2
	DC	CD	DD	DD	-	2
	DD	CC	CC	CC	1	-
$R < AT \leq T$	CC	**DD**	CC	CC	1	-
	CD	**DD**	DC	DD	-	2
	DC	**DD**	CD	DD	-	2
	DD	**CC**	CC	CC	1	-
$P < AT \leq R$	CC	CC	**CC**	CC	0	-
	CD	DD	**DC**	DD	-	2
	DC	DD	**CD**	DD	-	2
	DD	CC	**CC**	CC	1	-
$S < AT \leq P$	CC	**CC**	CC	CC	0	-
	CD	**DD**	DD	DD	-	1
	DC	**DD**	DD	DD	-	1
	DD	**DD**	DD	DD	-	0
$AT \leq S$	CC	CC	CC	CC	0	-
	CD	CD	CD	CD	-	-
	DC	DC	DC	DC	-	-
	DD	DD	DD	DD	-	0

Table 5-1. Decisions made by each of the two players in 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 player, the second letter to those of the other. C is cooperation and D is defection. The first imbalance between CC and DD for every value of AT has been shadowed. The meaning of x and y is explained in the text. The results shown in this table are independent of the memory length.

There are two states of the world that appear to be particularly important in the dynamics of the game. One is that where there have been *ml* successive bilateral cooperations (let us call it *mlBC*); the other is where there have been *ml* successive bilateral defections (let us call it *mlBD*). Whenever bilateral cooperation follows a visit to *mlBC*, then *mlBC* is immediately revisited (since players observe again that they both cooperated in the last *ml* time-steps). Similarly, whenever bilateral defection follows a visit to *mlBD*, then *mlBD* is immediately revisited (since players observe again that they both defected in the last *ml* time-steps). We can then define *x* as the number of times that *mlBC* has to be revisited after it has been abandoned before stable cooperation is reached, and *y* as the number of times that *mlBD* has to be revisited after it has been abandoned before stable defection is reached. As an example, when *AT* > *T*, if both players happen to cooperate when they observe *mlBC* for the first time, then they will both experience *mlBC* for the second time in the following time-step. Both of them will then defect (2nd visit to *mlBC*), and in doing so will abandon *mlBC*. If *mlBC* is then revisited (3rd visit), it will never be left again. In this hypothetical example, the number of times *x* that *mlBC* had to be revisited after it was abandoned before stable cooperation was reached was 1. This information is included in Table 5-1 and its significance will be explained later.

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 5-1. 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 (both players cooperate or defect at the same time), since if a player receives the S*ucker* payoff in any situation, they will never cooperate again in that situation. In this sense our players are particularly unforgiving. Players with aspiration thresholds greater than *Sucker* cannot be systemically exploited. The importance of this will be discussed later.
- When *AT* > *Punishment*, there is a qualitative jump in terms of average cooperation rates. This is because if *AT* > *Punishment*, when both Players

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* ≤ *Punishment*.

Taking into account the two previous points and looking at the "4th visit and onwards" column in Table 5-1, 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 5-3 shows the importance of aspiration thresholds and how they can modify the effect of the memory length.

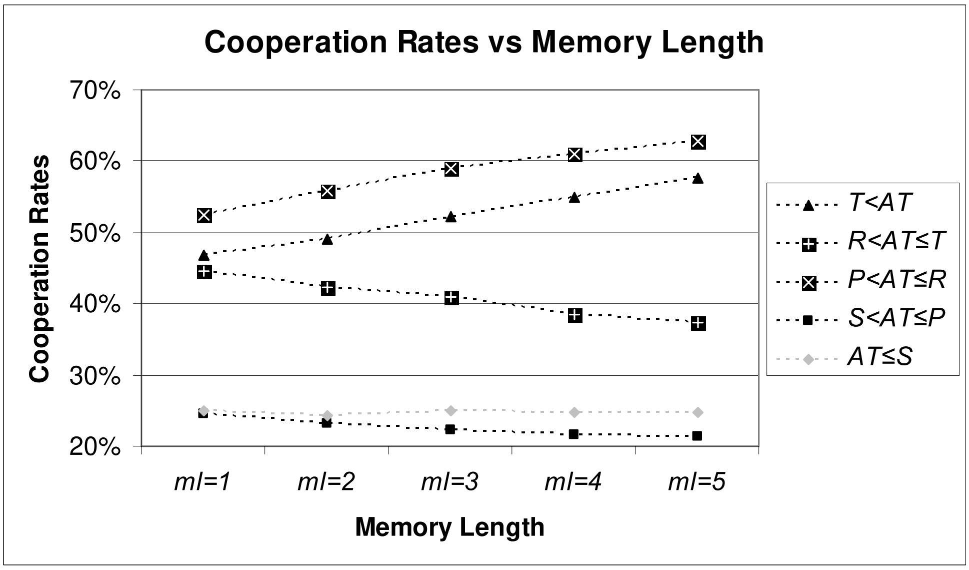


Figure 5-3. Average cooperation rates when modelling two players with Memory Length *ml* and Aspiration Threshold *AT*, playing the PD. The average cooperation rate shows the probability of finding both Players cooperating once they have finished the learning period (*i.e.* when the run locks in to a 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 %.

The interactions between the aspiration threshold and the memory length can be explained by taking into account two factors. Both factors are related to the fact that, as the memory length increases, the number of possible perceived states of the world grows exponentially and it becomes less likely for any given state of the world to be revisited. From now on let us refer to each payoff by its initial letter.

1. The first factor concerns only the relative frequency of stable bilateral cooperation and stable universal defection[27]. This factor is present for any $AT > S$ and represents a bias towards cooperation. Looking at Table 5-1, one could expect stable bilateral defection to be three times more likely than stable bilateral cooperation if $S < AT \leq P$, and as likely as stable bilateral cooperation if $AT > P$. However, as the memory length increases, there is a certain bias towards stable bilateral cooperation. For the simulation to lock in to stable bilateral cooperation, it is required that a bilateral decision (a bilateral cooperation if $S < AT \leq P$) follows the first visit to the state of the world formed by *ml* bilateral cooperations (*mlBC*) and that the same state of the world *mlBC* is revisited *x* more times after it is abandoned; similarly, stable bilateral defection requires a unilateral decision (or bilateral defection if $S < AT \leq P$) following the first visit to the state of the world formed by *ml* bilateral defections (*mlBD*) and *y* more visits to that state of the world *mlBD* after it is abandoned. As we can see in Table 5-1, except for the trivial case[28] where $AT \leq S$, the average *x* is always less than the average *y* for any given aspiration threshold. For high values of the memory length, revisiting a state can take a very long time and the fact that stable bilateral cooperation needs fewer visits (*x*) to settle down than stable bilateral defection does (*y*) is an important bias towards the frequency of stable bilateral cooperation.

2. The second factor explains why average cooperation rates not only fail to increase, but actually decrease with memory length for $S < AT \leq P$ and $R < AT \leq T$. This factor is present for $S < AT \leq T$ and it represents a bias towards cooperation if $P < AT \leq R$, and a bias towards defection if $S < AT \leq P$ or $R < AT \leq T$. For any $AT > S$, the simulation ends up in a cycle of bilateral decisions. Therefore, it is crucial to study whether there is a bias towards cooperative bilateral decisions (CC) or towards defective bilateral decisions (DD) in the players' learning process. Table 5-1 shows the history of decisions made by the players having observed any particular state of the world for different aspiration thresholds. The first imbalance between CC and DD for every value of *AT* has

[27] This effect is explained in detail by Izquierdo et al. (2003).

[28] If the Aspiration Threshold does not exceed *Sucker*, Agents repeat the same decision that they made at random the first time they visited a certain state of the world whenever they visit the same state again.

been shadowed (*e.g.* if $S < AT \leq P$ the first imbalance occurs in the second visit, where DD is three times more likely to happen than CC). Imbalances in the earlier visits to a state of the world are more important because those in later stages might never materialise if a cycle is reached before they can occur. Imbalances in the component parts of the state of the world (CC and DD) make certain states of the world more likely to occur than others, hence leading to biases in the cooperation rate. What is not obvious is why the importance of such imbalances (in terms of reward rates) increases with the value of memory length. This is so because, even ignoring the fact that some states of the world are more likely to occur than others, not all states of the world are equally likely to form part of a cycle; some states can form cycles more easily than others[29], and their relative frequency depends on the memory length. This is certainly the case for *mlBC* and *mlBD*. Not only are they the only states of the world that can form cycles just by themselves (assuming $AT > S$), but they also need fewer revisits to settle than the rest of the possible states of the world (see previous paragraph). Roughly half of the simulation runs reported in this paper with $AT > S$ ended up in cycles made up by either *mlBC* or *mlBD*. This means that an imbalance between the frequency of *mlBC* and *mlBD* can affect the reward rate substantially. The imbalance between *mlBC* and *mlBD* given an imbalance between CC and DD does depend on the memory length. To clarify this, assume that DD is always *z* times more likely than CC; then *mlBD* will be z^{ml} times more likely than *mlBC*. This analysis is not a proof since successive states of the world are not independent, but it clarifies why imbalances gain importance as the value of the memory length increases. As we can see in Table 5-1, if $S < AT \leq P$ or $R < AT \leq T$, the imbalance is towards the defective bilateral decision, making *mlBD* more likely to occur relative to *mlBC* as memory length increases, and thus reducing the average cooperation rate. On the other hand, if $P < AT \leq R$, the imbalance is towards cooperation.

The summary of the effect of each of the two factors depending on the *AT* outlined above is shown in Table 5-2, together with the total effect found in the simulations. We have not yet proved that the two effects explained here are the only operating factors.

[29] Or, conversely, some cycles comprise fewer different states of the world than others.

	$AT \leq S$	$S < AT \leq P$	$P < AT \leq R$	$R < AT \leq T$	$T < AT$
Effect of factor 1	No bias	Bias towards cooperation	Bias towards cooperation	Bias towards cooperation	Bias towards cooperation
Effect of factor 2	No bias	Bias towards defection	Bias towards cooperation	Bias towards defection	No bias
…	…	…	…	…	…
Total effect	No bias	Bias towards defection	Bias towards cooperation	Bias towards defection	Bias towards cooperation

Table 5-2. Effect on average cooperation rates of each of the two factors outlined in the text above depending on the value of AT, and results from the simulation runs.

It is clear that in CBR, not only *what* is learnt, but the actual *process* of learning can be of 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, players 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 5-3).

Importance of a common perception of the state of the world

To study the importance of having a shared perception of the state of the world in the PD, we studied the outcome of the game when played by players with partial representations of the state of world: players who only look at the other player's actions (only descriptor D1) and players who only look at their own actions (only descriptor D2). In both these cases, the two players may perceive the state of the world differently. Figure 5-4 shows the results obtained for $AT > T$. The results for other aspiration thresholds are very similar[30] so they are omitted.

[30] Except, again, for the trivial case where $AT \leq S$, in which the average cooperation rate is always 25%.

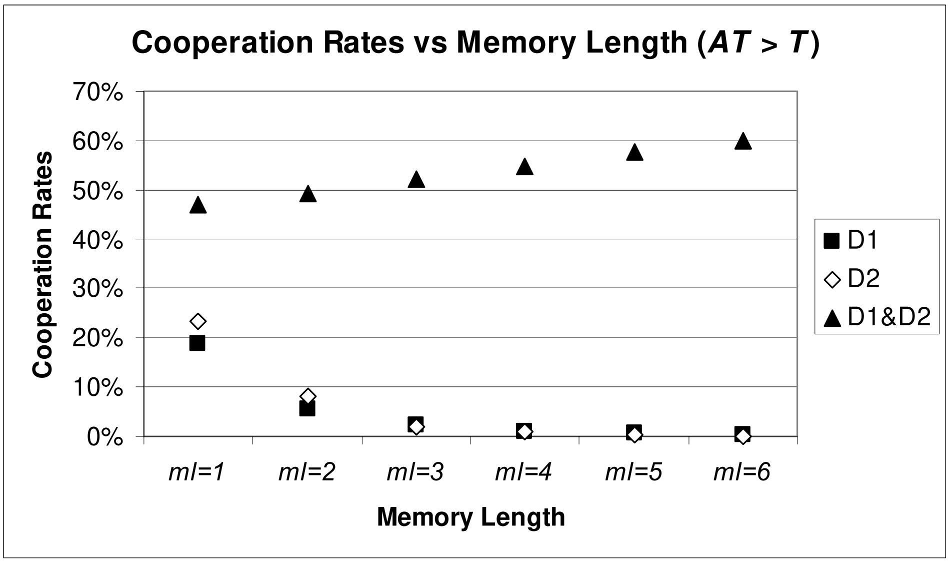


Figure 5-4. Average cooperation rates when modelling two players with Memory Length *ml*, Aspiration Threshold greater than *Temptation*, and with 3 different representations of the state of the world (D1, D2, and D1&D2), playing the PD. The values represented for *ml* = 1 have been computed exactly. The rest of the values have been estimated by running the model 10,000 times (*ml* = 2, 3, 4) or 1,000 times (*ml* = 5, 6) with different random seeds. All standard errors are less than 1%.

The difference in terms of average 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 memory length *ml* is. When both the player's own decisions and the other player's decisions form the perceived state of the world (D1&D2) the average cooperation rate is much higher than in the other cases.

As we saw in Table 5-1, except in the trivial case where $AT \leq S$, players will never cooperate again in a given state of the world after having received the *Sucker* payoff in that state. When using either of the two incomplete perceptions of the state of the world, there are sets of situations that are represented by the same perceived state of the world for one player but by different perceived states of the world for the other. The size of such sets of situations increases as the memory length *ml* increases. In these sets of situations, one of the players will make several decisions at random in situations which they perceive as novel, but which are represented by one single perceived state of the world for the other player.

This fact strongly increases the chances of the latter player getting a *Sucker* payoff and therefore not achieving a cooperative outcome.

5.4.2. The Tragedy of the Commons game

Aspiration Thresholds

The TC game is more complex to analyse than the PD since at any given time in the simulation players have not necessarily visited what they perceive as a distinct situation the same number of times[31]. Therefore, in a given time-step some players may be making decisions at random while some others may not. This means that we cannot build a table like Table 5-1 for the TC game.

Figure 5-5 shows the results obtained in the TC game when played by 10 players with memory length *ml* = 1, for different values of *M* (maximum number of defectors for which the reward is given). Similar results have been obtained when the game is played by 5 and by 25 players.

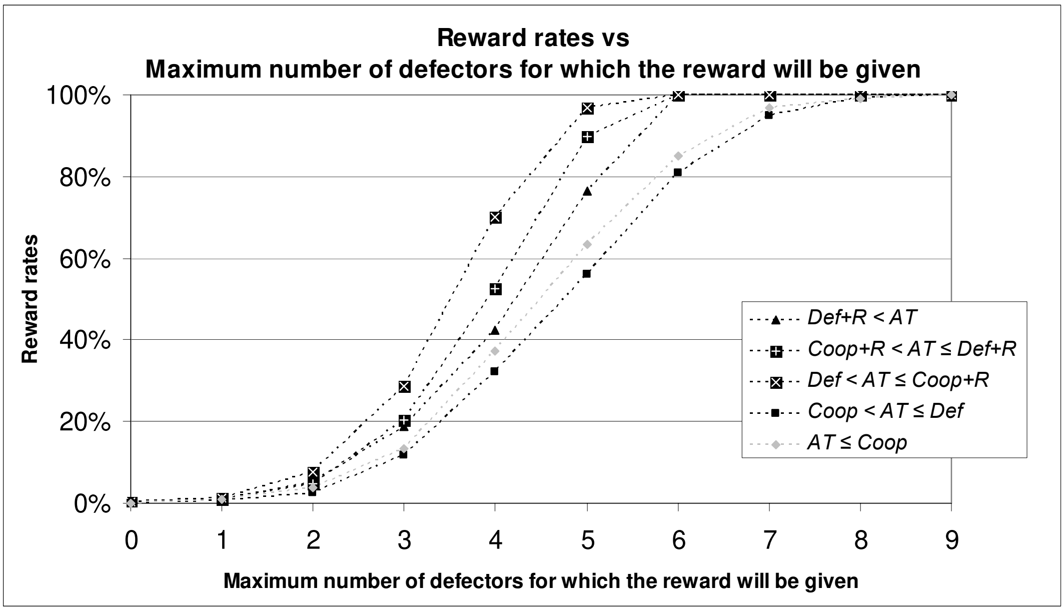


Figure 5-5. Average reward rates for different values of *M* in the Tragedy of the Commons game played by 10 Players with Memory Length *ml* = 1. Each represented value has been estimated by running the model 1,000 times. All standard errors are less than 1.5%.

[31] Recall that players know only whether they cooperated or defected, and *how many* others defected. In the TC game, the information provided to the players is thus not complete in the sense that they cannot identify who is defecting, as they could in the PD (since there was only one other player).

Figure 5-5 shows that levels of cooperation strongly depend on the maximum number of defectors for which the reward is given (*M*). When the requirement is too demanding (low values of *M*), levels of cooperation tend to be low and the reward is not usually given. On the other hand, for moderate and high values of *M* ($M \geq 6$), the reward is almost always given[32]. If players have aspiration thresholds greater than *Def-P* then the reward will be given more often than if they choose at random ($AT \leq$ *Coop-P*). The highest levels of cooperation are achieved when the aspiration thresholds are just above *Def-P*. Levels of cooperation then decrease as aspiration thresholds separate from the optimal value.

Importance of a common perception of the state of the world

To test the importance of a common perception of the state of the world, we put our players on a toroidal 2x5 grid so they could only observe their most immediate five neighbours in their Moore neighbourhood of radius 1. Results are shown in Figure 5-6.

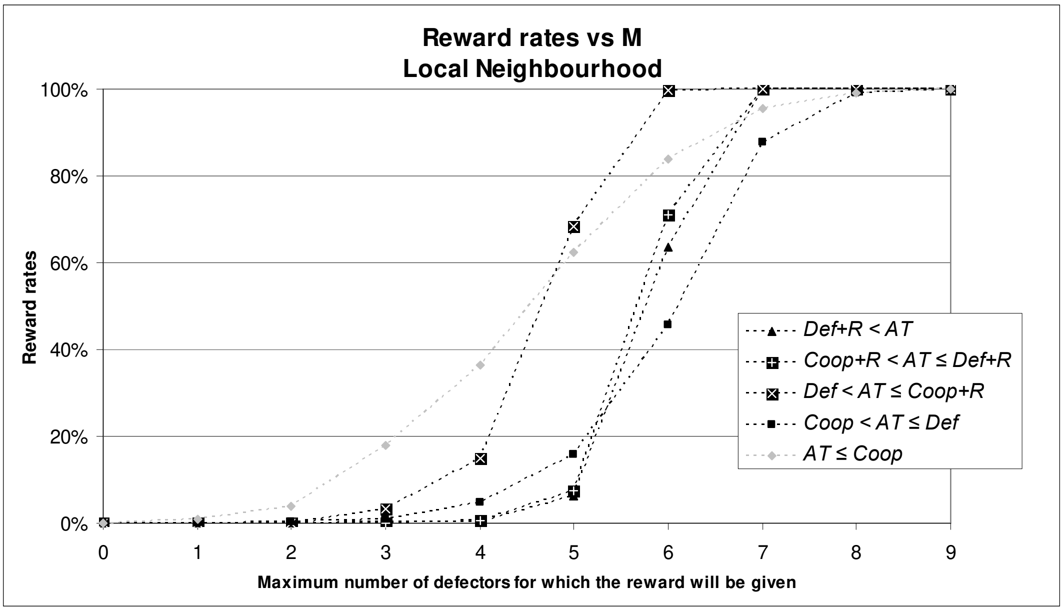


Figure 5-6. Average reward rates for different values of *M* in the Tragedy of the Commons game played by 10 Players with Memory Length $ml = 1$. Every player *A* can observe other 5 players only, who are the only ones that can observe player *A*. Each represented value has been estimated by running the model 1,000 times. All standard errors are less than 1.5%.

[32] When the game is played by 25 Agents, average reward rates are greater than 80% if $M \geq 15$ and greater than 99% if $M \geq 19$, for any aspiration threshold.

When players can observe only their local neighbourhood, the range of values of *M* to which the reward rate is sensitive is shifted and squeezed to the right. The use of local neighbourhoods sharpens the global movement from defection to cooperation. When players can only observe their neighbours, their global response to changes in the reward programme (parameterised by *M*) is not smooth anymore. Instead, the global behaviour is now better characterised by a hard threshold whose particular value depends on the aspiration threshold of the players forming the society. When players can only observe their neighbours there is a very narrow range of values for *M* where a very small change can make a huge difference.

As in the previous case, the highest levels of cooperation are achieved when the aspiration thresholds are just above *Def-P*. It is once again 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.

5.5. Discussion of the results obtained with the CBR model

The experiments conducted with the CBR model show that cooperation can emerge from the interaction of selfish and unforgiving (but satisficing) case-based reasoners. We are aware that the assumption that Agents make their decisions at random when confronted with a new situation is difficult to maintain. However, Table 5-1 shows that when *AT* > *Maximin*[13], any positive correlation between the random decisions taken by the Agents will tend to increase levels of cooperation. Similarly, we would expect negative correlations to lead to less cooperative outcomes. The experiments have also shown that the optimal value of the aspiration threshold is just above *Maximin*, and that sharing a common perception of the state of the world strongly increases levels of cooperation.

More importantly, the experiments conducted have revealed a concept of equilibrium which is more relevant than the Nash equilibrium for repeated games played by case-based reasoners: *strictly undominated outcomes* (or individually-rational outcomes). The concept of strictly undominated outcome is defined for one single stage of any game. Its defining property is that no player can be guaranteed a higher payoff by changing their decision[33] (*i.e.* every player is getting at least their *Maximin*). The concept of strictly undominated outcome is weaker (*i.e.* less restrictive) than the Nash equilibrium: A Nash equilibrium is always a strictly undominated outcome but the reverse is not necessarily true. In particular, in the one-shot PD, bilateral cooperation is a strictly undominated outcome 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 strictly undominated outcome accounts for every possible action that the other players might take. A strictly undominated outcome as equilibrium concept 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 other players' response, then the player has a clear incentive to deviate from that outcome, so that outcome cannot be an equilibrium (it is strictly dominated by other outcomes). If, on the contrary, no player has such incentive, the outcome could be an equilibrium. It comes as no surprise that this equilibrium concept is based on outcomes rather than strategies, since case-based reasoners place the emphasis on the case rather than on the rule.

In the PD, the only strictly undominated outcomes are the two bilateral decisions. In the TC the only strictly undominated outcome in which the reward is not given is universal defection; all the outcomes in which the reward is given are strictly undominated.

It can be mathematically shown that all the non-trivial simulations (i.e. those where aspiration thresholds are above the lowest payoff) reported in this chapter must end up in cycles made up of strictly undominated outcomes (Izquierdo et al.,

[33] A slightly more restrictive concept is that of an undominated outcome, in which no player can be guaranteed the same or a higher payoff by changing their decision. The concept of undominated outcome as equilibrium implies that players deviate from an outcome only if it is certain that they will not be worse off by doing so, whereas the *strictly* undominated concept implies that players move away from an outcome only if it is certain that they will be better off by doing so. The concept of undominated outcome as equilibrium is neither weaker nor stronger than the Nash equilibrium.

2004). As we have seen in the previous section, the actual selection among different strictly undominated outcomes can be strongly path-dependent and depends on the specific type of CBR algorithm that players use.

If their aspiration threshold is high enough, players in the CBR model will not accept outcomes in which they are guaranteed a higher payoff by changing their decision once their learning process is finished. However, they are quite naive in the sense that they are not able to infer that the game has locked in to a persistent cycle. In other words, they are not able to infer that the other players will not accept outcomes where they are not getting their *Maximin* either. We can conjecture what would happen if the players were sophisticated enough as to infer, through repeated interaction and learning, the *fact* that the rest of the players are also non-exploitable (*i.e.* they do not accept outcomes where they get a payoff lower than *Maximin*). Assuming (or learning) that the rest of the players are not exploitable can then enable a player X to infer that certain outcomes which give payoffs higher than *Maximin* to this player X will not be sustainable (because they do not yield payoffs higher than *Maximin* to some other player). This inference can make an outcome which was not initially strictly 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, we have seen that players with a high enough aspiration threshold who play the PD will end up in a cycle made up of bilateral cooperations and/or bilateral defections (the only two strictly undominated outcomes; see Figure 5-7b). If through repeated interaction the 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[34] that the opponent is not exploitable either. When only bilateral decisions are confronted,

[34] A proposition *A* is mutual belief among a set of players if each player believes that *A*. Mutual belief by itself implies nothing about what, if any, beliefs anyone attributes to anyone else (Vanderschraaf and Sillari, 2007).

the only strictly undominated outcome is bilateral cooperation (see Figure 5-7c). When confronted with bilateral cooperation as the only alternative, bilateral defection is not a strictly undominated outcome 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. In the TC game all the outcomes in which the reward is given survive two steps of outcome dominance, and they are the only outcomes that do so. It can be shown that in any game, after applying any number of steps of outcome dominance, the remaining outcomes are not Pareto-dominated by any of the outcomes which have been eliminated.

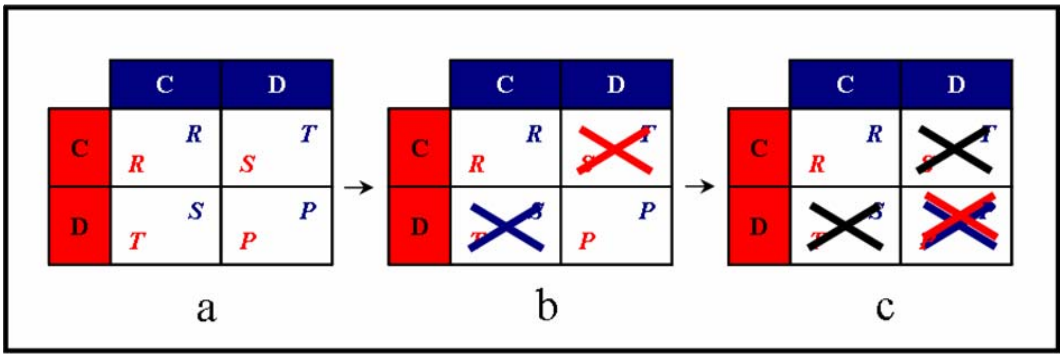


Figure 5-7. Elimination of dominated outcomes. Figure b shows the remaining outcomes after having applied one step of outcome dominance. Figure c shows the remaining outcomes after having applied two steps of outcome dominance. Red crosses represent outcomes which are unacceptable for player Red (row), blue crosses represent outcomes which are unacceptable for player Blue (column), and black crosses represent outcomes eliminated in previous steps.

How players would be able to move from bilateral defection to bilateral cooperation, if indeed they were, is not clear and is a matter for further research. We conjecture that this could be achieved by signalling processes to promote cooperation, or it could emerge from a form of learning by induction, since once the simulation has locked in to a cycle, it does show a general rule or pattern (players get a higher payoff when they cooperate than when they defect). Perhaps induction would then be produced by the simple forgetting of an episode's details and the consequent blurring together in memory of that episode with other similar episodes (Reisberg, 1999). In any case, the movement from bilateral defection to bilateral cooperation would require a non-trivial degree of coordination.

We have seen that if CBR players have a high enough aspiration threshold they are not exploitable in the sense that they do not accept outcomes where they are not getting at least *Maximin*. We find that a more useful definition of rationality in games is that of non-systemic-exploitability. Rational players are not systemically exploitable. According to this definition, cooperation emerges among selfish rational players as soon as it becomes mutual (not necessarily common) belief that the game is being played among rational players. Using Macy's words, cooperation would then emerge among self-interested agents "not from the shadow of the future but from the *lessons of the past*" (Macy, 1998).

5.6. Trembling hands process: the N-CBR model

While useful as a "tool to think with", the CBR model is admittedly rather unrealistic in the sense that simulations end up necessarily with players locked in to a persistent cycle. In this section we consider an extension of the CBR model where players may suffer from "trembling hands" (Selten, 1975) –i.e. they occasionally experiment (or make mistakes) with small probability. Importantly, we also significantly relax the assumptions made about what defines a perceived state of the world and about the decision-making algorithm used by players. These changes make the model more general, slightly more realistic, and the introduction of noise allows us to make more specific predictions. In particular, as in chapter 4, 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, i.e. the set stochastically stable outcomes. 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.

The definition of a case is substantially more general in the *noisy* CBR model (henceforth N-CBR) than in the CBR model. A case (an experience) lived by player *i* in the N-CBR model 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 a subset of the decisions undertaken by every player in the game (potentially all decisions by all players, including the case-holder *i*) in

the preceding ml_i (for *memory length*) time-steps. (Note that different players may have different memory lengths.) When comparing the N-CBR model with the CBR-model it will be assumed that players in the N-CBR model build their perceived state of the world as in the CBR model (see section 5.3).

- 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.

As in the CBR model, players in the N-CBR model 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. The definition of the decision-making algorithm in the N-CBR model is also substantially more general than in the CBR model. In a certain situation (i.e. for a given perceived state of the world) any particular player i will face one of two possibilities:

- Not every action available to player i is represented in C_i. Given the fact that players in the N-CBR model suffer from trembling hands (this is explained in detail below), this is a temporary situation. No assumptions are made in the N-CBR about how players make decisions in this situation. When comparing the N-CBR model with the CBR-model it will be assumed that players in the N-CBR model use, for this situation, the same decision-making algorithm as in the CBR model (see section 5.3).
- Every action available to player i is represented in C_i. As in the CBR model, in this situation player i selects randomly among those actions with the highest payoff obtained in the set C_i.

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 (Young, 1993). This completes the specifications of the N-CBR model.

This chapter 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.

5.7. Dynamics of the N-CBR model

The following explains why the N-CBR model has a unique limiting distribution. First, note that any N-CBR model can be formulated as a Markov chain where the state of the system is defined by every player's set of most recent cases that occurred in every possible perceived state of the world for each one of the actions available to her. Given the definition of the set of different states of the world possibly perceived by every player and the nature of the trembling hands noise, it is clear that this Markov chain is finite and has a unique recurrent class (where all actions available to each player i are represented in the set C_i for every state of the world possibly perceived by i). The trembling hands noise guarantees that it is possible to go from any recurrent state to any other recurrent state in a finite number of steps. This basically means that the N-CBR model can be formulated as a uni-reducible Markov chain, which has a unique limiting distribution (Janssen and Manca, 2006, Corollary 5.2, pg. 117).

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

5.7.1. Transient dynamics

To explore the transient dynamics of the N-CBR model we focus on the particular N-CBR process merely consisting of adding noise to the CBR model, and we study the Prisoner's Dilemma (PD). As one would expect, the short-term dynamics of this N-CBR process –i.e. when only a few trembles have taken place– are initially similar to the dynamics of the CBR 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. Figure 5-8 shows the proportion of outcomes where both players are cooperating (cooperation rate) in the PD for different values of both players' memory $ml_i = ml$ and aspiration threshold AT, and for different values of the overall probability of trembles ε. The cooperation rates shown in Figure 5-8 are calculated over time-steps 1001 to 1100.

A word of caution about Figure 5-8 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. $ml_i = 5$) 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 $ml_i = 0$, and $\varepsilon \neq 0.001$).

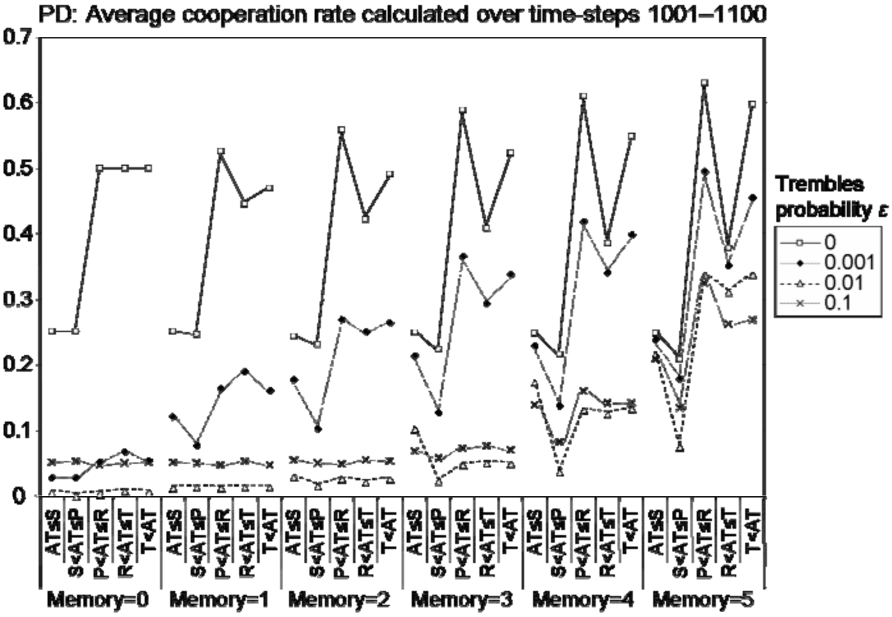


Figure 5-8. Average proportion of outcomes where both players are cooperating in the Prisoner's Dilemma (PD), calculated over 100 time-steps starting at time-step 1001, and using 500 simulation runs for each data point. The payoffs in the game are represented by its initial letter: S for Suckers, P for Punishment, R for Reward, and T for Temptation.

5.7.2. Asymptotic behaviour

Once enough trembles have taken place in every situation distinctively perceived by any player, the dynamics of the N-CBR model approach its asymptotic behaviour. The following proposition shows that a very broad range of N-CBR models share the same asymptotic behaviour:

Proposition 5-1: Assuming that every player has a common perception of the state of the world[35], the asymptotic behaviour of the N-CBR process is independent:

1. of the specific structure of the perceived state of the world (i.e. the algorithm used to construct it), and
2. of the decision-making algorithm employed by each player *i* when she has not explored every action available to her in a *similar* situation (i.e. when not every action available to player *i* is represented in C_i).

[35] This means that any two situations that look the same to one player will also look the same to every other player and any two situations that look different to one player will also look different to every other player.

Proposition 5-1, which is proved in Appendix B, implies that the asymptotic dynamics of all the simulations shown in Figure 5-8 are independent of the players' memory (see point 1 in the proposition) and of their aspiration thresholds (see point 2 in the proposition). 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 Figure 5-8, the series with low memory ($ml_i = 0$ or $ml_i = 1$) 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 reach the long-run behaviour of the process. If we represented the data in Figure 5-8 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 Figure 5-8 is that whereas mutual cooperation usually forms part of the cycles in the CBR model, it cannot be sustained in the long-term when small trembles occur.

To summarise, the dynamics of the N-CBR model follow a transition from a very path-dependent distribution similar to that corresponding to the CBR model, to a very different distribution which is only dependent on the probabilities with which trembles occur.

5.7.3. Stochastic stability

Having seen that the asymptotic behaviour of the N-CBR model is only dependent on the structure of trembles (assuming a common perception of the state of the world), 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 in N-CBR models (it is not assumed that players must share a common perception of the state of the world).

Proposition 5-2: In all N-CBR models, every stochastically stable outcome is individually rational.

The proof of Proposition 5-2 can be found in appendix B. Proposition 5-2 is a useful necessary condition to identify outcomes which cannot be stochastically stable but, except in very simple games (e.g. see Figure 5-9A), 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 program in Mathematica© that calculates the exact long-run probability that any 2-player game is in each possible outcome when the probability of trembles tends to zero. To calculate such probabilities, we did have to assume that players share a common perception of the state of the world. Using the computer program, we came to the following conclusions:

- Stochastically stable outcomes are not necessarily Nash equilibria (e.g. see the game of Chicken in Figure 5-9B).
- In fact, some players in some stochastically stable outcomes may be choosing strictly dominated strategies (e.g. see the game represented in Figure 5-9C).

- Nash equilibria are not necessarily stochastically stable (e.g. see the game of Stag Hunt in Figure 5-9D).
- Stochastically stable outcomes can be Pareto dominated by outcomes which are not stochastically stable (e.g. see the Prisoner's Dilemma game in Figure 5-9E). 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 Figure 5-9C, for example, if we 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 Figure 5-9A).

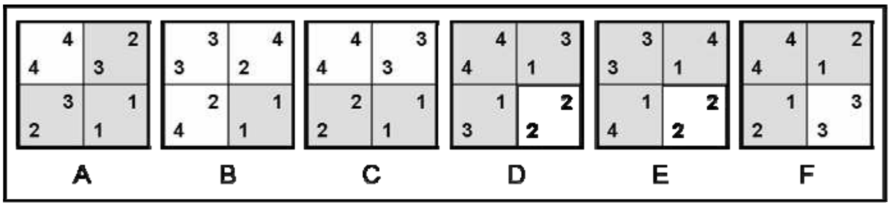


Figure 5-9. Stochastically stable outcomes (highlighted in white) in various 2-player 2-strategy games. Payoffs are numeric for the sake of clarity, but only their relative order for each player is relevant.

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 payoff[36]. The first possibility is related to the concept of Nash equilibrium, whilst the second is related to the concept of "protection" (Bendor et al., 2001b). As explained in section 4.7 when studying the Bush-Mosteller learning algorithm, an outcome is protected if unilateral deviations by any player do not hurt any of the other players. Bendor et al. (2001b) 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

[36] 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.

equilibrium. The same is not true for the model we study in this chapter (see the game represented in Figure 5-9F), but protected strict Nash equilibria are very relevant here too (as they were proved to be in the Bush-Mosteller model too; see section 4.7): 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 a 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.

5.8. Conclusions of this chapter

This chapter 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. The importance of the impact of unilateral deviations on non-deviators for the stability of outcomes was also highlighted in chapter 4. This seems to be a recurring observation in learning game theory: if a unilateral deviation harms another player, the non-deviator who has been hurt may choose to select a different strategy in the subsequent period, thus compromising the stability of the original strategy profile. A unilateral deviation that does not hurt any non-deviator is less likely to trigger a change of strategy in the non-deviators.

In broader terms, this chapter has proposed a new algorithm to narrow the set of expected outcomes in games. This method, i.e. iterative elimination of dominated outcomes, is a logical process through which outcome-based reasoners can arrive at sensible (*i.e.* Pareto optimal) outcomes in games. The only outcome that survives two steps of iterative elimination of dominated outcomes in the Prisoner's Dilemma is mutual cooperation. Thus, this chapter has shown that reasoning by outcomes leads to solution concepts significantly different from those present in the classical game theory literature (where reasoning is conducted using strategies as the key concept). Interestingly, one could argue that there is no a priori logical argument why rationality in game theory should be defined in terms of strategies rather than outcomes. Players in game theory do select a strategy (rather than an outcome), but the payoff they receive (i.e. their measure of performance) is determined by the resulting outcome, which is only partially determined by their selection of strategy. Thus, when defining rationality in game theory, it seems as natural to define it in terms of outcomes as the key concept (i.e. rational players do not choose dominated outcomes), as to define it using strategies (i.e. rational players do not accept dominated strategies). Reasoning by outcomes may even be a more natural way of modelling real human behaviour. Admittedly, the definition of rationality by outcomes proposed here implies some dynamicity (note the sentence: "players *do not accept* dominated outcomes"), whereas the definition of dominance reasoning does not. However, it is also true that, as explained in section 2.2.2, the concept of dominance reasoning is hardly ever enough to narrow the set of expected outcomes in games significantly, and when stronger concepts of rationality based on strategies are brought into play, issues at least as worrying as those that may be raised when defining outcome-based rationality often appear. These issues will be discussed further in chapter 7.

6. Structural Robustness of Evolutionary Models in Game Theory[♣]

6.1. Introduction

Naturally, the method that scientists have traditionally followed to advance our formal understanding of evolutionary social interactions has been to design and study models that were tractable with the tools of analysis available at the time. Until not long ago, such tools have derived almost exclusively from the realm of mathematics, and they have given rise to mainstream Evolutionary Game Theory (EGT). Mainstream EGT has proven to be tremendously useful (Weibull, 1995), but it is founded on many assumptions made to ensure that the resulting models could be mathematically analysed (e.g. infinite and homogeneous populations, random encounters, infinitely repeated interactions…). The aim of this chapter is to assess the extent to which some of these assumptions are affecting the conclusions obtained in mainstream EGT.

The assumptions made in EGT for the sake of mathematical tractability have had important implications both in terms of the *classes of systems* that have been investigated, and in terms of the *kind of conclusions* that have been drawn concerning such systems.

In terms of *classes of systems*, in order to achieve mathematical tractability, EGT has traditionally analysed *idealised systems*, i.e. systems that *cannot* exist in the real world (e.g. a system where the population is assumed to be infinite). Typically, mainstream EGT has also imposed various other assumptions that simplify the analysis, but which do not necessarily make the system ideal in our terminology (i.e. unable to exist in the real world). Some examples of common

[♣] Some parts of the material presented in this chapter have been published in Izquierdo, L. R., Izquierdo, S. S., & Polhill, J. G. (2006), “EVO-2x2: a modelling framework to study the evolution of strategies in 2x2 symmetric games under various competing assumptions”, in *Proceedings of the First World Congress on Social Simulation*, Kyoto, Japan, Vol. 2, pp. 273-280, and in Izquierdo, S.S. and Izquierdo, L.R. (2006). On the Structural Robustness of Evolutionary Models of Cooperation. *Lecture Notes in Computer Science* 4224, pp. 172-182.

assumptions in EGT are: populations are *well-mixed* (each individual is equally likely to interact with any other individual), interactions are *infinitely repeated*, strategies are *deterministic* and there is a *finite* set of them, individuals are selected with probabilities *proportional* to their fitness, and invasions are *homogenous* and *arbitrarily small*. Applying mainstream EGT to non-idealised systems can be very problematic because the validity for non-idealised systems of conclusions drawn from extremely similar idealised systems is not as straightforward as one may think. As an example, Beggs (2002) demonstrates that when analysing some types of evolutionary idealised systems, results can be widely different depending on the order in which certain limits are taken: if one takes the limit as population size becomes (infinitely) large and then considers the limit as the force of selection becomes strong, then one obtains different results from those attained if the order of the limits is inverted. Thus, Beggs (2002) warns that "care is therefore needed in the application of these approximations".

The need to achieve mathematical tractability has also influenced the *kind of conclusions* obtained in mainstream EGT. Thus, mainstream EGT has focused on analysing the stability of incumbent strategies to arbitrarily small mutant invasions, but has not paid much attention to the overall dynamics of the system in terms of e.g. the size of the basins of attraction of different evolutionary stable strategies, or the average fraction of time that the system spends in each of them.

Nowadays it has just become possible to start addressing the limitations of mainstream EGT outlined above. The current availability of vast amounts of computing power through the use of computer grids is enabling us to conduct formal and rigorous analyses of the dynamics of non-idealised systems through an adequate exploration of their sensitivity both to basic parameters and to their structural assumptions. These analyses can complement previous studies by characterising dynamic aspects of (idealised and non-idealised) systems beyond the limits of mathematical tractability. It is this approach that we follow in this chapter.

The structure of this chapter is as follows: section 6.2 outlines the general research question that EGT is mainly concerned with, and explains how our approach can complement the work conducted in mainstream EGT. Section 6.3 describes EVO-2x2, a computer simulation modelling framework designed to formally assess the impact of various assumptions commonly made in mainstream EGT. The subsequent two sections illustrate the use and the usefulness of EVO-2x2 with a particular example. The specific application selected here is a study of the structural robustness of evolutionary models of cooperation. To put our work into context, section 6.4 provides a brief and critical review of some of the most relevant work conducted on the evolutionary emergence of cooperation within the realms of game theory. Section 6.5 summarises some of the most interesting results we have obtained and the method we followed to analyse and summarise them. Finally, section 6.6 presents the conclusions of this investigation.

6.2. Overall research question and approach

In very broad terms, the question that EGT tries to answer is usually of the form: "In a population of individuals who repeatedly interact with each other, what sort of behavioural traits are likely to emerge and be sustained under evolutionary pressures?". Naturally, the answer to such a question may depend on a number of assumptions regarding population size, population structure (i.e. how individuals meet to interact), the specific nature of each interaction, the mechanisms through which natural selection occurs, and how mutations take place. In this chapter we present a formal modelling framework (EVO-2x2) designed to address this general question from different angles, i.e. using various different assumptions. EVO-2x2 provides a single coherent framework within which results obtained from different models can be contrasted and compared with analytical approaches. Thus, EVO-2x2 can be used to investigate the impact of various assumptions which may all be valid when trying to answer the general question posed above.

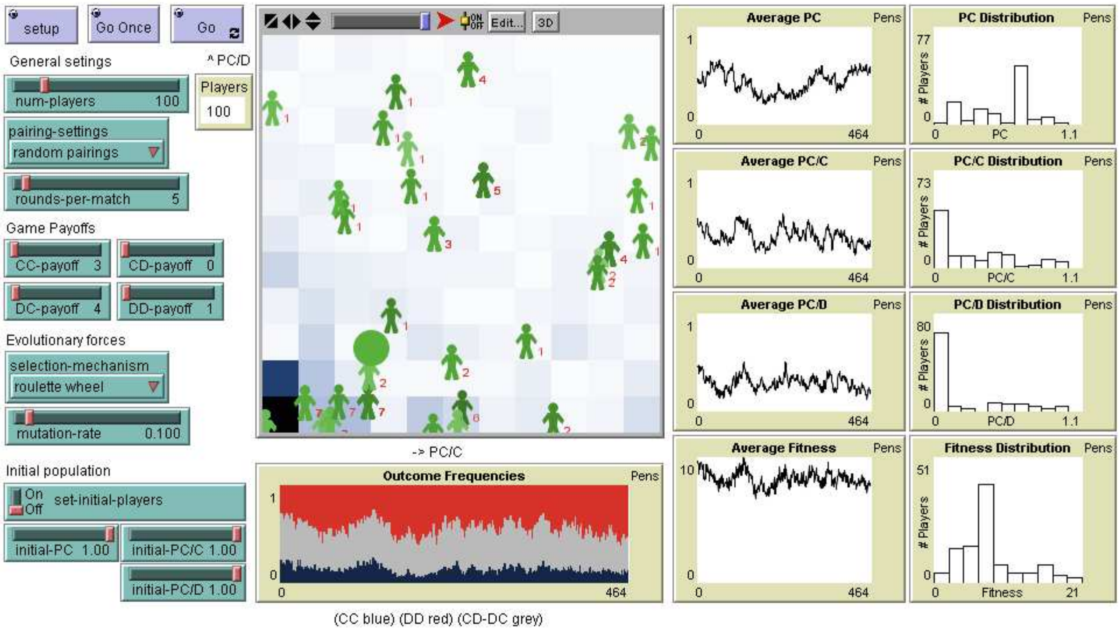
EVO-2x2 implements a wide range of competing plausible assumptions, all of which are fully consistent with the most basic principles of the theory of evolution. Logically, the assumptions embedded in EVO-2x2 limit its applicability. The most stringent assumption in EVO-2x2 is arguably the fact that interactions are modelled as 2-player 2-strategy (2x2) symmetric games. We will see in the next section, however, that individuals in EVO-2x2 are explicitly and individually represented, so any simulation conducted in EVO-2x2 is a non-idealised system (i.e. a system that could potentially exist in the real world). This move towards greater realism implies some loss of mathematical tractability, e.g. closed-form analytical solutions for the systems modelled in EVO-2x2 are not readily available. Nevertheless, EVO-2x2 is simple enough so many insights can be gained by using the theory of stochastic processes to analyse the results obtained by performing many simulation runs with it, as will be shown later. The following section explains all the assumptions embedded in EVO-2x2 in detail. Subsequently we illustrate the use of EVO-2x2 by studying the structural robustness of evolutionary models of cooperation.

6.3. Description of EVO-2x2

EVO-2x2 is a computer simulation modelling framework designed to formally investigate the evolution of strategies in 2x2 symmetric games under various competing assumptions. EVO-2x2 enables the user to set up and run many computer simulations (effectively many different models) aimed at investigating the same question using alternative assumptions. The specific question to be addressed is: "In a population of individuals who interact with each other by repeatedly playing a certain 2x2 symmetric game, what strategies are likely to emerge and be sustained under evolutionary pressures?".

6.3.1. The conceptual model

In this section we explain the conceptual model that EVO-2x2 implements. The information provided here should suffice to re-implement the same conceptual model on any platform. Figure 6-1 provides a snapshot of EVO-2x2 interface, which is included here to clarify the explanation of the model. The reader may also want to consider following the explanation of the model using it at the same time; EVO-2x2 is included in the Supporting Material of this thesis. We use bold red italicised arial font to denote ***parameter*** names.



General setings

num-players 100

Players 100

pairing-settings random pairings

rounds-per-match 5

Game Payoffs

CC-payoff 3

CD-payoff 0

DC-payoff 4

DD-payoff 1

Evolutionary forces

selection-mechanism roulette wheel

mutation-rate 0.100

Initial population

On Off set-initial-players

initial-PC 1.00

initial-PC/C 1.00

initial-PC/D 1.00

Figure 6-1. Snapshot of the interface in EVO-2x2.

Overview of EVO-2x2

In EVO-2x2, there is a population of ***num-players*** players. Events occur in discrete time-steps, which can be interpreted as successive generations. At the beginning of every generation every player's payoff (which denotes the player's fitness) is set to zero. Then, every player is paired with another player, according to some customisable procedure (***pairing-settings***), to play a 2-player match.

Each match consists of a number of sequential rounds (***rounds-per-match***). In each round, the two members of the pair play a symmetric 2x2 game once, where each of them can undertake one of two possible actions. These two possible actions are called cooperate (C) and defect (D). The action selected by each of the players determines the magnitude of the payoff that each of them receives in that round (***CC-payoff***, ***CD-payoff***, ***DC-payoff***, ***DD-payoff***). The total payoff that a player obtains in a match is the sum of the payoffs obtained in each of the rounds.

Players differ in the way they play the match, i.e. they generally have different strategies. The strategy of a player is determined by three numbers in the interval [0 , 1]:

- *PC*: Probability to cooperate in the first round.

- *PC/C*: Probability to cooperate in round *n* (*n* > 1) given that the other player has cooperated in round (*n* – 1).
- *PC/D*: Probability to cooperate in round *n* (*n* > 1) given that the other player has defected in round (*n* – 1).

Once every player has played one –and only one– match (except when the pairing mechanism is *round robin*, as explained below), two evolutionary processes (i.e. natural selection (***selection-mechanism***) and mutation (***mutation-rate***)) come into play to replace the old generation with a brand new one. Successful players (those with higher payoffs) tend to have more offspring than unsuccessful ones. This marks the end of a generation and the beginning of a new one, and thus the cycle is completed.

Parameters

The value of every parameter in EVO-2x2 can be modified at run-time, with immediate effect on the model. This enables the user to interact closely with the model by observing the impact of changing various assumptions during the course of one single run.

Population parameters

num-players: Number of players in the population. This number is necessarily even for pairing purposes.

set-initial-players: This is a binary variable that is either *on* or *off*. If *on*, every player in the initial population will have the same strategy, which is determined using the following parameters: ***initial-PC***, ***initial-PC/C***, and ***initial-PC/D***. If *off*, the initial population of strategies will be created at random using a uniform distribution.

Rounds and Payoffs

rounds-per-match: Number of rounds in a match.

CC-payoff: Payoff obtained by a player who cooperates when the other player cooperates too.

CD-payoff: Payoff obtained by a player who cooperates when the other player defects.

DC-payoff: Payoff obtained by a player who defects when the other player cooperates.

DD-payoff: Payoff obtained by a player who defects when the other player also defects.

Pairing settings

This parameter (***pairing-settings***) determines the algorithm that should be used to form pairs of players. There are three options:

- *random pairings*: Pairs are made at random, without any bias. Every player plays one and only one match in a generation.
- *round robin*: Every player is paired with every other player once, so every player plays exactly (***num-players*** – 1) matches per generation.
- *children together*: Players are paired preferentially with their siblings (and at random among siblings). Once all the possible pairs between siblings have been made, the rest of the players are paired at random. Every player plays one and only one match in a generation. This procedure was implemented because it seems plausible in many biological contexts that individuals belonging to the same family tend to interact more often among them than with individuals from other families. The algorithm is formally equivalent to simple applications of tags (Holland, 1993) in evolutionary models (see Hales, 2000).

Evolutionary forces

selection-mechanism: This parameter determines the algorithm used to create the new generation. There are four options:

- *roulette wheel*: This procedure involves conducting ***num-players*** replications, which form the new generation. In each replication, players from the old generation are given a probability of being chosen to be replicated that is proportional to their total payoff (which denotes their fitness).
- *Moran process*: In each time-step (i.e. generation), one player is chosen for replication with a probability proportional to its fitness. The offspring replaces a randomly chosen player (possibly its parent). Payoff totals are set to zero at the beginning of every time-step.

- *winners take all*: This method selects the player(s) with the highest total payoff (i.e. the "winners"). Then, for **num-players** times, a random player within this "winners set" is chosen to be replicated. The **num-players** replications constitute the new generation. Note that this mechanism (which is sometimes called "cultural imitation", e.g. see Traulsen et al., 2006) violates the proportional fitness rule.
- *tournament*: This method involves selecting two agents from the population at random and replicating the one with the higher payoff for the next generation. In case of tie, one of them is selected at random. This process is repeated **num-players** times. The **num-players** replications form the new generation.

mutation-rate: This is the probability that any newly created player is a mutant. A mutant is a player whose strategy (the 3-tuple formed by *PC*, *PC/C*, and *PC/D*) has been determined at random.

6.3.2. Displays

EVO-2x2 provides various displays which are shown in Figure 6-1. Some of these displays are time-series plots showing the historical evolution of the value of a particular variable throughout generations (e.g. frequency of outcomes and population average values of *fitness*, *PC*, *PC/C*, and *PC/D*), whereas others refer only to the last generation (e.g. population distributions of *fitness*, *PC*, *PC/C*, and *PC/D*).

The large square in the middle of the interface is the representation in the strategy space of every individual player in a generation. This representation is 2-dimensional in EVO-2x2 due to constraints in the modelling platform (NetLogo 3.0.2), but we also provide in the Supporting Material a 3D version of EVO-2x2, called EVO-2x2-3D (implemented in NetLogo 3-D Preview 1), where the three dimensions of the strategy space (*PC*, *PC/C*, and *PC/D*) are explicitly represented. This is the only difference between EVO-2x2-3D and EVO-2x2: EVO-2x2-3D represents players in the *PC–PC/C–PC/D* 3-dimensional strategy space, while EVO-2x2 displays the projection of such a space on the *PC/C–PC/D* plane (Figure 6-2). In Figure 6-2, the sphere (in the left-hand image) and its circular projection (in the right-hand image) indicate population averages.

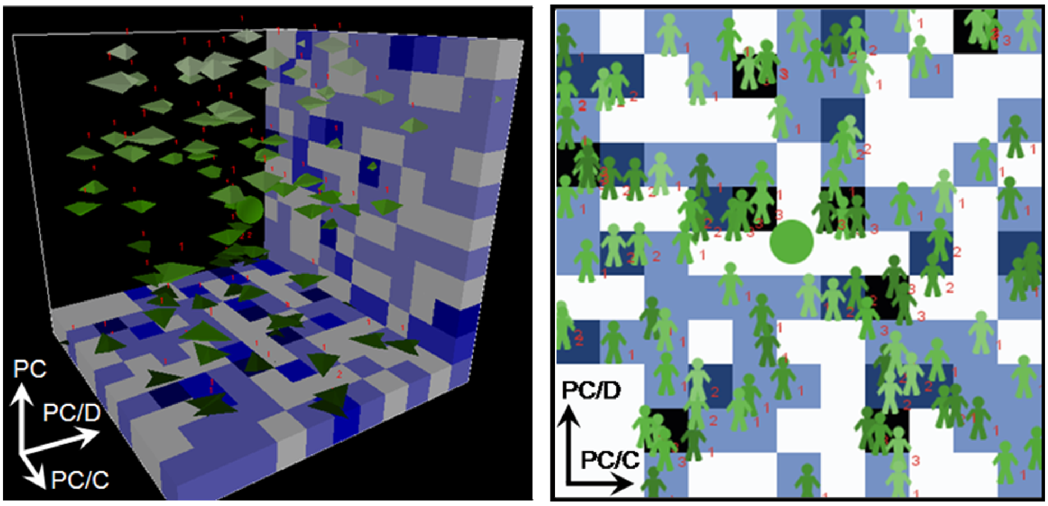


Figure 6-2. Representation of players in the strategy space using EVO-2x2-3D (left) and EVO-2x2 (right). The image on the right shows the top-down projection of the representation on the left.

The cells in the background of the 2-dimensional projections of the strategy space are coloured in shades of blue according to the number of players that have spent some time on them. Each player that has visited a certain part of the strategy space leaves a mark that is used to create the density plots shown in Figure 6-2. The more players who have stayed for longer in a certain area, the darker its shade of blue.

6.3.3. Exploration of the parameter space

The rationale behind EVO-2x2 was to conduct a systematic exploration of the impact of various competing assumptions. An exploration of the parameter space is something that can be easily conducted within NetLogo using a tool called BehaviorSpace. This tool allows the user to set up and run experiments. Running an experiment consists in running a model many times, systematically varying the model's settings and recording the results of each model run.

The problem when undertaking experiments that involve large parameter sweeps is to organise, analyse, and summarise the vast amount of information obtained from them so the results can be meaningfully interpreted. To do that, we have created a set of supporting scripts (written in Perl and Mathematica, and available in the supporting Material) that are able to read in the definition of the experiment setup and all its results in the format used by NetLogo. The output of these scripts is:

- an automatically generated directory structure that reflects all the combinations of parameter values explored in the experiment (e.g. /100/random-pairings/roulette-wheel/0.001/…/), and
- a customisable summary of the results of each model run, which is placed in the appropriate folder.

An example of a useful summary of the results produced in a simulation run is the accumulated frequency of different types of strategies throughout the course of a simulation run. This is something that can be plotted in a 3D contour plot, and in complementary 2D density plots, as shown in Figure 6-3. The relationship between the 3D contour plot and the accompanying 2D density plots is sketched in Figure 6-4.

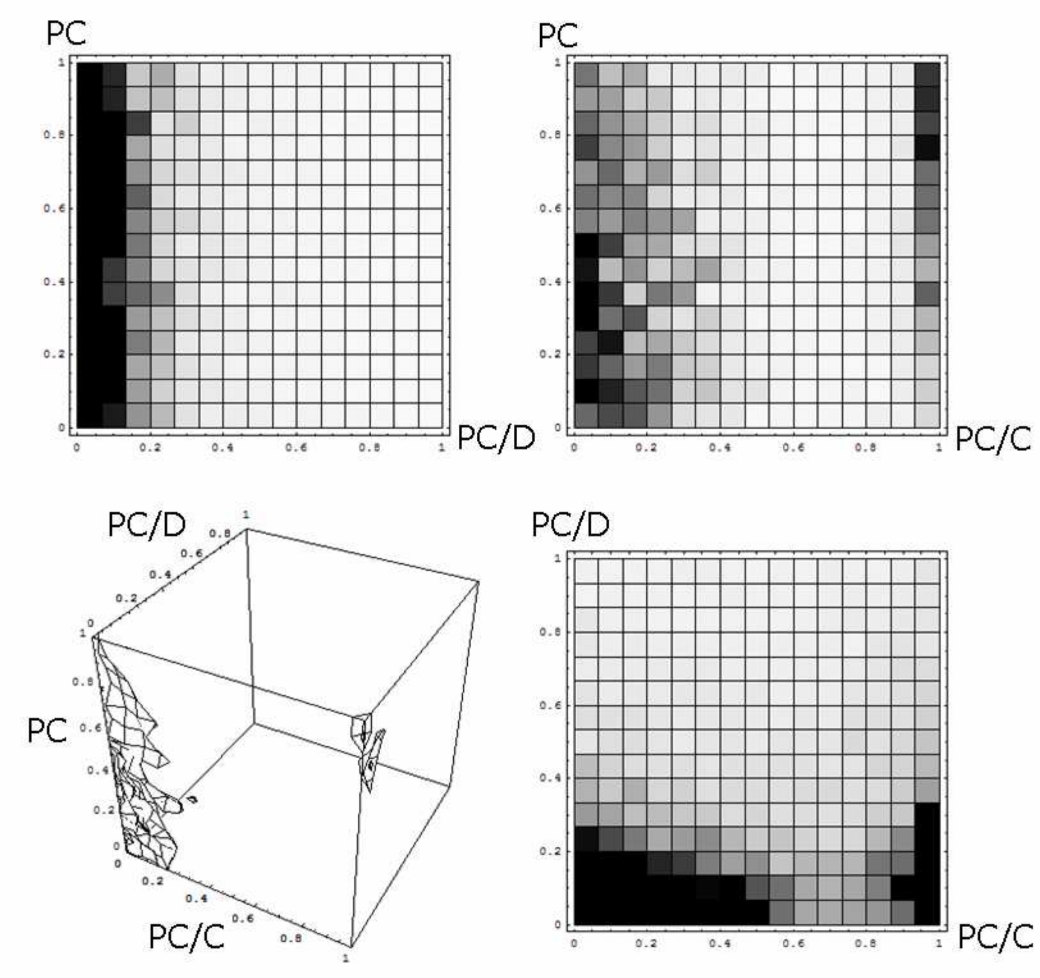


Figure 6-3. Example of a graphical summary of the results obtained with EVO-2x2. This figure is automatically created and placed in the appropriate folder by the supporting scripts.

6.3.4. Implementation details

EVO-2x2 has been implemented in NetLogo 3.0.2 (Wilensky, 1999). We also provide a 3-D version of EVO-2x2, called EVO-2x2-3D, which has been implemented in NetLogo 3-D Preview 1 (Wilensky, 1999). The two programs are available in the Supporting Material together with a user guide under the GNU General Public Licence.

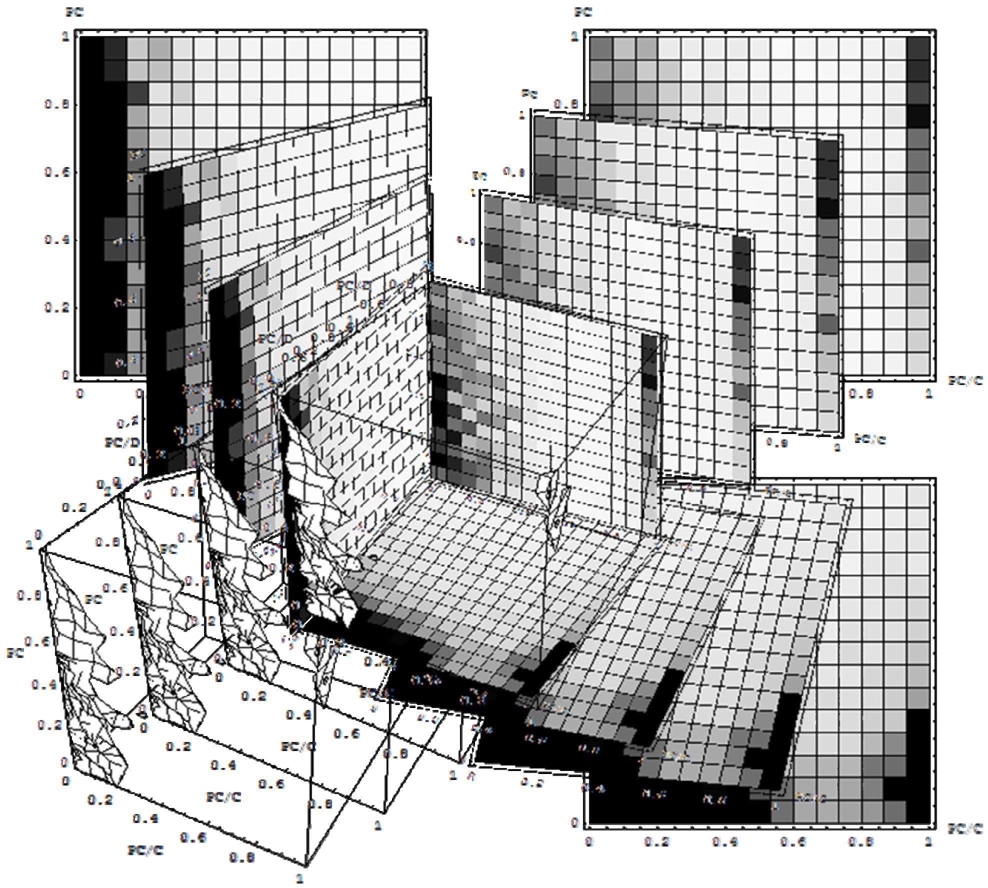


Figure 6-4. Sketch showing the relationship between the 3D contour plot and the accompanying 2D density plots created by the supporting scripts.

6.4. Evolutionary emergence of cooperation

The fundamental challenge of understanding the evolutionary emergence and stability of cooperation can be illuminated, at the most elementary level, by identifying the conditions under which a finite number of units that interact by playing the Prisoner’s Dilemma (PD) may cooperate. These units might be able to adapt their individual behaviour (i.e. learn), or the population of units as a whole may adapt through an evolutionary process (or both). While formalizing the problem of cooperation in this way significantly decreases its complexity (and generality), the question still remains largely unspecified: how many units form the population? How do they interact? What strategies can they use? What is the value of each of the payoffs in the game? and, crucially, what are the processes governing the dynamics of the system?

It has been well known since the early years of the study of the evolution of cooperation that, in general, the question of how –if at all– cooperation emerges in a particular system significantly depends on all of the above defining characteristics of the system (see e.g. Axelrod, 1984; Bendor and Swistak, 1995, 1997, 1998; Gotts et al., 2003b). Here we report previous work that has shed light on the robustness of evolutionary models of cooperation. We find it useful to place these models in a fuzzy spectrum that goes from mathematically tractable models with strict assumptions that limit their applicability (e.g. work on idealised systems), to models with the opposite characteristics. The rationale behind the construction and use of such a spectrum is that when creating a formal model to investigate a certain question (e.g. the evolution of cooperation), there is often a trade-off between the applicability of the model (determined by how constraining the assumptions embedded in the model are) and the mathematical tractability of its analysis (i.e. how deeply the functioning of the model can be understood given a certain set of available tools of analysis).

The former end is mostly populated by models *designed to* ensure mathematical tractability. Near this end we find papers that study the impact of some structural assumptions, whilst still keeping others which ensure the model remains tractable and which, unfortunately, also tend to make the model retain its idealised nature. Gotts et al. (2003b) review many such papers in sections 2 and 4. Some of these investigations have considered finite vs. infinite populations (Nowak et al., 2004; Taylor et al., 2004; Imhof et al., 2005), different pairing settings or population structures (see section 6 in Gotts et al. (2003b) for a review, and Santos et al. (2006) for the most recent advances in this field), deterministic vs. stochastic strategies (Nowak, 1990; Nowak and Sigmund, 1990; Nowak and Sigmund, 1992), finite vs. infinitely repeated games (Nowak and Sigmund, 1995), and arbitrary intensities of selection (Traulsen et al., 2006). While illuminating, the applicability of most of these studies is somewhat limited since, as mentioned before, the models investigated there tend to retain their idealised nature.

Near the opposite end, we find models that tend to be slightly more applicable (e.g. they consider non-idealised systems), but they are often mathematically intractable. It is from this end that we start in our investigation. To our knowledge, the first relevant study with these characteristics was conducted by Axelrod (1987). As explained in section 3.1, Axelrod had previously organized two open tournaments in which the participant strategies played an iterated PD in a round robin fashion (Axelrod, 1984). Tit for Tat (TFT) was the winner in both tournaments, and also in an *ecological analysis* that Axelrod (1984) conducted after the tournaments. Encouraged by these results, Axelrod (1987) investigated the generality of TFT's success by studying the evolution of a randomly generated population of strategies (as opposed to the arguably arbitrary set of strategies submitted to the tournament) using a particular genetic algorithm. The set of possible strategies in this study consisted of all deterministic strategies able to consider the 3 preceding actions by both players. From this study, Axelrod (1987) concluded that in the long-term, "reciprocators […] spread in the population, resulting in more and more cooperation and greater and greater effectiveness". However, the generality of Axelrod's study (1987) is doubtful for two reasons: (1) he used a very specific set of assumptions, the impact of which was not tested, and (2) even if we constrain the scope of his conclusions to his particular model, the results should not be trusted since Axelrod only conducted 10 runs of 50 generations each. As a matter of fact, Binmore (1994, p. 202; 1998) cites unpublished work by Probst (1996) that contradicts Axelrod's results.

In a more comprehensive fashion, Linster (1992) studied the evolution of strategies that can be implemented by two-state Moore machines in the infinitely repeated PD. He found a strategy called GRIM remarkably successful. In particular, GRIM was significantly more successful than TFT. GRIM always cooperates until the opponent defects, in which case it switches to defection forever. Linster (1992) attributed the success of GRIM over TFT to the fact that GRIM is able to exploit poor strategies while TFT is not. Linster's investigation was truly remarkable at its time, but technology has advanced considerably since then, and we are now in a position to expand his work significantly by conducting parameter explorations beyond what was possible before. As an example, note that Linster (1992) could only consider deterministic strategies and one specific value for the mutation rate; furthermore, in the cases he studied where the dynamics were not deterministic, there is no guarantee that his simulations had reached their asymptotic behaviour.

Another important part of the literature on the study of the evolutionary emergence of cooperation using computer simulation comes from the use of tags. Tags are socially recognisable marks or signals that, in principle, are not necessarily linked to any particular form of behaviour (Holland, 1993). Tags do, however, influence the way individuals interact: individuals with similar tags have a preference to interact with each other (see e.g. Riolo (1997), Hales (2000), Riolo et al. (2001), Edmonds and Hales (2003)). Tags, like strategies, are also assumed to be passed from parents to their kin. Thus, tags and strategies follow a very similar evolutionary process. The resulting correlation between tags and strategies leads to a tendency for individuals with similar strategies to interact with each other. In the context of social dilemmas this correlation clearly favours cooperative behaviours, as it effectively diminishes the chances of exploitation.

Riolo (1997) developed the first tag model in the study of the evolutionary emergence of cooperation in the PD. He showed that real-valued tags can promote high levels of cooperation in the iterated PD. Hales (2000) developed Riolo’s work and studied discrete tags, with preferential pairings occurring only if tags matched exactly. With this exact tag matching constraint, cooperation can emerge even when players interact for only one round. Hales’ pairing mechanism is formally equivalent to “children-together” in EVO-2x2 (see section 6.3.1). Tags as a useful mechanism to promote cooperation were further explored by Riolo et al. (2001). This piece of work, however, turned out to be flawed, as it relied upon the fact that individuals were forced to donate to others with an identical tag (see Roberts and Sherratt (2002) and Edmonds and Hales (2003) for a much more in-depth investigation). Since then research using tags has worked towards making this cooperation mechanism more robust, so it can be usefully applied in real-world contexts (see e.g. Hales and Edmonds (2005), and Edmonds (2006)).

In the following section we use EVO-2x2 to conduct a consistent and systematic exploration of the impact of competing assumptions in non-idealised evolutionary models of cooperation.

6.5. Robustness of evolutionary models of cooperation

In this section we illustrate the usefulness of EVO-2x2 by applying it to advance our formal understanding of the structural robustness of evolutionary models of cooperation. To do this, we analyse simple non-idealised models of cooperation and we study their sensitivity to small structural changes (e.g. slight modifications in the way players are paired to interact, or in how a generation is created from the preceding one). Specifically, we aim to determine what behavioural traits are likely to emerge and be sustained under evolutionary pressures in the Prisoner's Dilemma (PD). To do this rigorously, we have run many computer simulations (effectively many different models) aimed at addressing the same question: "In a population of individuals who interact with each other by repeatedly playing the PD, what strategies are likely to emerge and be sustained under evolutionary pressures?". Given the amount of computing power required to conduct this research, all the simulations have been run on computer grids.

6.5.1. Method followed to analyse the simulation results

Defining a state of the system as a certain particularisation of every player's strategy, it can be shown that all simulations in EVO-2x2 with positive mutation rates can be formulated as irreducible positive recurrent and aperiodic discrete-time finite Markov chains. Thus, ergodicity is guaranteed. This observation enables us to say that there is a unique long-run distribution over the possible states of the system, *i.e.* initial conditions are immaterial in the long-run (Theorem 3.15 in Kulkarni (1995)). Although calculating such (dynamic) distributions analytically is infeasible, we can estimate them using the computer simulations. The problem is to make sure that a certain simulation has run for long enough, so the limiting distribution has been satisfactorily approximated. To make sure that this is the case, for each possible combination of parameters considered, we ran 8 different simulations starting from widely different initial conditions. These are the 8 possible initial populations where every individual has the same pure strategy (the 8 corners of the strategy space). Then, every simulation run is conducted for 1,000,000 generations. Thus, in those cases where the 8 distributions are similar, we have great confidence that they are showing a distribution close to the limiting distribution[37]. As an example, consider Figure 6-5, where distributions starting from the 8 different initial conditions are compared.

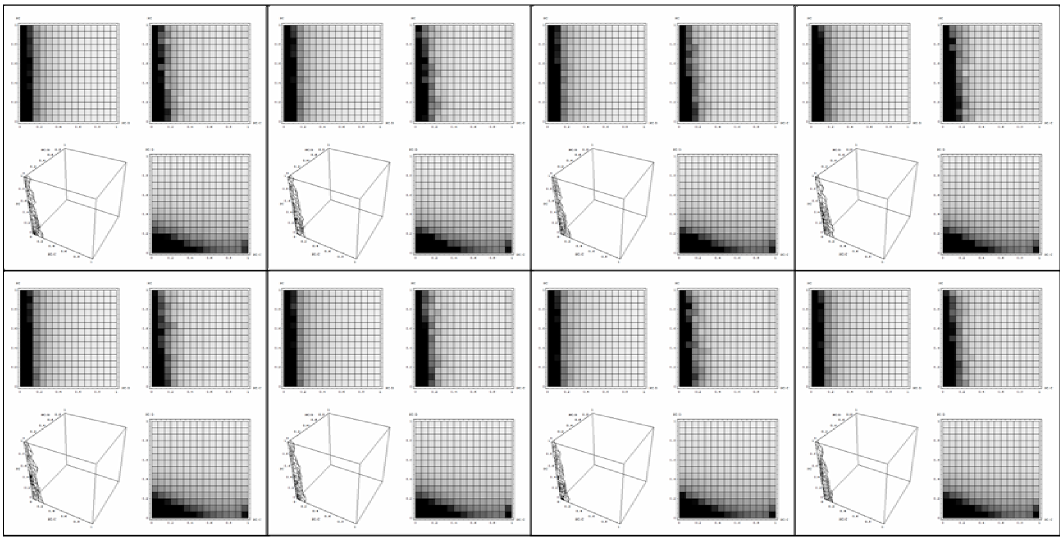


Figure 6-5. Accumulated frequency of different types of strategies in 8 simulation runs starting from different initial conditions. Axes are as in Figure 6-3.

6.5.2. Results and discussion

In this section we report several cases where it can be clearly seen that some of the assumptions in EGT that are sometimes thought to have little significance (e.g. mutation-rate, number of players, or population structure) can have a major

[37] The appropriateness of the inductive method used here (which is not formal proof) to infer the asymptotic distribution of the system can be qualitatively checked by thinking what would happen if this method were to be applied to study the system characterised in chapter 4. In that case, the method would consist in running 4 simulations starting from the corners of the strategy space. Clearly, simulations starting in an SRE would stay there forever. Thus, only in those cases where there is really a unique asymptotic distribution, would the 4 simulations eventually look similar, and only when very close to the limiting distribution. In other words, the method used here would work perfectly well for the system characterised in chapter 4: the 4 cumulative distributions would look similar if and only if they were close to the limiting distribution.

impact on the type of strategies that emerge and are sustained throughout generations. The following are parameter values that are common to all the simulations reported here[38]:

CC-payoff = 3; ***CD-payoff*** = 0; ***DC-payoff*** = 5; ***DD-payoff*** = 1;

selection-mechanism = *roulette wheel*;

Consider first the two distributions in Figure 6-6, which only differ in the value of the mutation rate used (0.01 on the left, and 0.05 on the right). The distribution on the left shows the evolutionary emergence and (dynamic) permanence of strategies similar to TFT ($PC \approx 1$, $PC/C \approx 1$, and $PC/D \approx 0$; average time ≈ 3.3%). Such strategies are observed one order of magnitude less frequently for slightly higher mutation rates (distribution on the right; average time ≈ 0.3%). The other parameter values used were ***num-players*** = 100; ***pairing-settings*** = *random pairings*; ***rounds-per-match*** = 50.

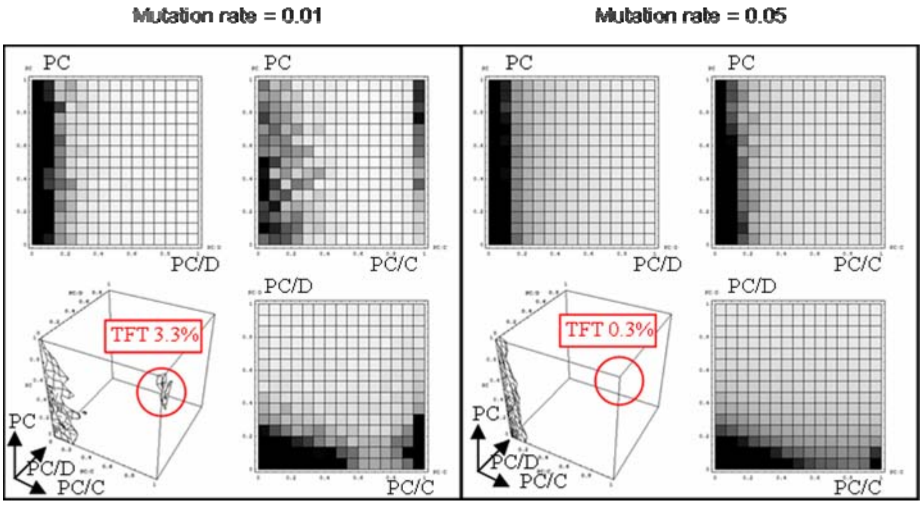


Figure 6-6. Influence of the mutation rate on the dynamics of the system. TFT measures the average time that strategies with $PC \geq (13/15)$, $PC/C \geq (13/15)$ and $PC/D \leq (2/15)$ were observed.

The two distributions in Figure 6-7 only differ in the number of players in the population (100 on the left, and 10 on the right). The distribution on the left shows

[38] The payoffs used in this chapter are those employed by Axelrod (1984), and consequently those used in most simulation papers on the evolution of cooperation. They are used here too to facilitate comparisons with previous research.

the evolutionary emergence and (dynamic) permanence of strategies similar to TFT (average time ≈ 3.3%), whereas –again– such strategies are observed one order of magnitude less frequently in smaller populations (average time ≈ 0.4%). The other parameter values are: **pairing-settings** = *random pairings*; **rounds-per-match** = 50; **mutation-rate** = 0.01.

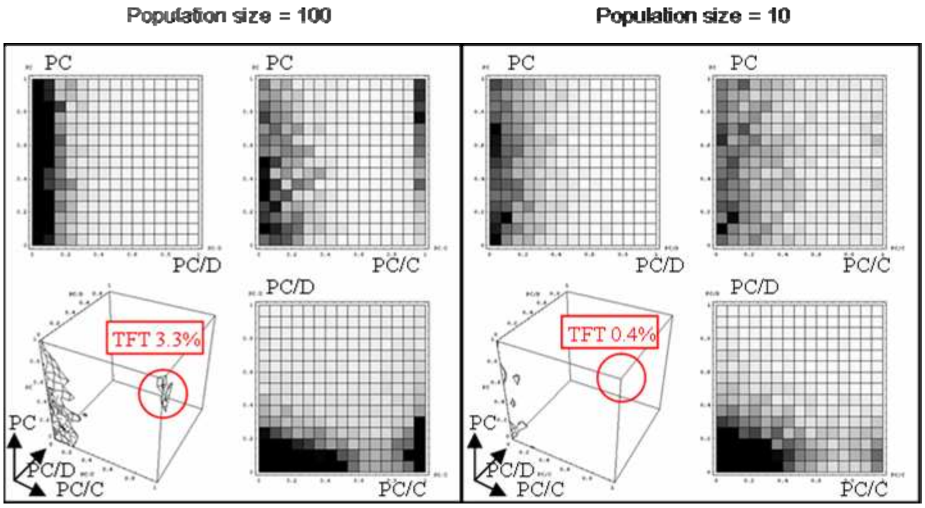


Figure 6-7. Influence of the number of players in the population. TFT measures the average time that strategies with $PC \geq (13/15)$, $PC/C \geq (13/15)$ and $PC/D \leq (2/15)$ were observed.

The two distributions in Figure 6-8 only differ in the algorithm used to form the pairs of players (*random pairings* on the left, and *children together* on the right). On the left, strategies tend to be very similar to ALLD ($PC \approx 0$, $PC/C \approx 0$, and $PC/D \approx 0$), i.e. strongly uncooperative (average time ALLD ≈ 72%). In stark contrast, the distribution on the right is concentrated around strategies similar to TFT (average time TFT ≈ 23%; average time ALLD ≈ 1%). The other parameter values used were: **num-players** = 100; **rounds-per-match** = 5; **mutation-rate** = 0.05. The underlying reason behind the dramatic increase in cooperation when using the pairing algorithm "children together" (which is formally equivalent to simple applications of tags, see e.g. Hales, 2000) is that this mechanism promotes mimicry. Children, who have inherited the same strategy from their parents, tend to be paired together. This confers a great evolutionary advantage to cooperation, since it effectively rules out the possibility of exploitation: cooperators (and defectors) play only with each other.

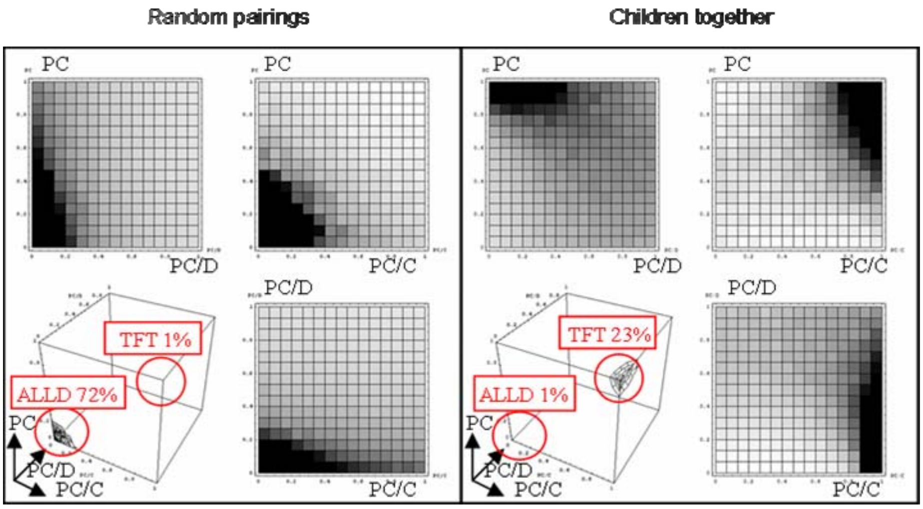


Figure 6-8. Influence of different pairing mechanisms. TFT measures the average time that strategies with *PC* ≥ (10/15), *PC/C* ≥ (10/15) and *PC/D* ≤ (5/15) were observed; ALLD measures the average time that strategies with *PC* ≤ (5/15), *PC/C* ≤ (5/15) and *PC/D* ≤ (5/15) were observed.

Figure 6-9 shows a very interesting result. The two distributions in Figure 6-9 only differ in the set of possible values that *PC*, *PC/C* or *PC/D* can take. For the distribution on the left the set of possible values is any (floating-point) number between 0 and 1, and the strategies are mainly uncooperative, similar to ALLD (average time ALLD ≈ 60%). For the distribution on the right, the set of possible values is only {0, 1}, and the distribution is concentrated in TFT (average time TFT ≈ 58%). The other parameter values used were: ***num-players*** = 100; ***mutation-rate*** = 0.05; ***rounds-per-match*** = 10; ***pairing-settings*** = *random pairings*.

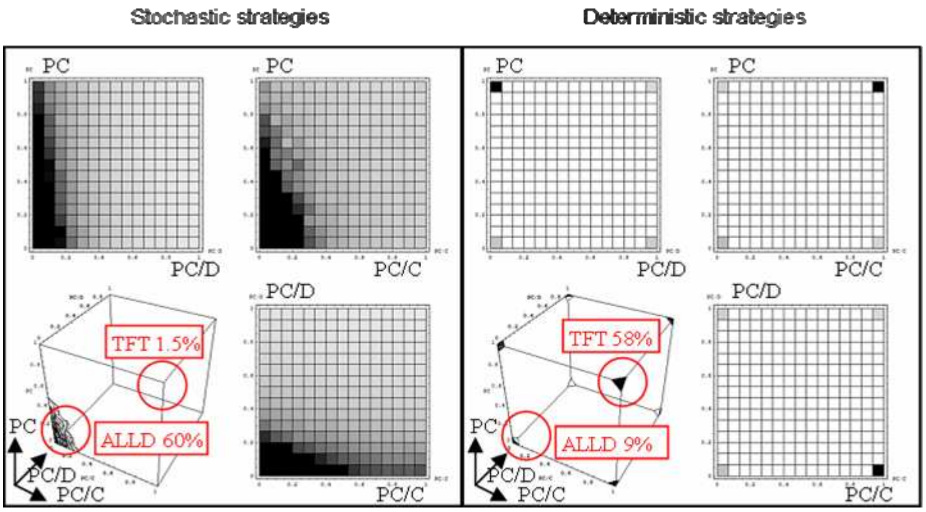


Figure 6-9. Stochastic (mixed) strategies vs. deterministic (pure) strategies: influence in the system dynamics. TFT measures the average time that strategies with *PC* ≥ (10/15), *PC/C* ≥ (10/15) and *PC/D* ≤ (5/15) were observed; ALLD measures the average time that strategies with *PC* ≤ (5/15), *PC/C* ≤ (5/15) and *PC/D* ≤ (5/15) were observed.

Given the clarity and importance of the results presented in Figure 6-9 we investigated this issue further. In Figure 6-10 and Figure 6-11 we show the effect of gradually increasing the set of possible values for *PC*, *PC/C* and *PC/D* (i.e. ***num-strategies***). Figure 6-10 shows the (average) number of each possible outcome of the game (CC, CD/DC or DD) in observed series of 10^6 matches (this number of matches is selected so the effect of changing the initial state is negligible, i.e. results are close to the stationary limiting distribution).

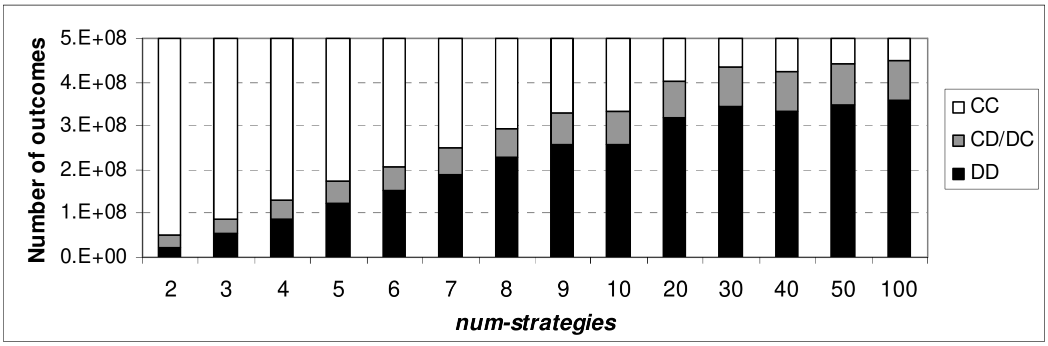


Figure 6-10. Influence in the distribution of outcomes (CC, CD/DC or DD) of augmenting the set of possible values for *PC*, *PC/C* and *PC/D*.

Figure 6-11 shows the average values of *PC*, *PC/C* and *PC/D* observed in the same series. Augmenting the set of possible values for *PC*, *PC/C* and *PC/D* undermines cooperation and favors the emergence of ALLD-like strategies. The other parameter values used were: **num-players** = 100; **mutation-rate** = 0.01; **rounds-per-match** = 10; **pairing-settings** = *random pairings*.

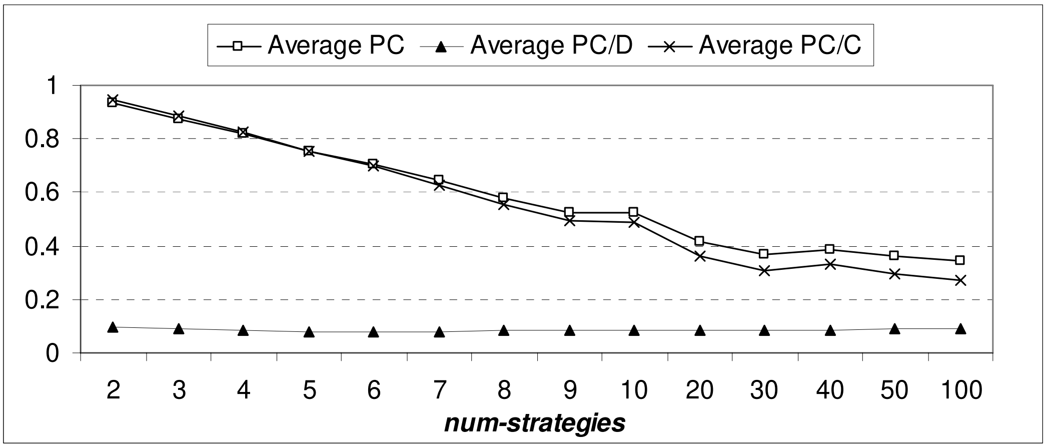


Figure 6-11. Influence of augmenting the set of possible values for *PC*, *PC/C* and *PC/D* in the average values of these variables in the population.

Thus, it is clear that the number of possible strategies has a tremendous effect on the evolutionary stability of cooperation. This is mainly due to the fact that the emergence of TFT-like behaviour crucially relies on perfect reciprocation. A single defection in a contest between two TFT-like strategies with high –but lower than 1– values of *PC/C* will result in a chain of uncoordinated outcomes CD-DC, thus losing much of their evolutionary advantage over ALLD.

6.6. Conclusions of this chapter

In this chapter we have shown by example that some of the assumptions made in mainstream evolutionary game theory for the sake of mathematical tractability can have a greater effect than what has been traditionally thought. In particular, the granularity of the strategy space and the assumption of well-mixed populations have proved to be critical in determining the type of strategies that are likely to emerge and be sustained in evolutionary contexts.

More specifically, this chapter has studied the structural robustness of evolutionary models of cooperation, i.e. their sensitivity to small structural changes. To do this, we have focused on the Prisoner's Dilemma game and on the set of stochastic strategies that are conditioned on the last action of the player's counterpart. Strategies such as Tit-For-Tat (TFT) and Always-Defect (ALLD) are particular and classical cases within this framework; here we have studied their potential appearance and their evolutionary robustness, as well as the impact of small changes in the model parameters on their evolutionary dynamics. Our results show that strategies similar to ALLD tend to be the most successful in most environments, whereas strategies similar to TFT tend to spread best in large populations, where individuals with similar strategies tend to interact more frequently, when only deterministic strategies are allowed, with low mutation rates, and when interactions consist of many rounds.

7. Discussion

In broad terms, most of the results presented in the previous 3 chapters can be seen as logical deductive inferences of the form:

“Set of assumptions A” IMPLIES “Set of (deduced) statements B” [7-1]

As a matter of fact, any computer simulation and any mathematical derivation can be seen as a logical inference that establishes the truth of a set of statements B (e.g. the output of a model, or a derived mathematical result) given the assumption that a set of statements A (expressed in e.g. computer code, or as a set of equations) are true.

Deductive logical inferences are more useful the greater the generality of the set of assumptions A, and the greater the scope and level of detail of the set of deduced statements B. As an example, consider the results presented in chapter 4 on the dynamics of the Bush-Mosteller reinforcement model. These results advance previous work by Cross (1973) and by Börgers and Sarin (1997) because the results derived in this thesis are valid not only for positive stimuli, but also for negative ones; thus, the generality of the set of assumptions investigated in this thesis is greater. Similarly, the results presented in that same chapter are an advancement of (parts of) the work conducted by Macy and Flache (2002) and Flache and Macy (2002) on the Bush-Mosteller model because the level of detail of the characterisation of this model’s dynamics is significantly greater in this thesis.

The logical inferences derived in this thesis can be applied in a number of useful ways. This chapter outlines 5 ways in which the research conducted in the previous chapters can be usefully applied to contribute to the advancement of human knowledge.

7.1. Direct application of the derived inferences

The simplest application of the logical statement “A implies B” relates to the case where A is thought, postulated, or demonstrated to be true. If a set of individuals are playing a certain game using one of the decision-making algorithms investigated in this thesis (e.g. the Bush-Mosteller reinforcement learning algorithm), then the results obtained in the previous chapters can be used to *predict* the (dynamic) outcome of the game, and also how this outcome may change when certain conditions (e.g. the magnitude of the payoffs or the speed at which players learn) are modified. Similarly, since “A implies B” is logically equivalent to “Not B implies Not A”, if the observed results are deemed significantly different from B, then logical statement [7-1] can be used to infer that A cannot be true.

7.2. Assessment of the importance of assumptions in similar models

Another way in which logical statement [7-1] can be meaningfully used concerns the identification of crucial assumptions in inferences of the type “Set of assumptions A2 implies set of statements B2”. Consider the case where sets A and A2 contain a large number of identical assumptions. An example of this would be two models of the same game: one of the models (A2) assumes common knowledge of rationality among the players, whereas the other model (A) assumes that players make decisions following the Bush-Mosteller reinforcement learning approach. Comparing the set of deduced results B and B2 will be illuminating: any difference between B and B2 can be attributed to the differences between A and A2. Thus, inference [7-1] can be used to assess the impact of various assumptions in models that are similar to the one defined by the set of assumptions A, but not the same. A clear illustration of this type of inference in the literature is given by Flache and Hegselmann (1999), who compare two models that differ only in the decision-making algorithm used by a set of players confronting the same spatial social dilemma setting: in one of the models, players use (partially) rational strategies that cooperate whenever reciprocal cooperation can be sustained as a rational equilibrium in the 2-player game they play (i.e. whenever the “shadow of the future” (Axelrod, 1984) is powerful enough); in the other model, players use a reinforcement learning rule based on Bush and

Mosteller's (1955) principles. In particular, Flache and Hegselmann (1999) show that under a wide range of conditions, the reinforcement learners need more time than the (partially) rational players to form stable cooperative relationships. This line of work was further developed by Hegselmann and Flache (2000), who compared rational behaviour and the Bush-Mosteller reinforcement learning rule over all possible symmetric 2x2 prisoner's dilemma games.

7.3. Selection, parameterisation, and validation of models

A third way in which the research conducted in this thesis contributes to the advancement of human knowledge concerns the interdependent processes of selecting, parameterising, and validating a model. A model is an abstraction of a real-world system that allows us to establish inferences about how the real-world system or certain aspects of it operate. Any model represents a compromise between realism and manageability (Intriligator et al., 1996, p. 13). Ideally, one would like to have a model that captures the essence of the target system (i.e. the model is realistic) and, at the same time, enables us to draw insights and conclusions that could not be derived from direct observation of the target system (i.e. the model is manageable). A perfectly manageable model that is not realistic is not useful; similarly, a realistic model that is not manageable (i.e. it does not yield new insights) is useless. This thesis has increased the manageability of several models that have received empirical support, thus improving their applicability. In this way, the work reported in this thesis enhances game theorists' toolkit of models that can be usefully employed to study real-world systems.

The task of selecting one particular model often includes considering various different alternatives. Naturally, the choice of criteria for the comparison of models depends on the purpose of the modelling exercise. Models in game theory are often compared with the aim of understanding what decision-making processes may be generating an observed pattern of play (see e.g. Feltovich (2000) and Camerer (2003)). For that purpose, one is often interested in studying the models' ability to reproduce observed statistical signatures and to predict patterns of play to a satisfactory extent. To conduct this assessment, the models to be compared need to be parameterised first. The following section outlines how to do this.

7.3.1. Parameterisation of models

As explained in section 3.2.2, the models investigated in this thesis can all be meaningfully formalised as Markov processes. The implicit assumption when parameterising a model with a set of observed data is that such data have been generated by the (appropriately parameterised) model. The challenge when parameterising the models studied in this thesis is that they represent systems where the state is not a variable that can be observed, i.e. the Markov chain is *hidden*. What is available to an observer is the pattern of play (i.e. the decisions made by the players), which is a stochastic process governed by the underlying Markov chain, but different from it. As an example, consider the Bush-Mosteller model of reinforcement learning. As explained in chapter 4, the model can be formalised as a Markov chain $\{X_k\}_{k\geq 0}$ whose state is fully specified by a two-dimensional vector [$p_{1,\mathrm{C}}$, $p_{2,\mathrm{C}}$], where $p_{i,\mathrm{C}}$ is player i's probability to cooperate. The sequence of actual decisions made by the players is another stochastic process $\{Y_k\}_{k\geq 0}$ which is linked to the hidden Markov chain $\{X_k\}_{k\geq 0}$ in the sense that X_k governs the distribution of the corresponding Y_k. Since only $\{Y_k\}$ is observed, any statistical inference about the unknown parameters of the Markov chain $\{X_k\}$ must be done in terms of $\{Y_k\}$. Fortunately, methods to parameterise hidden Markov chains have been developed remarkably in the last few years. An excellent introduction to conduct this type of parameterisation is given by Cappé et al. (2005). In addition to the analysis of the pattern of play, it could well be the case that the value of certain parameters can be inferred using various other methods, like purpose-designed experiments, questionnaires or interviews with the players. These methods may be more reliable, simpler and, in any case, constitute a source of potentially very useful information that does not decrease the validity of the quantitative methods described above; thus, it seems most advisable to conduct them, if at all feasible.

7.3.2. Selection, validation, and applicability of models

Once the models to be compared have been parameterised, the process of selecting one can proceed. This is an activity that is strongly linked with the process of model validation. In broad terms, models are compared with the aim of selecting the best one of them according to some set of criteria, whereas validating the selected model is studying whether this (best) model is “good enough” for the intended purpose. Thus, it seems natural that the same techniques used to pick out the best model are appropriate to assess its validity too.

A model is valid to the extent that it provides a satisfactory range of accuracy consistent with the intended application of the model (Kleijnen, 1995)[39]. As mentioned above, models in game theory are often constructed with the aim of understanding what decision-making processes may be generating an observed pattern of play. In that context, validation often refers to the process of assessing how well a model is capturing the essence of its empirical referent. As mentioned above, one should not forget that a simple approach to validate a model about how certain individuals played a game is actually asking that same question to the individuals themselves[40]. Unfortunately, this does not seem to be a common approach in the literature of experimental game theory, even though it seems clear that it has the potential to contribute significantly to the design of more realistic models. The long tradition of introspective theoretical work in classical game theory may be at the root of this apparent lack of interaction with experimental subjects.

One common technique to quantify the extent to which a model is capturing the essence of a pattern of play consists in studying the models’ ability to reproduce observed statistical signatures and to predict patterns of play to a satisfactory extent. This is an issue extensively studied in the systems identification literature (Söderström and Stoica, 1989; Ljung, 1999). The general approach to validate a model is based on an in-depth analysis of its prediction error, which is a measure of the disparity between the observed data and the model’s predicted output. If possible, the preferred option is to evaluate the model performance using a set of

[39] See a complete epistemic review of the validation problem in Kleindorfer et al. (1998).

[40] Work outside the literature in experimental game theory suggests that players’ responses may vary depending on *when* they are asked to describe their reasoning processes (Ericsson and Simon, 1980). People tend to verbalise what they are doing more accurately when asked *while they solve a problem* rather than when asked *some time after having tackled the problem*.

data different from the data employed to parameterise the model (i.e. the estimation data). If, on the other hand, the prediction error has to be calculated using the estimation data, there are a number of model selection criteria (e.g. Akaike's information criterion (Akaike, 1969) and minimum description length (Rissanen, 1978)) designed to avoid biases and pitfalls (e.g. overparameterisation and overfitting) by adding certain correcting terms to the computed prediction error (Ljung, 1999, p. 507). These correcting approaches are especially relevant when comparing models that have different number of parameters. An important part of the validation exercise is then the analysis of residuals (i.e. the part of the validation data that the model could not reproduce). This analysis minimally consists in plotting the residuals, computing basic statistics on them, analysing their structure, and conducting tests of independence. The precise purpose of the modelling exercise will dictate what other tests will be useful.

At this point it is worth addressing a criticism that the Bush-Mosteller model investigated in chapter 4 of this thesis has recently received, and which relates to its applicability. Bendor et al. (2007) argue that the BM model (and many others) have "little empirical content" because "such models imply that virtually anything can happen" (see reply by Macy and Flache (2007)). They prove their point showing that any outcome of the game can be sustained as a stable outcome by some pure SRE. Their proof of this result consists in setting an aspiration threshold below the lowest payoff of the game. As shown in chapter 4, once a certain value for the aspiration threshold is chosen, it is not generally true that any outcome can be sustained by an SRE. In fact, it is straightforward to see that any value for the aspiration threshold above the minimum payoff will preclude at least one outcome from being sustained by an SRE. Thus, their criticism refers to a Bush-Mosteller model where players have aspiration thresholds below the minimum payoff they can receive. In our view, the aspiration threshold is a parameter whose value can be estimated using empirical methods by e.g. using the theory of inference in hidden Markov chains mentioned in the previous section. The fact that it is possible to find a specific value for the aspiration threshold such that any outcome can be supported by an SRE is not a drawback of the model, since the value of the aspiration threshold can be inferred from empirical observation, and most of the values this parameter can take induce a process where not every outcome can be sustained by an SRE. An analogy that comes to mind is Newton's theory of gravitation: this theory provides (in particular) a mapping between the height at which an object is released and the time that the object takes to hit the ground (time = f(height)). Similarly, this thesis has characterised the (non-trivial) mapping between the parameters of the Bush-Mosteller model (in particular, the aspiration threshold) and the dynamics of the resulting process (in particular, the characterisation of the set of SREs):

Set_of_SREs = function(Aspiration_Threshold).

It is indeed true that for any given outcome, one can always find an aspiration threshold so the outcome is supported by an SRE. Similarly, in Newton's theory of gravitation, for any time t_0 one can always find a height h_0 such that $f(h_0) = t_0$, but this does not seem to be a drawback of the theory.

Bendor et al.'s (2007) criticism seems to be unjustified even in the case where aspiration thresholds are so low that any outcome can be sustained by an SRE. As explained in chapter 4, even in the case where there is a positive probability that any outcome will be played indefinitely, this probability is generally different for different outcomes and depends on a number of factors (e.g. initial conditions, aspiration thresholds, and learning rates). The exact probability of approaching each possible SRE can be estimated to any degree of accuracy using the methods explained in chapter 4. Thus, the Bush-Mosteller model yields predictions that can be falsified, even when aspiration thresholds are below the minimum payoff.

7.4. Modelling frameworks

As explained in chapter 2, there is nowadays a whole universe of models that abandon the demanding assumptions of classical game theory on players' rationality and beliefs. These models make different assumptions regarding the meaning of payoffs, the amount of information that players can access, players' computational capabilities, and the level at which the dynamics are described (i.e. population adaptation vs. individual learning), to mention a few. The formal analysis of these models is often quite challenging, and consequently most of the research conducted until now has focused on characterising the dynamics of each of these non-trivial models in relative isolation. There is obviously a lot to be gained from comparing different models, but our lack of in-depth knowledge of their dynamics has meant that this comparison has had to be postponed. Fortunately, nowadays the number of models that have been thoroughly analysed seems to be sufficient to justify initiating the process of creating frameworks –i.e. meta-models– where alternative models would arise as particular cases.

An example of a useful framework that has been proposed within the field of learning game theory is Flache and Macy's (2002) general reinforcement learning (GRL) framework. Flache and Macy's (2002) framework integrates a smoothed version of the Erev-Roth model (see section 4.1) and the Bush-Mosteller model as particular cases. The GRL framework has a parameter that measures the level of fixation in the decision-making algorithm. When this fixation parameter equals 0, the framework reduces to the Bush-Mosteller model, whereas if the parameter equals 1, the obtained model is Erev and Roth's. The use of the GRL framework enabled Flache and Macy to conduct a transparent and fruitful comparison of the two models and also to uncover hidden assumptions in both models.

An example of a framework within the field of evolutionary game theory is EVO-2x2. As explained in chapter 6, EVO-2x2 is a computer simulation modelling framework designed to formally investigate the evolution of strategies in 2x2 symmetric games under various competing assumptions. EVO-2x2 enables the user to set up and run many computer simulations (effectively many different models) aimed at investigating the same question using alternative assumptions. Thus, EVO-2x2 provides a single coherent framework within which results obtained from different stochastic finite models can be contrasted and compared, as illustrated in section 6.5.2.

The development of frameworks is useful not only to assess the impact of various assumptions in theoretical terms, but also to inform experimental research. By making differences between models explicit, frameworks can facilitate the design of experiments targeted at identifying the type of models that may be most adequate in a certain situation. Frameworks can also help to identify the factors (i.e. types of assumption) that may have the greatest impact in the outcome of a social interaction. Thus, the use of frameworks may facilitate the interaction between game theorists and empirically-driven social scientists, from which game theory would benefit so much. The ideal result of this interaction would be a framework encompassing various models as particular cases, where the differences between the models were made explicit, and where each model were annotated with indications about the type of context for which the model may be most adequate.

A discussion about frameworks raises the question of whether evolutionary and learning game theory could be integrated into a single discipline. The derivation of a significant number of theoretical results relating various learning models with different versions of the replicator dynamics (e.g. Börgers and Sarin, 1997; Posch, 1997; Hopkins, 2002; Hopkins and Posch, 2005) would seem to suggest that the integration of these two fields may be within reach (Weibull, 1998). However, the integrative theoretical results tend to establish analogies at a very high level of abstraction. A representative example is given by Börgers and Sarin (1997), who demonstrate that the continuous time limit approximation of the dynamics of the Bush-Mosteller learning model (which cannot be used to characterise its asymptotic behaviour, as demonstrated in chapter 4) converges to the replicator dynamics of evolutionary game theory. These types of result are certainly useful, as they provide non-biological interpretations of evolutionary models, and evolutionary interpretations of learning models. However, the number of assumptions that are needed to align models from the two disciplines tend to decrease the applicability of the obtained inferences significantly. Thus, it seems that there are many frameworks that can be usefully developed at lower level of abstractions before the integration of learning and evolutionary game theory can take place.

7.5. Models as 'tools to think with'

The formal models developed in this thesis have also been useful as 'tools to think with'. The clearest example of this use of a model is illustrated in section 5.5, where the concept of iterative elimination of dominated outcomes was put forward. Iterative elimination of dominated outcomes is a logical process through which players 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*. 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 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 that 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.

In this section we expand the philosophical basis of this process of reasoning by outcomes a bit further. As mentioned several times in this thesis, the history of classical game theory has been marked by the assumption that agents are instrumentally rational. However, except in strictly competitive games, defining rational behaviour in games is by no means straightforward (Colman, 1995). The challenge in game theory is that, in general, the definition of rational behaviour for any one player depends on the behaviour of potentially every other player in the game. As an example, in an iterated Prisoner's Dilemma game, the rational strategy against a player who always defects is to defect, but the rational strategy against a player who is known to play Tit for Tat may be to cooperate, if the number of rounds is sufficiently large.

Thus, in order to identify the rational course of action in a game, one is bound to partition the infinite set of possible behaviours that the other players may take according to some criterion, and then try to compute the best reply to each type of behaviour identified. Classical game theory partitions this universe of possible behaviours according to strategies. In this way, classical game theory defines rationality in terms of beliefs about the *strategy* that the other players may use: rational players do not choose dominated *strategies* because there is no belief about the other players' *strategies* such that selecting the dominated *strategy* is optimal. The partition of the "behaviour space" according to strategies is quite natural since, after all, it is strategies that players can choose.

On the other hand, players' measure of success –i.e. the obtained payoff– is not determined solely by their strategy, but by every player's strategy, i.e. by the resulting *outcome* of the interaction. Thus, it may also seem natural to assume that players do not think in terms of *strategies*, but in terms of *outcomes*. In other words, players may be willing to accept certain *outcomes* but not others. The models developed in chapter 5 triggered the idea of defining rationality partitioning the universe of possible behaviours according to *outcomes*, instead of strategies. This leads to the definition of the so-called outcome-based rationality. According to this definition, rational players do not accept dominated *outcomes*. Note that this definition is somewhat problematic, since the words "do not accept" already imply the existence of some dynamics. Remember, however, that the definition of rationality based on strategies also led to similarly worrying problems (e.g. the existence of many possible Nash equilibria).

Once outcome-based rationality is defined, one can develop the same concepts that were explained in section 2.2.2 using the new definition of rationality. Thus, one can define the process of iterative elimination of dominated outcomes, and also the concept of rationalisable outcomes.

The definition of outcome-based rationality has a certain intuitive appeal which becomes apparent when studying the Prisoner's Dilemma. The process of iterative elimination of dominated outcomes leaves mutual cooperation as the unique surviving outcome. The reasoning behind this logical process goes as follows: players are rational and therefore they will not accept the outcome where they receive the sucker's payoff. They also know that the other player is rational, so they acknowledge the fact that their counterpart is not going to be exploitable either. Once this is recognised by the two players, the rational course of action is to try to achieve mutual cooperation rather than mutual defection.

It seems clear that even though there is a clear causal link between strategies and outcomes, defining rationality in terms of outcomes rather than in terms of strategies leads to completely different results even in the simplest games. Section 5.2 explained how rational strategies may lead to outcomes that are not rational, whereas rational outcomes may be generated by strategies that are not rational. A more thorough account of the implications of outcome-based rationality is left for future work.

8. Conclusions

This thesis was initiated with the overall aim of advancing game theory by formally studying the implications of dropping some of its most stringent assumptions, which have been made for the sake of tractability and are not generally supported by empirical evidence.

Naturally, the first part of this research consisted in clearly identifying the most relevant and prevalent assumptions made in the different branches of game theory. This investigation led to the critical dissection of deductive game theory presented in chapter 2, which served as a guiding framework to structure the rest of the research conducted in this thesis. In particular, this critical review enabled a precise identification of those assumptions of game theory that are abandoned and those that are retained in the models developed in this thesis. Specifically, all the research conducted here abandons the strong assumptions made in classical game theory regarding player's rationality, players' beliefs about their counterparts' behaviour, and the alignment of such beliefs across players. The research conducted in this thesis also abandons the assumption of one single *infinite* population, which is commonly made in evolutionary game theory, and which was shown in chapter 2 to have wider implications than may be initially suspected.

The abandonment of several assumptions that are made in game theory to allow for mathematical tractability has meant that new methodologies were needed to formally analyse the models developed in this thesis. In particular, computer simulation has proven to be particularly useful to enhance and complement mathematical derivations. The combined use of analytical work and computer simulation has enabled me to draw some methodological conclusions that are also included in this chapter.

The structure of this final chapter is particularly simple. Section 8.1 summarises the main contributions of this thesis to the advancement of game theory. These are presented at two different levels of abstraction for the sake of clarity: subsections 8.1.1 and 8.1.2 present the specific contributions of this thesis to the advancement of learning and evolutionary game theory respectively (and the implications of these for the study of social dilemmas), whereas subsection 8.1.3 discusses in more general terms the wider implications of the research conducted here for game theory as a whole. The methodological conclusions derived from the symbiotic use of computer simulation and mathematical analysis are then summarised in section 8.2. Finally, the last section of this chapter (8.3) identifies areas for future research.

8.1. Contributions to the advancement of game theory

8.1.1. Specific contributions to learning game theory

Chapter 4 of this thesis provided an in-depth analysis of the transient and asymptotic dynamics of the Bush-Mosteller reinforcement learning algorithm, whereas chapter 5 explored cased-based reasoning as decision-making process in strategic contexts. The specific insights obtained for each of these learning algorithms were summarised in sections 4.10 and 5.8 respectively. The following presents the main conclusions that can be drawn from this investigation in more general terms:

- The transient dynamics of models in learning game theory can be substantially different from their asymptotic behaviour. Moreover, some systems may take an extraordinarily long time to reach their asymptotic dynamics (see e.g. Figure 4-8 and Figure 5-8). This is especially important because most theoretical research focuses on the characterisation of asymptotic equilibria exclusively, whereas studies using computer simulation tend to explore only the short-term dynamics of models.
- The transient dynamics of models in learning game theory tend to be very complex and highly path-dependent (see e.g. section 5.4). Players learn from each other's actions in a very dynamic fashion, and their individual responses affect every player's payoff (and –consequently– their subsequent behaviour). This means that one single decision made by one player may change the evolution of the whole system substantially and have a permanent effect on its overall dynamics (especially in models without "trembling hands noise").

- It has been long known that the inclusion of “trembling hands noise” can affect the dynamics of models in learning game theory. This thesis has illustrated that this type of noise can *completely* change the dynamics of a model by showing that some outcomes that are observed with arbitrarily high probability in unperturbed models can effectively lose all their attractiveness if players make occasional mistakes in selecting their actions (see e.g. sections 4.8 and 5.7).
- In general, occasional mistakes made by players can destabilise outcomes in two different ways: by giving the deviator a higher payoff, or by giving any of the non-deviators a lower payoff. Thus, outcomes where unilateral deviations hurt the deviator (strict Nash) but not the non-deviators (protected) tend to be the most stable (see sections 4.8 and 5.7.3).

The application to social dilemmas of the models developed in this thesis (and the review of similar models in the literature) has enabled me to draw the following general conclusions in this regard:

- Cooperation in social dilemmas is not only a common outcome in models where players learn from each other’s behaviour, but also the unique asymptotic outcome in many cases (see sections 4.1, 4.5 and 5.4).
- Cooperative outcomes are most commonly observed in models where players satisfice to some extent: they have an aspiration threshold that divides the set of outcomes into two classes: satisfactory and unsatisfactory outcomes. Naturally, aspiration thresholds that make the cooperative outcome satisfactory and the non-cooperative outcome unsatisfactory tend to promote the highest rates of cooperation (see sections 4.7 and 5.4).
- Cooperative outcomes tend to be particularly susceptible to be destabilised by small trembles. This is so because deviations have two undesirable effects: they favour the deviator *and* they hurt the non-deviators. Therefore trembles in cooperative outcomes encourage all cooperating players to change their behaviour. On the other hand, non-cooperative outcomes are particularly robust to trembles because deviations from them hurt the

deviator *and* benefit the non-deviators, thus encouraging everyone to keep defecting (see sections 4.8 and 5.7.3).

8.1.2. Specific contributions to evolutionary game theory.

Chapter 6 described EVO-2x2, the modelling framework developed in this thesis to assess the impact of various assumptions made in mainstream evolutionary game theory for the sake of mathematical tractability. The following summarises the main conclusions that can be drawn from this investigation in general terms (for more specific conclusions see section 6.6):

- The study of the evolution of finite populations is significantly different from that of infinite populations (both in terms of the methods that are adequate for their analysis and on the results obtained with them). This fact has serious implications, since most of our intuitions about evolutionary dynamics come from analyses of models where populations are infinite.
- Stochastic effects (e.g. the potential occurrence of two or more mutations at the same time) play an important role in the analysis of finite evolutionary systems (see sections 2.3.4 and 6.5).
- The type of strategies that are likely to emerge and be sustained in finite evolutionary contexts is strongly dependent on assumptions that traditionally have been thought to be unimportant or secondary (e.g. number of players, continuity of the strategy space, mutation rate, and population structure). See results presented in section 6.5.2.
- There seems to be great value in developing general frameworks that facilitate rigorous and transparent comparisons between different stochastic finite models and the results obtained with them.

The use of EVO-2x2 was illustrated by conducting an investigation on the structural robustness of evolutionary models of cooperation. The results obtained in that research (and other papers in the literature – see e.g. Imhof et al., 2005) showed that stochastic evolution of *finite* populations need not select the strict Nash equilibrium (as is the case when making the assumptions of mainstream evolutionary game theory) and can therefore favour cooperation over defection. Stochastic finite systems exhibit dynamics over the strategy space with time averages that –for some parameterisations– are concentrated around cooperative strategies (e.g. TFT; see section 6.5.2).

8.1.3. General contributions to game theory

The dissection of game theory made in chapter 2 of this thesis (and some of the issues discussed in section 7.5) showed that classical game theory is founded on rather problematic assumptions that may have deeper philosophical implications than commonly assumed. Fortunately, this has been increasingly acknowledged in the last few years, and several models that abandon the demanding assumptions of classical game theory on players' rationality and beliefs have been put forward and analysed in depth. This reasonably new programme of research, to which the present thesis contributes, is starting to provide fruitful insights.

This thesis in particular has thoroughly analysed the dynamics of two models of learning that have received notable empirical support (see chapters 4 and 5). In this way, the work reported here enhances game theorists' toolkit of models that can be usefully employed to study real-world systems. One of the main challenges that game theory faces nowadays derives from the need of managing and synthesising the various insights obtained with a number of disparate models that abandon the stringent assumptions of game theory through different avenues. This diversity of new assumptions and results calls for the creation of frameworks aimed at facilitating a clear and transparent comparison between models and the results obtained with them. This thesis has tried to meet this challenge by placing its contributions in an overall framework that can encompass, in admittedly very broad terms, most of the research conducted in game theory until now (see chapter 2). In the particular context of evolutionary game theory, the modelling framework developed in chapter 6, i.e. EVO-2x2, represents a step forward in this direction too. Using EVO-2x2, it has been demonstrated here that some of the assumptions made in mainstream evolutionary game theory for the sake of mathematical tractability can have a greater effect than has been traditionally thought. Specifically, the granularity of the strategy space and the assumption of well-mixed populations have proved to be critical in determining the type of strategies that are likely to emerge and be sustained in evolutionary contexts (see section 6.5).

Thus, in general terms, this thesis has contributed to game theory (a) by examining the formal implications of replacing some of the unsupported assumptions in mainstream game theory with assumptions that stem from empirical research, and (b) by creating frameworks aimed at making differences between models explicit and at facilitating the comparison of results obtained with different models.

8.2. Methodological contributions

Before the development of computational modelling, the formal analysis of game theoretical models could be conducted using mathematical analyses only, and this may have distorted our understanding of such models to some extent. This thesis has shown that computer modelling can greatly enhance and complement mathematical derivations. These two techniques to analyse formal systems are both extremely useful, and they are complementary in the sense that they can provide fundamentally different insights on the same issue. Chapter 4 is a clear illustration of the fact that the level of understanding gained by using these two techniques together could not have been obtained using either of them on their own. Thus, the use of only one of these techniques may lead to an incomplete picture of the dynamics of a model. Chapter 4 also illustrates how each technique can produce both problems and hints for solutions for the other.

This thesis has also shown that most models in learning and evolutionary game theory can be usefully formalised as Markov processes. In the absence of noise, these tend to have many different recurrent classes (i.e. areas of the state space that cannot be escaped once entered). In such cases, one single (stochastic) decision made by one player may lead the system to one or another recurrent class (and completely change the properties of the resulting dynamics), making the formal analysis of these models very challenging (see e.g. section 5.4). The inclusion of some kind of noise (e.g. mutations or trembling hands) tends to simplify the analysis to a great extent, since it often means that all the states of the system communicate (and this most often implies that the stochastic process is ergodic). On a slightly more negative note, this fact also demonstrates that very small changes in the assumptions of a model may have quite an important effect on its dynamics. In any case, this thesis has illustrated that the theory of Markov processes can be particularly useful to analyse formal models of social interactions, and it has also provided various indications on which specific mathematical results may be most valuable depending on the properties of the system to be analysed (see e.g. sections 3.2.2 and 4.1).

8.3. Areas for future work

8.3.1. Assessment of the philosophical foundations of game theory

As noted by some authors (see e.g. Hargreaves Heap and Varoufakis, 1995, pp. 14-18), game theory is rooted in philosophical foundations that are not free from controversy. One of the most contentious issues in this regard concerns the concept of instrumental rationality used in classical game theory (see section 2.2.2). Critically studying the philosophical foundations of game theory seems to be a matter of great importance for at least two reasons: because most economists and many game theorists seem to be almost unaware that the foundations of game theory are at the very least debatable, and because a richer notion of rationality may provide game theory with the intuitive appeal and logical coherence that some of its analyses lack (Hargreaves Heap and Varoufakis, 1995, p. 14). This thesis in particular (see section 7.5) has outlined the basis of a potential line of future research based on a new form of reasoning, i.e. reasoning by outcomes. This proposed area of research could potentially lead to more plausible solution concepts that could capture more of the intuitional knowledge (i.e. heuristics) that people seem to implicitly use in their social interactions.

8.3.2. Learning algorithms vs. Rationality

As explained in section 2.4.1, a current limitation of learning game theory is that most models assume that every player in the game follows the same decision-making algorithm. Thus, in many of these models the observed dynamics may be very dependent on the fact that the game is played among “cognitive clones”, and the extent of this effect is not often evaluated. Confronting the investigated learning algorithms with alternative decision-making algorithms seems to be a promising way forward in learning game theory. In particular, confronting learning algorithms with highly rational players seems to have the potential to be very illuminating.

8.3.3. Evolution of learning algorithms

As explained in section 2.4, one of the main differences between evolutionary and learning game theory is the level at which adaptation takes place[41]. Adaptation processes in evolutionary models occur at the population level: populations are subject to evolutionary pressures (and therefore the population adapts), but the individual components of populations may not adapt at all (i.e. they may have a predefined fixed behaviour). On the other hand, adaptation processes in learning models take place at the individual level through learning, and it is this learning process that is formally described[42]. Most current efforts to integrate these two branches of game theory aim at drawing similarities between the (mean-field) dynamics of certain learning algorithms and an appropriate version of the replicator dynamics (see e.g. Börgers and Sarin (1997), Laslier et al. (2001), Hopkins (2002), Laslier and Walliser (2005), Hopkins and Posch (2005), Beggs (2005)). A complementary (and less pursued) way in which these two branches can be integrated to some extent consists in analysing models that incorporate adaptation processes both at the individual and at the population level, i.e. studying the evolution of different learning algorithms (Kirchkamp, 1999, 2000). Playing with the relative strength of these two levels at which adaptation may take place is likely to offer new insights on the conditions that may favour the evolutionary emergence of certain reasoning processes over others.

8.3.4. Stochastic approximation theory

This thesis and a significant number of papers in the literature (see the brief review presented in section 4.1) have benefited immensely from recent developments in the theory of stochastic approximation. This theory is devoted, in particular, to identifying the conditions under which the actual dynamics of a stochastic system can be approximated by an appropriately constructed deterministic model. Further developments in the theory of stochastic

[41] Another important difference relates to the interpretation of payoffs in each of these branches of game theory (see section 2.1).

[42] Another difference between these two branches of game theory relates to the *nature* of the adaptation process that is modelled. Adaptation in evolutionary models takes place through processes of selection and mutation (see section 2.3), while this is not necessarily the case in learning models (see section 2.4).

approximation theory will undoubtedly enable game theorists to better understand their models, and also to analyse the dynamics of models that were previously intractable. Furthermore, developing our understanding of the relations between stochastic and deterministic models is likely to provide new insights on the relation between learning and evolutionary game theory (Weibull, 2002).

8.3.5. Development of frameworks

This thesis has extensively argued for the value of frameworks at several points (see e.g. sections 2.4.1, 6.1 and 7.4). The wide variety of models developed in the last few years in game theory calls for the creation of frameworks aimed at facilitating the process of model comparison, both in terms of their assumptions and in terms of the results obtained with them. As argued in section 7.4, the development of frameworks is useful not only to assess the impact of various assumptions in theoretical terms, but also to inform experimental research. Thus, the use of frameworks may facilitate the interaction between game theorists and other social scientists, an area for future work that is outlined below.

8.3.6. Greater interaction with other social sciences

There is clearly a lot to gain from the interaction of game theory and other social sciences. Traditionally, game theory has developed almost entirely from introspection and theoretical concerns. Whilst the work developed in game theory up until now has proven to be tremendously useful, it seems clear that game theory will not fulfil all its potential as a useful practical tool to analyse real-world social interactions unless a greater effort is made to interact with other social sciences. In particular, a closer interaction with more empirically-driven social scientists is likely to increase the applicability and relevance of game theory for the study of real-world social interactions. Ideally, this interaction should not be postponed until the stage in the research where a theoretical model is to be validated; on the contrary, empirical research (both experimental and field work) can suggest exciting and relevant avenues where theoretical research may be most needed. In this way, empirical and theoretical work can usefully drive, shape, and benefit from each other. As Weibull (2002) says, "perhaps this is the beginning of a new phase in economic research where economists get together with psychologists, sociologist, and social anthropologists". Let us make it happen.

Appendix A. Proofs of propositions in chapter 4

***Notation*:** Since most of the proofs follow Norman (1968) we adopt his notation. The state of the system in iteration n, characterized in the BM model by the mixed-strategy profile in iteration n, is denoted S_n. The set of possible states is called the *state space* and denoted S. The realization of both players' decisions in iteration n is referred to as an event and denoted E_n. The set of possible events is called the *event space* and denoted E. S_n and E_n are to be considered random variables. In general, s and e denote elements of the state and event spaces, respectively. The function of S into S that maps S_n into S_{n+1} after the occurrence of event e is denoted $f_e(\cdot)$. Thus, if $E_n = e$ and $S_n = s$, then $S_{n+1} = f_e(s)$. Let $T_n(s)$ be the set of values that S_{n+1} takes on with positive probability when $S_1 = s$. Let us say that a state *s is associated* with an event e if s is a pure state (where all probabilities are either 0 or 1) and the occurrence of e pushes the system towards s from any other state. In any system, only one state is associated with a certain event, but the same state may be associated with several events. Finally, use $d(A, B)$ for the minimum Euclidean distance between two subsets A and B of S.

$$d(A,B) = \inf_{s\in A, s'\in B} d(s,s')$$

Lemma 1. Assuming players' aspiration levels are different from their respective payoffs, the 2-player 2-strategy BM model can be formulated as a strictly distance diminishing model (Norman, 1968, p.64).

Proof. Proving that the BM model can be formulated as a strictly distance diminishing model involves checking that hypotheses H1 to H8 in Norman (1968) hold. Define the state of the system S_n in iteration n in the BM model as the mixed-strategy profile in iteration n. The state space is then the mixed-strategy space of the game, and the event space E is the space of pure-strategy profiles, or possible outcomes of the game; consider also the Euclidean distance $d(s, s')$ in S. Having stated that, hypotheses H1 to H6 (which are included here for the sake of completeness) are immediate:

H1. The occurrence of an event effects a change of state such that if $E_n = e$ and $S_n = s$, then $S_{n+1} = f_e(s)$. Thus, $S_{n+1} = f_{E_n}(S_n)$ for $n \geq 1$.

H2. E is a finite set.

H3. The learning situation is memory-less and temporally homogeneous, in the sense that the probabilities of the various possible events on trial *n* depend only on the state on trial *n*, and not on earlier states or events, or on the trial number. That is, there is a real valued function $\varphi.(\cdot)$ on $E \times S$ such that

$$P_s(E_1 = e_1) = \varphi_{e_1}(s) \ ,$$

and $$P_s(E_{n+1} = e_{n+1} \mid E_j = e_j, \ 1 \le j \le n) = \varphi_{e_{n+1}}(f_{e_1 \ldots e_n}(s)) \quad , \text{ for } n \ \geq \ 1,$$

where $f_{e_1 \ldots e_n}(s) = f_{e_n}(f_{e_{n-1}}(\ldots(f_{e_1}(s)))$

H4. (*S*,*d*) is a metric space.

H5. (*S*,*d*) is compact.

H6. Let us use the following notations. If *h* and *g* are mappings of *S* into the real numbers and into *S*, respectively, their maximum "difference quotients" *m*(*h*) and *u*(*g*) are defined by

$$m(h) = \sup_{s \neq s'} \frac{|h(s) - h(s')|}{d(s,s')} \qquad \text{and} \qquad u(g) = \sup_{s \neq s'} \frac{d(g(s), g(s'))}{d(s,s')}$$

whether or not these are finite. H6 is the following regularity condition:

$$m(\varphi_e) < \infty \quad \text{for all } e \in E$$

This is easily proven by defining $\varphi_e(s) \equiv d(s,0)$

H7. For strictly distance diminishing models H7 reads

$$\sup_{s \neq s'} \frac{d(f_e(s), f_e(s'))}{d(s,s')} < 1 \qquad \text{for all } e \in E$$

Given that learning rates are strictly within 0 and 1 and stimuli are always non-zero numbers between −1 and 1 (since players' aspiration levels are different from their respective payoffs by assumption), it can easily be checked that H7 holds. The intuitive idea is that after any event *e*, the distance from any state *s* to the pure state s_e associated with event *e* is reduced by a fixed proportion in each of the components of *s* which is not already equal to the corresponding component in s_e. For the strict inequality in H7 to hold, it is instrumental that every state of the system (except at most one for each event) changes after any given event occurs (i.e. $f_e(s) \neq s$ for all $s \neq s_e$). The assumption that players' aspiration levels are different from their respective payoffs guarantees such a requirement. Without that assumption, H7 does not necessarily hold in its strict form.

Finally, H8 reads:

H8. For any $s \in S$ there is a positive integer k and there are k events e_1 ,…, e_k such that

$$\sup_{s \neq s'} \frac{d(f_{e_1 \dots e_n}(s), f_{e_1 \dots e_n}(s'))}{d(s,s')} < 1 \quad \text{and} \quad P(E_j = e_j, 1 \leq j \leq n \mid S_1 = s) > 0$$

where $f_{e_1 \dots e_n}(s) = f_{e_n}(f_{e_{n-1}}(\dots(f_{e_1}(s)))$

H8 is immediate having proved H7 in its strict form, since at least one event is possible in any state.█

Lemma 2. Consider any 2-player 2-strategy BM system where players' aspiration levels differ from all their respective payoffs. Let s_e be the state associated with event e. If e may occur when the system is in state s ($\Pr\{E_n = e \mid S_n = s\} > 0$), then

$$\lim_{n \to \infty} d(T_n(s), s_e) = 0$$

Proof. The BM model specifications guarantee that if event e may occur when the system is in state s, then it will also have a positive probability of happening in any subsequent state. Mathematically,

$$\Pr\{E_n = e \mid S_n = s\} > 0 \quad \rightarrow \quad \Pr\{E_{n+t} = e \mid S_n = s\} > 0 \qquad \text{for any } t \geq 0$$

This means that any finite sequence of events {e, e…e} has positive probability of happening. Note now that if the system is in state $s \neq s_e$ and event e occurs, the distance from s to s_e is reduced by a fixed proportion in each of the components of s which is not already equal to the corresponding component in s_e. This proportion of reduction is, for each player, the product of the player's absolute stimulus magnitude generated after e and the player's learning rate. Both proportions are strictly between 0 and 1 since players' aspiration levels are different from their respective payoffs by assumption. Let k be the minimum of those two proportions. Imagine then that event e keeps occurring, and note the following bound.

$$d(T_n(s), s_e) \leq (1-k)^n \cdot d(s, s_e)$$

The proof is completed taking limits in the expression above.

$$0 \leq \lim_{n \to \infty} d(T_n(s), s_e) \leq \lim_{n \to \infty} (1-k)^n \cdot d(s, s_e) = 0 \quad █$$

Proof of Proposition 4-1. Statement (i) is an application of Theorem 1 in chapter 2 of Benveniste et al. (1990, p. 43). Statement (ii) follows from Theorem 8.1.1 in Norman (1972, p. 118). The assumptions to apply this Theorem are listed in Norman (1972, p. 117). Here we show that with the hypotheses in Proposition 4-1, all these assumptions hold. In this section, following Norman (1972), the state of the system in iteration n is denoted X_n, and the letter θ denotes the learning rate. Since the state space $I_\theta = I$ is independent of θ, (a.1) is satisfied. $H_n^\theta = \Delta X_n^\theta / \theta$ does not depend on θ, so (a.2) and (a.3) hold. All components of the functions $w(x) = \mathrm{E}(H_n^\theta | X_n^\theta = x)$ and $s(x) = \mathrm{E}((H_n^\theta - w(x))^2 | X_n^\theta = x)$ are polynomials, so every assumption (b) is satisfied. Finally, since H_n^θ does not depend on θ the supremum over θ can be omitted in (c), and also the module of each of the components of H_n^θ is bounded by the maximum learning rate, so (c) is also satisfied. Thus Theorem 8.1.1 is applicable. Finally, Statement (iii) is an application of Theorem 4.1 in chapter 8 of Kushner and Yin (1997).█

Proof of Proposition 4-2. Proposition 4-2 follows from Theorem 2.3 in Norman (1968, p.67), which requires the model to be distance-diminishing and one extra assumption H10.

H10. There are a finite number of absorbing states a_1 ,…, a_N , such that, for any $s \in S$, there is some $a_{j(s)}$ for which

$$\lim_{n\to\infty} d(T_n(s), a_{j(s)}) = 0$$

Given the assumptions of Proposition 4-2, Lemma 1 can be used to assert that the BM model is distance diminishing, with associated stochastic processes $\boldsymbol{S}_n$ and $\boldsymbol{E}_n$. Proving that H10 prevails will then complete the proof. The proof of H10 rests on the following three points:

a) If in state s there is a positive probability of an event e occurring, then, applying Lemma 2:

$$\lim_{n\to\infty} d(T_n(s), s_e) = 0$$

where s_e is the state associated with the event e.

b) The state s_e associated with a Mutually Satisfactory (MS) event e is absorbing. Note also that there are at most four absorbing states.

c) From any state there is a positive probability of playing a MS event within three iterations.

Points (a) and (b) are straightforward. To prove (c) we define strictly mixed strategies as those that assign positive probability to both actions, and mixed states as states where both players' strategies are strictly mixed. Note that after an unsatisfactory event, every player modifies her strategy so the updated strategy is strictly mixed, and that strictly mixed strategies will always remain so.

Since players' aspiration levels are below their respective *maximin* by assumption, there is at least one MS event. Hence from any mixed state there is a positive probability for a MS event to happen. We focus then on non-mixed states where no MS event can occur in the first iteration. This implies that the event in the first iteration is unsatisfactory for at least one player, so at least one player will have a strictly mixed strategy in the second iteration. Without loss of generality let us say that player 1 has a strictly mixed strategy in the second iteration. If player 2's strategy were also strictly mixed, then the state in the second iteration would be mixed, and a MS event could occur. Imagine then that the state in the second iteration is not mixed. Given that player 1's aspiration is below its *maximin*, there is a positive probability that the event in iteration 2 will be satisfactory for player 1. If such a possible event is also satisfactory for player 2, an MS event has occurred. If not, then both players will have a strictly mixed strategy in iteration 3, so a MS event could happen in iteration 3. This finishes the proof of point (c).

The proof of the fact that every SRE can be asymptotically reached with positive probability if the initial state is completely mixed rests on two arguments: (a) there is a strictly positive probability that an infinite sequence of any given MS event *e* takes place (this can be proved using Theorem 52 in Hyslop (1965, p.94)), and (b) such an infinite run would imply convergence to the associated (SRE) state s_e. We also provide here a theoretical result to estimate with arbitrary precision the probability L_∞ that an infinite sequence of a MS event $\boldsymbol{e} = (d_1, d_2)$ begins when the system is in mixed state $\boldsymbol{p} = (p_{1,d_1}, p_{2,d_2})$.

$$L_\infty = \lim_{n\to\infty}\prod_{t=0}^{n}[1-(1-p_{1,d_1})(1-l_1 s_1(d_1))^t][1-(1-p_{2,d_2})(1-l_2 s_2(d_2))^t]$$

The following result can be used to estimate L_∞ with arbitrary precision:

Let $P_k = \prod_{t=0}^{k-1}(1-xy^t)$ and let $P_\infty = \lim_{k\to\infty} P_k$. Then, for $x, y \in (0, 1)$,

$$P_k > P_\infty > P_k(1-\frac{xy^k}{1-y})$$

We are indebted to Professor Jörgen W. Weibull for discovering and providing the lower bound in this result (personal communication).■

Proof of Proposition 4-3. Each statement of Proposition 4-3 will be proved separately. Statement (i) is an immediate application of Theorem 2.3 in Norman (1968, p.67), which requires the model to be distance-diminishing and the extra assumption H10 (see proof of Proposition 4-2). Having proved in Lemma 1 that the model is distance-diminishing, we prove here that H10 holds. The proof of H10 rests on the same three points (a-c) exposed in the proof of Proposition 4-2. The terminology defined there is also used here. Points (a) and (b) are straightforward. To prove (c), remember that after an unsatisfactory event, every player modifies her strategy so the updated strategy is strictly mixed, and that strictly mixed strategies always remain so. By assumption, there is at least one absorbing state, which means that there must be at least one MS event. This implies that from any mixed state there is a positive probability for a MS event to happen.

Since players' aspirations are above their respective *maximin*, given any action for player *i*, there is always an action for her opponent such that the resulting event would be unsatisfactory for player *i*. In other words, if one of the players has a strictly mixed strategy, then there is a positive chance that the system will be in a mixed state in the next iteration. We focus then on states where no player has strictly mixed strategies and a MS event cannot occur in the first iteration. This implies that the event in the first iteration is unsatisfactory for at least one player, who will have a strictly mixed strategy in the second iteration and, as just shown, this implies a positive probability that the system will be in a mixed state in the third iteration. The proof of statement (i) is then finished.

Statement (ii) follows from Theorem 2.2 in Norman (1968, p.66), which requires the model to be distance-diminishing and one extra assumption H9.

H9. $\lim_{n\to\infty} d(T_n(s), T_n(s')) = 0$ for all $s, s' \in S$

Having proved in Lemma 1 that the model is distance-diminishing, we prove here that H9 holds. Since, by assumption, there are no absorbing states, there cannot be MS events. This implies that the event in the first iteration is unsatisfactory for at least one player, who will have a strictly mixed strategy in the second iteration. As argued in the proof of statement (i), this implies a positive probability that the system will be in a mixed state in the third iteration. Therefore at the third iteration any event has a positive probability of happening, so we can select any one of them, the state s_e associated with it, and then, by Lemma 2, we know that $\lim_{n\to\infty} d(T_n(s), s_e) = 0$ for any state s, so H9 holds. ■

<u>Proof of Proposition 4-4</u>. The reasoning behind this proof follows Sastry et al. (1994). Statement (i) can be proved considering one player i who benefits by deviating from the SRE by increasing her probability $p_{i,q}$ to conduct action q. The expected change in probability $p_{i,q}$ can then be shown to be strictly positive for all $p_{i,q} > 0$ while keeping the other player's strategy unchanged. Statement (ii) can be proved considering the Jacobian of the linearization of ODE [2]. Without loss of generality, assume that $Y_i = \{\text{A, B}\}$ and the certain outcome at the SRE is $\boldsymbol{y}_{\mathbf{SRE}}$ = (A, A). Choose $p_{1,\text{B}}$ and $p_{2,\text{B}}$ as the two independent components of the system, so the SRE is $[p_{1,\text{B}}, p_{2,\text{B}}] = [0, 0]$. The Jacobian J at the SRE is then as follows:

$$J = \begin{bmatrix} l_1(\delta(s_1(\text{B,A})) - s_1(\text{A,A})) & l_1 \cdot \delta(-s_1(\text{A,B})) \\ l_2 \cdot \delta(-s_2(\text{B,A})) & l_2(\delta(s_2(\text{A,B})) - s_2(\text{A,A})) \end{bmatrix}$$

$$\text{where } \delta(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases}$$

It is then straightforward that if $\boldsymbol{y}_{\mathbf{SRE}}$ = (A,A) is a mutually satisfactory (s_i(A,A) > 0) strict Nash equilibrium (s_1(A,A) > s_1(B,A); s_2(A,A) > s_2(A,B)) and at least one unilateral deviation leads to a satisfactory outcome for the non-deviating player (s_1(A,B) ≥ 0 or s_2(B,A) ≥ 0), then the two eigenvalues of J are negative real, so the SRE is asymptotically stable.■

Notes to extend the theoretical results to populations of players. All the lemmas and propositions in chapter 4 and this appendix can be easily extended to finite populations from which two players are randomly drawn to play a 2×2 game taking into account the following points: (1) the state of the system S_n in iteration n is the mixed-strategy profile of the whole population. (2) An event E_n in iteration n comprises an identification of the two players who have played the game in iteration n and their decisions. (3) Pure states are now associated (in the sense given in the notation of the appendix) with *chains* of events, rather than with single events. A pure state s is associated with a finite chain of events c (where every player must play the game at least once) if the occurrence of c pushes the system towards s from any other state.

Proof of Proposition 4-5. Let Θ be the *mixed-strategy space* of the finite normal-form game. The proof consists in applying Brouwer's Fixed Point theorem to the function $\mathbf{W}(\boldsymbol{p}) \equiv \mathbf{E}(\boldsymbol{P}_{n+1} \mid \boldsymbol{P}_n = \boldsymbol{p})$ that maps the mixed-strategy profile $\boldsymbol{p} \in \Theta$ to the *expected* mixed-strategy profile $\mathbf{W}(\boldsymbol{p})$ after the game has been played once and each player has updated her strategy $\boldsymbol{p}_i$ accordingly. Since the mixed-strategy space Θ is a non-empty, compact, and convex set, it only remains to show that $\mathbf{W}:\Theta \to \Theta$ is a continuous function. Let $\mathbf{w}_i(\boldsymbol{p})$ be the *i*th component of $\mathbf{W}(\boldsymbol{p})$, which represents player *i*'s expected strategy for the following iteration. Therefore:

$$\mathbf{w}_i(\boldsymbol{p}) = \sum_{y \in Y} \Pr\{\boldsymbol{d}\} \cdot \boldsymbol{r}_i^y(\boldsymbol{p}) = \sum_{y \in Y} (\prod_{i \in I} p_{i,y_i}) \cdot \boldsymbol{r}_i^y(\boldsymbol{p})$$

Since all $\boldsymbol{r}_i^y(\boldsymbol{p})$ are continuous for every $\boldsymbol{y}$ and every i by hypothesis, $\mathbf{W}(\boldsymbol{p})$ is also continuous. Thus, applying Brouwer's Fixed-Point theorem, we can state that there is at least one $\boldsymbol{p}^* \in \Theta$ such that $\mathbf{W}(\boldsymbol{p}^*) = \boldsymbol{p}^*$. This means that the *expected change* in all $(p_{i,j})^*$ (probability of player i following her jth pure strategy) is zero. ∎

Appendix B. Proofs of propositions in chapter 5

Proof of Proposition 5-1. Proving the second part of proposition 5-1 –i.e. that the asymptotic behaviour of the N-CBR model is independent of the decision-making algorithm employed by each player *i* when she has not yet explored every action available to her in a *similar* situation– is straightforward, since this is a transient situation. Given the definition of the set of different states of the world possibly perceived by any player, the trembling hands noise guarantees that sooner or later every possible state of the world perceived by any player will happen infinitely often. The trembling hands noise also guarantees that every player will choose every possible action available to her in any given situation. Thus, sooner or later, every player will have selected every action available to her in every possible state of the world she can perceive (i.e. every action available to player *i* will be represented in her set of cases C_i, for every state of the world possibly perceived by *i*). Therefore, sooner or later, no player will be using the decision-making algorithm that the second part of proposition 5-2 refers to, so the asymptotic behaviour of the model is independent of such algorithms.

The following proves part 1 of proposition 5-1, i.e. that if every player has a common perception of the state of the world, then the asymptotic behaviour of the N-CBR process is independent of the specific structure of the perceived state of the world. The previous paragraph demonstrates that sooner or later the state of the system in the N-CBR model is fully characterised by every player's set of most recent cases that occurred in every possible perceived state of the world for each one of the actions available to her. Thus, this second proof (which refers to the asymptotic behaviour of the system) assumes that every player has already selected every action available to her at least once in every possible state of the world she can perceive. Consider the following two points:

- The assumption that players have a common perception of the state of the world implies that all players perceive that any particular state of the world has occurred in exactly the same time-steps. In other words, all players would unanimously agree or disagree with any statement of the form "The situations lived in time-steps $\{x, y,\ldots,z\}$ looked all similar to me (i.e. they correspond to the same perceived state of the world)".

- Note also that the decision made by each player *i* in any particular situation is only affected by decisions (made by all players) that took place in a previous *similar* situation (i.e. having perceived the same state of the world).

Thus, one can view the dynamics of the whole model (where players can perceive various different states of the world) as a collection of parallel dynamic processes, each of them corresponding to one specific state of the world (perceived by all players at once). The dynamics observed for each individual perceived state of the world are governed by the same decision-making processes and are independent of each other. Each of these individual threads, if observed on its own, induces the same dynamics that one would observe in a model where players cannot distinguish between different states of the world. The following table illustrates this interpretation with an example.

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
SW_t	sw_3	sw_1	sw_4	sw_3	sw_2	sw_3	sw_4	sw_4	sw_1	sw_2	sw_1	sw_3	sw_2	sw_4	sw_1
THREAD $SW = sw_1$		1							2		3				4
THREAD $SW = sw_2$					1					2			3		
THREAD $SW = sw_3$	1			2		3						4			
THREAD $SW = sw_4$			1				2	3						4	

where SW_t is the random variable that denotes the state of the world perceived by every player at time-step t, sw_i are particular values of that variable, and the numbers on coloured backgrounds inside the table indicate the number of times that the corresponding state of the world has been visited.

Let $T_n^{(sw)}$ be the state of the thread $\{SW = sw\}$ (where the perceived state of the world is sw), defined by the payoffs each player obtained the last time that she selected each of the actions available to her having observed state of the world sw, when state of the world sw has been observed n times. It is clear then that the sequence of random variables $\{T_n^{(sw)}\}_{n\geq 1}$ (for any fixed sw) corresponds to a model where players cannot distinguish between different states of the world. Following the reasoning presented in the first paragraph of section 5.7, it is also straightforward to show that $\{T_n^{(sw)}\}_{n\geq 1}$ can be formulated as a uni-reducible Markov chain, which has a unique limiting distribution (Janssen and Manca, 2006, Corollary 5.2, pg. 117). Finally, it should also be apparent that all threads have the same limiting distribution:

$$\lim_{n\to\infty} \mathbf{Pr}(T_n^{(i)} = \alpha) = \lim_{n\to\infty} \mathbf{Pr}(T_n^{(j)} = \alpha) \quad \forall i, j$$

For clarity of notation, let $\{T_n\}_{n\geq 1}$ denote the sequence of states corresponding to a model where players cannot distinguish different states of the world. Thus,

$$\lim_{n\to\infty} \mathbf{Pr}(T_n^{(i)} = \alpha) = \lim_{n\to\infty} \mathbf{Pr}(T_n^{(j)} = \alpha) = \lim_{n\to\infty} \mathbf{Pr}(T_n = \alpha) \qquad \forall i, j$$

The fact that remains to be proven is that the overall dynamics of the model (i.e. the ensemble of threads) also show the same limiting distribution as the individual threads. To show that, let X_t denote the state of the thread corresponding to the state of the world observed at time t. Formally:

$$X_t = \left\{T_{N_i(t)}^{(i)} : SW_t = i\right\}$$

where $N_i(t)$ denotes the number of times that the event $\{SW_t = i\}$ has occurred up until time-step t. Formally: $N_i(t) = \#\{\, k \in \{1,...,t\} : SW_k = i\}$

With this notation, the proof of the second part of proposition 5-2 will be concluded once it is demonstrated that:

$$\lim_{t\to\infty} \mathbf{Pr}(X_t = \alpha) = \lim_{t\to\infty} \mathbf{Pr}(T_t = \alpha)$$

The following, which is conditioned to a set of (arbitrary) initial conditions, concludes the proof.

$$\lim_{t\to\infty} \mathbf{Pr}(X_t = \alpha) = \lim_{t\to\infty} \sum_i \mathbf{Pr}(\{SW_t = i\} \ \& \ \{T_{N_i(t)}^{(i)} = \alpha\}) =$$

$$= \lim_{t\to\infty} \sum_i \mathbf{Pr}(T_{N_i(t)}^{(i)} = \alpha \mid SW_t = i) \cdot \mathbf{Pr}(SW_t = i)$$

It has been argued previously that states of the world are visited infinitely often, thus:

$$\lim_{t\to\infty} \mathbf{Pr}(T_{N_i(t)}^{(i)} = \alpha \mid SW_t = i) = \lim_{n\to\infty} \mathbf{Pr}(T_n^{(i)} = \alpha) = \lim_{n\to\infty} \mathbf{Pr}(T_n = \alpha)$$

(Regardless of the set of (arbitrary) initial conditions)

and it is also clear that $\sum_{i} \mathbf{Pr(}SW_t = i\mathbf{)} = 1 \qquad \forall t$

Using the two results above the first part of proposition 5-1 is finally proved:

$$\lim_{t\to\infty} \mathbf{Pr(}X_t = \alpha\mathbf{)} = \lim_{t\to\infty} \sum_{i} \mathbf{Pr(}T^{(i)}_{N_i(t)} = \alpha \mid SW_t = i\mathbf{)} \cdot \mathbf{Pr(}SW_t = i\mathbf{)} = \lim_{t\to\infty} \mathbf{Pr(}T_t = \alpha\mathbf{)}$$

Proof of Proposition 5-2. As argued in the proof of proposition 5-1, sooner or later, every player will have selected every action available to her in every possible state of the world she can perceive (i.e. every action available to player *i* will be represented in her set of cases C_i, for every state of the world possibly perceived by *i*). Thus, sooner or later, the state of the system in the N-CBR model is fully characterised by every player's set of most recent payoffs she obtained for each one of the actions available to her in every possible state of the world she can perceive. 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 $\varepsilon = 0$ and all players have explored all their available actions for every possible state of the world they can perceive. Note that P^0 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 (i.e. the model without noise). The second argument is that the *outcome* (i.e. the set of decisions made by players) that is induced by any recurrent *state* of P^0 is necessarily individually rational. The following proves an alternative (but equivalent) formulation of the second argument: if state *x* in P^0 induces an outcome that is not individually rational, then *x* is a transient state of P^0. We will prove this second argument by showing that if state *x* induces an outcome that is not individually rational, then *x* will never be revisited.

Let *A* be one of the players who has received a payoff below her maximin $Maximin_A$ in the outcome induced by state *x*, and let sw_A be the state of the world perceived by *A* in state *x*. Let *a* be the action that *A* chose in state *x*, and $p_x(A, a)$ be the payoff that *A* had obtained the previous time she had perceived sw_A and selected action *a*; this payoff $p_x(A, a)$ is part of the definition of *x*. Note that a necessary condition for *x* to be revisited is that player *A* perceives sw_A again, and also that the payoff that *A* has obtained the previous time she has perceived sw_A and selected action *a* is $p_x(A, a)$. This can never be the case for the following argument:

1. The fact that player *A* selected action *a* in state *x* implies that $p_x(A, a) \geq Maximin_A$. In more informal terms, the payoff player *A* believed she would obtain by selecting action *a* (having observed state of the world sw_A) was the maximum over all her possible actions, and therefore it was necessarily no less than $Maximin_A$.
2. Player *A* obtained a payoff strictly below her $Maximin_A$ when, after having perceived state of the world sw_A, she selected action *a*. Thus, from then onwards she will remember that the last time she selected action *a* having observed state of the world sw_A, she obtained a payoff strictly below $Maximin_{A.}$
3. There is at least one action that gives player *A* a payoff no less than $Maximin_A$ regardless of the actions of her counterparts. When perceiving state of the world sw_A again, player *A* will always select this (maximin) action over action *a*. Thus, player *A* will never update her belief that selecting action *a* when she perceives state of the world sw_A will give her a payoff below $Maximin_A$.

State *x* required player *A* to believe that selecting action *a* would give her a payoff no less than $Maximin_A$. Thus, state *x* cannot be revisited, and this fact concludes the proof.

Supporting Material

All the software, parameter files, and documentation required to easily replicate every experiment presented in this thesis are included in the accompanying CD. This section outlines the file structure of this CD.

The root directory contains 3 folders, one for each of the chapters where results from computational experiments are presented:

Folder "chapter4"

This folder contains an HTML file named "index.html" that can be used to easily access every program that was used to create each of the figures in chapter 4. All these programs were coded using Mathematica©. There is no need to make any alterations to the source code to obtain each of the figures presented in chapter 4.

Folder "chapter5"

This folder contains the following files and directories:

- "analyticalCalculation.nb" is the Mathematica© program used in section 5.7.3 to identify features that make outcomes stochastically stable. As explained in section 5.7.3, this program also calculates the exact long-run fraction of time that any 2-player system spends in each possible outcome when the probability of trembles tends to zero.
- "CBR-model" is a directory that contains an Objective-C implementation of the CBR-model, a detailed user guide that explains how to use the model (casd-0-userGuide.pdf), and several parameter files for demonstration.
- "N-CBR-model" is a directory that contains the Objective-C implementation of the specific N-CBR model that was used to produce figure 5-8, and several parameter files for demonstration.
- "dataForFigures" is a directory that contains all the parameter files and the data that were used to generate each of the figures in chapter 5.

Folder “chapter6”

This folder contains the following files:

- “index.html” is an HTML file that contains an applet of EVO-2x2 and detailed instructions on how to use it.
- “EVO-2x2.nlogo” is the NetLogo 3.0.2 (Wilensky, 1999) implementation of EVO-2x2. It also contains all the parameter files required to replicate all the experiments presented in chapter 6. These can be accessed using the “BehaviorSpace” tool that forms part of NetLogo.
- “EVO-2x2-3D.nlogo” is the NetLogo 3-D Preview 1 (Wilensky, 1999) implementation of EVO-2x2-3D.
- “NetLogoLite.jar” is a file required to run the applet in the HTML file “index.html”.
- “extraSoftware” is a directory that contains the Perl script (“trimmer.pl”) and the Mathematica© program (“graphGenerator-1.nb”) used to conduct the automatic analyses explained in section 6.3.3.

Glossary of game theory terms

Action: a pure strategy.

Deficient equilibrium: An equilibrium is deficient if there exists another outcome which is preferred by every player.

Common interest game: a game where there is a unique *payoff profile* that strongly Pareto dominates all other payoff profiles (and this payoff profile may be achieved via several strategy profiles). See Aumann and Sorin (1989).

Common knowledge: Common knowledge (CK) in game theory often comes with a certain order: zero-order CK of X is just the assumption that X prevails for every player (e.g. zero-order common knowledge of complete information (CKCI) means that every player has complete information); first-order CK is the assumption that every player knows that X prevails for every player (e.g. first-order CKCI means that every player knows that every player has complete information); in general, (n)th-order CK is the assumption that (n-1)th-order CK is known by every player. If no order is specified, it is assumed that the order is infinite (this produces an infinite recursion of shared assumptions).

Common knowledge of rationality (CKR): Following the definition of common knowledge (see above), first-order CKR is the assumption that every player knows that every player is rational; (n)th-order CKR is the assumption that (n-1)th-order CKR is known by every player. If no order is specified, it is assumed that the order of CKR is infinite. See Aumann (1976) for a formal definition.

Complete information: In a game of complete information it is assumed that players not only know the rules of the game and their own payoff function, but also their counterparts' payoff functions (see section 2.2.1).

Evolutionary stable strategy: Informally, an evolutionarily stable strategy is a strategy which, if adopted by a population of players, cannot be invaded by any alternative strategy (see section 2.3.1).

Finite game: a game with finitely many players, each of which has a finite set of pure strategies.

Individually-rational outcome: An outcome giving each player at least their maximin payoff, i.e. the largest payoff that they can guarantee obtaining (regardless of the opponents' moves) in a single-stage game using pure strategies.

Instrumentally rational: An instrumentally rational player has unlimited computational capacity devoted to maximise her individual payoff function. There are various degrees of rationality in game theory; see section 2.2.2.

Maximin payoff: the largest possible payoff a player can guarantee herself (regardless of the opponents' moves) in a single-stage game using pure strategies. The maximin payoff for each player in the one-shot Prisoner's Dilemma is the payoff obtained when both players defect.

Mixed strategy: A probability distribution *P* over the set of pure strategies. It is understood that a player using a mixed strategy chooses one pure strategy randomly according to *P*.

Mutual belief: A proposition *X* is mutual belief among a set of players if each player believes *X*. Mutual belief by itself implies nothing about what, if any, beliefs anyone attributes to anyone else (Vanderschraaf and Sillari, 2007).

Mutual interest game: a game where there exists a unique *pure strategy profile* that gives the highest possible payoff to every player. All mutual interest games are, in particular, common interest games (Aumann and Sorin, 1989).

NxM game: A normal form game for two players, where one player has N possible actions and the other one has M possible actions. The payoff function in NxM games can be neatly represented with a matrix.

Nash equilibrium (Nash, 1951): a set of strategies such that no player, knowing the strategy of the other(s), could improve her expected payoff by changing her own strategy. Every finite game has at least one Nash Equilibrium (possibly in mixed strategies).

Outcome: a particular combination of pure strategies, one for each player, and their associated payoffs.

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

Perfect information: Informally, in (sequential) games of perfect information, the actions taken by every player are instantaneously known by every other player (e.g. chess). Complete information does not imply perfect information.

(Strictly) dominated strategy: For a player A, strategy S_A is (strictly) dominated by strategy S^*_A if for each 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).

Subgame: Informally, a subgame is a subset or piece of a sequential game beginning at some node such that every previous action undertaken by every player at every point is common knowledge.

Subgame perfect equilibrium (Selten, 1975): A strategy profile is a subgame perfect equilibrium if it represents a Nash equilibrium of every subgame of the original game (whether or not the subgame is reached along the equilibrium path induced). Subgame perfect equilibrium is a refinement of the concept of Nash equilibrium that eliminates non-credible threats in sequential games.

Tit-for-Tat (TFT): This is the strategy consisting of starting by cooperating, and thereafter doing what the other player did on the previous move.

List of acronyms

ABM:	Agent-Based Modelling.
AT:	Aspiration Threshold.
BM:	Bush-Mosteller (model).
CBDT:	Case-Based Decision Theory.
CBR:	Case-Based Reasoning.
CGT:	Classical Game Theory.
CK:	Common Knowledge.
CKCI:	Common Knowledge of Complete Information.
CKR:	Common Knowledge of Rationality.
CogGT:	Cognitive Game Theory.
EGT:	Evolutionary Game Theory.
ER:	Erev-Roth (model).
ESS:	Evolutionary Stable Strategy.
EUT:	Expected Utility Theory.
EWA:	Experience Weighted Attraction (model).
FP:	Fictitious Play.
GPL:	General Public Licence.
GRL:	General Reinforcement Learning (framework).
LGT:	Learning Game Theory.
N-CBR:	Noisy Case-Based Reasoning (model).
NE:	Nash Equilibrium.
ODE:	Ordinary Differential Equation.
PC:	Personal Computer.
PD:	Prisoner's Dilemma.
RD:	Replicator Dynamics.
SCE:	Self-Correcting Equilibrium.
SFP:	Smooth Fictitious Play.
SRE:	Self-Reinforcing Equilibrium.
SREUP:	Self-Reinforcing Equilibrium of the associated Unperturbed Process.
TC:	Tragedy of the Commons (game).
TFT:	Tit For Tat (strategy).
UML:	Unified Modelling Language

List of figures

List of tables

References

Aamodt, A. and Plaza, E. (1994). "Case-based reasoning: foundational issues, methodological variations, and system approaches". *AI Communications* 7(1), pp. 39-59.

Akaike, H. (1969). "Fitting autoregressive models for prediction". *Annals of the Institute of Statistical Mathematics* 21, pp. 243-247.

Alexander, J. M. (2003). "Evolutionary Game Theory". In *The Stanford Encyclopedia of Philosophy*, Zalta, E. N. (ed.). Summer 2003 Edition.

Arthur, W. B. (1991). "Designing economic agents that act like human agents: A behavioral approach to bounded rationality". *American Economic Review* 81(2), pp. 353-359.

Aumann, R. (1976). "Agreeing to disagree". *Annals of Statistics* 4(6), pp. 1236-1239.

Aumann, R. J. and Hart, S. (1992). *Handbook of Game Theory with Economic Applications*. Amsterdam: North-Holland.

Aumann, R. J. and Sorin, S. (1989). "Cooperation and Bounded Recall". *Games and Economic Behavior* 1(1), pp. 5-39.

Axelrod, R. (1984). *The Evolution of Cooperation*. Basic Books USA.

Axelrod, R. (1987). "The evolution of strategies in the iterated prisoner's dilemma". In *Genetic Algorithms and Simulated Annealing*, Davis, L. (ed.), pp. 32-41. Los Altos, CA: Morgan Kaufman.

Beggs, A. (2002). "Stochastic evolution with slow learning". *Economic Theory* 19(2), pp. 379-405.

Beggs, A. W. (2005). "On the convergence of reinforcement learning". *Journal of Economic Theory* 122(1), pp. 1-36.

Benaim, M. and Weibull, J. W. (2003). "Deterministic Approximation of Stochastic Evolution in Games". *Econometrica* 71(3), pp. 873-903.

Bendor, J., Diermeier, D. and Ting, M. (2004). "The Empirical Content of Adaptive Models". *Stanford GSB Working Paper*, 1877.

Bendor, J., Diermeier, D. and Ting, M. (2007). "Comment: Adaptive Models in Sociology and the Problem of Empirical Content". *The American Journal of Sociology* 112(5), pp. 1534-1545.

Bendor, J., Mookherjee, D. and Ray, D. (2001a). "Aspiration-Based Reinforcement Learning In Repeated Interaction Games: An Overview." *International Game Theory Review* 3(2-3), pp. 159-174.

Bendor, J., Mookherjee, D. and Ray, D. (2001b). "Reinforcement Learning in Repeated Interaction Games". *Advances in Theoretical Economics* 1(1), Article 3.

Bendor, J. and Swistak, P. (1995). "Types of Evolutionary Stability and the Problem of Cooperation". *Proceedings of the National Academy of Sciences of the United States of America* 92(8), pp. 3596-3600.

Bendor, J. and Swistak, P. (1997). "The evolutionary stability of cooperation". *American Political Science Review* 91(2), pp. 290-307.

Bendor, J. and Swistak, P. (1998). "Evolutionary equilibria: Characterization theorems and their implications". *Theory and Decision* 45(2), pp. 99-159.

Benveniste, A., Métivier, M. and Priouret, P. (1990). *Adaptive Algorithms and Stochastic Approximations*. Berlin: Springer-Verlag.

Bernheim, B. D. (1984). "Rationalizable strategic behavior". *Econometrica* 52(4), pp. 1007-1028.

Binmore, K. (1994). *Playing Fair: Game Theory and the Social Contract*. Cambridge, MA: MIT Press.

Binmore, K. and Samuelson, L. (1993). "An Economist's Perspective on the Evolution of Norms". *Journal of Institutional Theoretical Economics* 150, pp. 45-63.

Binmore, K., Samuelson, L. and Vaughan, R. (1995). "Musical Chairs: Modeling Noisy Evolution". *Games and Economic Behavior* 11(1), pp. 1-35.

Binmore, K. G. (1998). "Review of "The Complexity of Cooperation" by Robert Axelrod". *Journal of Artificial Societies and Social Simulation* 1(1).

Börgers, T. and Sarin, R. (1997). "Learning through reinforcement and replicator dynamics". *Journal of Economic Theory* 77(1), pp. 1-14.

Börgers, T. and Sarin, R. (2000). "Naive reinforcement learning with endogenous aspirations". *International Economic Review* 41(4), pp. 921-950.

Boylan, R. T. (1992). "Laws of large numbers for dynamical systems with randomly matched individuals". *Journal of Economic Theory* 57(2), pp. 473-504.

Boylan, R. T. (1995). "Continuous Approximation of Dynamical Systems with Randomly Matched Individuals". *Journal of Economic Theory* 66(2), pp. 615-625.

Brenan, G. and Lomasky, L. (1984). "Inefficient Unanimity". *Journal of Applied Philosophy* 1, pp. 151-163.

Brown, G. W. (1951). "Iterative Solutions of Games by Fictitious Play". In *Activity Analysis of Production and Allocation*, Koopmans, T. C. (ed.) New York: Wiley.

Bush, R. R. and Mosteller, F. (1955). *Stochastic Models for Learning*. New York: John Wiliey & Son.

Camerer, C. (2003). *Behavioral Game Theory: Experiments on Strategic Interaction*. New York: Russell Sage Foundation.

Camerer, C. and Ho, T. H. (1999). "Experience-weighted attraction learning in normal form games". *Econometrica* 67(4), pp. 827-874.

Cappé, O., Moulines, E. and Rydén, T. (2005). *Inference in Hidden Markov Models*. New York: Springer.

Castellano, C., Marsili, M. and Vespignani, A. (2000). "Nonequilibrium phase transition in a model for social influence". *Physical Review Letters* 85(16), pp. 3536-3539.

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". *Behavioral and Brain Sciences* 26(2), pp. 139-153.

Cross, J. G. (1973). "A Stochastic Learning Model of Economic Behavior". *The Quarterly Journal of Economics* 87(2), pp. 239-266.

Chen, Y. and Tang, F. F. (1998). "Learning and Incentive-Compatible Mechanisms for Public Goods Provision: An Experimental Study". *Journal of Political Economy* 106(3), pp. 633-662.

Dawes, R. M. (1980). "Social Dilemmas". *Annual Review of Psychology* 31, pp. 169-193.

Dawes, R. M. and Thaler, R. H. (1988). "Anomalies: Cooperation". *Journal of Economic Perspectives* 2(3), pp. 187-197.

Dawkins, R. (1989). *The Selfish Gene*. 2nd edition. New York: Oxford University Press.

Diekmann, A. (1985). "Volunteer's dilemma". *Journal of Conflict Resolution* 29, pp. 605-610.

Diekmann, A. (1986). "Volunteer's dilemma: A social trap without a dominant strategy and some empirical results". In *Paradoxical Effects of Social Behavior: Essays in Honor of Anatol Rapoport*, Diekmann, A. and Mitter, P. (eds.), pp. 187-197. Heidelberg and Vienna: Physica-Verlag.

Doebeli, M. and Hauert, C. (2005). "Models of cooperation based on the Prisoner's Dilemma and the Snowdrift game". *Ecology Letters* 8, pp. 748-766.

Doran, J. (1997). "Analogical Problem Solving". In *Artificial Intelligence Techniques: A Comprehensive Catalogue*, Bundy, A. (ed.), p. 4. Springer-Verlag.

Duffy, J. (2006). "Agent-Based Models and Human Subject Experiments". In *Handbook of Computational Economics II: Agent-Based Computational Economics.*, Tesfatsion, L. and Judd, K. L. (eds.), pp. 949-1011. Elsevier.

Edmonds, B. (2000). "The Use of Models - Making MABS More Informative". In *Lecture Notes in Computer Science 1979/2000: Multi-Agent-Based Simulation: Second International Workshop, MABS 2000, Boston, MA, USA, July. Revised and Additional Papers*, Moss, S. and Davidsson, P. (eds.), pp. 269-282.

Edmonds, B. (2006). "The emergence of symbiotic groups resulting from skill-differentiation and tags". *Journal of Artificial Societies and Social Simulation* 9(1), Article 10.

Edmonds, B. and Hales, D. (2003). "Replication, replication and replication: Some hard lessons from model alignment". *Journal of Artificial Societies and Social Simulation* 6(4), Article 11.

Edwards, M., Huet, S., Goreaud, F. and Deffuant, G. (2003). "Comparing an individual-based model of behaviour diffusion with its mean field aggregate approximation". *Journal of Artificial Societies and Social Simulation* 6(4).

Elster, J. (1982). "Marxism, functionalism and game theory". *Theory and Society* 11(4), pp. 453-482.

Ellison, G. (2000). "Basins of Attraction, Long Run Equilibria, and the Speed of Step-by-Step Evolution". *Review of Economic Studies* 67, pp. 17-45.

Erev, I., Bereby-Meyer, Y. and Roth, A. E. (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), pp. 111-128.

Erev, I. and Roth, A. E. (1998). "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria". *American Economic Review* 88(4), pp. 848-881.

Erev, I. and Roth, A. E. (2001). "Simple reinforcement learning models and reciprocation in the prisoner's dilemma game". In *Bounded Rationality: The Adaptive Toolbox*, Gigerenzer, G. and Selten, R. (eds.), pp. 216-231. Cambridge, MA: MIT Press.

Ericsson, K. A. and Simon, H. A. (1980). "Verbal reports as data". *Psychological Review* 87(3), pp. 215-251.

Eshel, I. and Cavalli-Sforza, L. L. (1982). "Assortment of Encounters and Evolution of Cooperativeness". *Proceedings of the National Academy of Sciences of the United States of America* 79(4), pp. 1331-1335.

Feltovich, N. (2000). "Reinforcement-based vs. Belief-based learning models in experimental asymmetric-information games". *Econometrica* 68(3), pp. 605-641.

Flache, A. and Hegselmann, R. (1999). "Rationality vs. Learning in the Evolution of Solidarity Networks: A Theoretical Comparison". *Computational & Mathematical Organization Theory* 5(2), pp. 97-127.

Flache, A. and Macy, M. W. (2002). "Stochastic collusion and the power law of learning: A general reinforcement learning model of cooperation". *Journal of Conflict Resolution* 46(5), pp. 629-653.

Foster and Young (1990). "Stochastic evolutionary game dynamics". *Theoretical Population Biology* 38, pp. 219-232.

Fudenberg, D. and Kreps, D. (1993). "Learning mixed equilibria". *Games and Economic Behavior* 5(3), pp. 320-367.

Fudenberg, D. and Levine, D. K. (1998). *The theory of learning in games*. Cambridge, MA: MIT Press.

Galán, J. M. and Izquierdo, L. R. (2005). "Appearances can be deceiving: Lessons learned re-implementing Axelrod's 'evolutionary approach to norms'". *Journal of Artificial Societies and Social Simulation* 8(3), Article 2.

Gayer, G., Gilboa, I. and Lieberman, O. (2007). "Rule-Based and Case-Based Reasoning in Housing Prices". *The B.E. Journal of Theoretical Economics* 7(1), Article 10.

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

Gilboa, I., Lieberman, O. and Schmeidler, D. (2006). "Empirical Similarity". *The Review of Economics and Statistics* 88(3), pp. 433-444.

Gilboa, I. and Schmeidler, D. (1995). "Case-based decision theory". *Quarterly Journal of Economics* 110(3), pp. 605-639.

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

Gintis, H. (2000). *Game Theory Evolving: A Problem-Centered Introduction to Modeling Strategic Interaction.* Princeton, New Jersey: Princeton University Press.

Gotts, N. M., Polhill, J. G. and Adam, W. J. (2003a). "Simulation and analysis in agent-based modelling of land use change". *Online Proceedings of the First Conference of the European Social Simulation Association*, Groningen, The Netherlands, 18-21 September 2003.

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(1), pp. 3-92.

Hales, D. (2000). "Cooperation without Memory or Space: Tags, Groups and the Prisoner's Dilemma". In *Lecture Notes in Computer Science 1979/2000: Multi-Agent-Based Simulation: Second International Workshop, MABS 2000, Boston, MA, USA, July. Revised and Additional Papers*, Moss, S. and Davidsson, P. (eds.), pp. 157-166.

Hales, D. and Edmonds, B. (2005). "Applying a socially inspired technique (tags) to improve cooperation in P2P networks". *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans.* 35(3), pp. 385-395.

Hamilton, W. D. (1967). "Extraordinary sex ratios". *Science* 156(3774), pp. 477-488.

Hardin, G. (1968). "The tragedy of the commons. The population problem has no technical solution; it requires a fundamental extension in morality". *Science* 162(859), pp. 1243-1248.

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

Harsanyi, J. C. (1967a). "Games with Incomplete Information Played by "Bayesian" Players I-III. Part I: The basic model." *Management Science* 14(3), pp. 159-182.

Harsanyi, J. C. (1967b). "Games with Incomplete Information Played by "Bayesian" Players I-III. Part II: Bayesian equilibrium points." *Management Science* 14(5), pp. 320-334.

Harsanyi, J. C. (1968). "Games with Incomplete Information Played by "Bayesian" Players I-III. Part III: The basic probability distribution of the game." *Management Science* 14(7), pp. 486-502.

Hauert, C. and Doebeli, M. (2004). "Spatial structure often inhibits the evolution of cooperation in the snowdrift game". *Nature* 428(6983), pp. 643-646.

Hegselmann, R. and Flache, A. (2000). "Rational and Adaptive Playing". *Analyse & Kritik* 22(1), pp. 75-97.

Hofbauer, J., Schuster, P. and Sigmund, K. (1979). "A note on evolutionary stable strategies and game dynamics". *Journal of Theoretical Biology* 81(3), pp. 609-612.

Holt, C. A. and Roth, A. E. (2004). "The Nash equilibrium: A perspective". *Proceedings of the National Academy of Sciences of the United States of America* 101(12), pp. 3999-4002.

Holland, J. (1993). "The Effect of Labels (Tags) on Social Interactions". *Santa Fe Institute Working Paper*, 93-10-064. Santa Fe, NM

Hopkins, E. (2002). "Two competing models of how people learn in games". *Econometrica* 70(6), pp. 2141-2166.

Hopkins, E. and Posch, M. (2005). "Attainability of boundary points under reinforcement learning". *Games and Economic Behavior* 53(1), pp. 110-125.

Huet, S., Edwards, M. and Deffuant, G. (2007). "Taking into Account the Variations of Neighbourhood Sizes in the Mean-Field Approximation of the Threshold Model on a Random Network". *Journal of Artificial Societies and Social Simulation* 10(1), Article 10.

Ianni, A. (2001). "Reinforcement learning and the power law of practice: Some analytic results". *Mimeo. University of Southampton.*

Imhof, L. A., Fudenberg, D. and Nowak, M. A. (2005). "Evolutionary cycles of cooperation and defection". *Proceedings of the National Academy of Sciences of the United States of America* 102(31), pp. 10797-10800.

Intriligator, M. D., Bodkin, R. G. and Hsiao, C. (1996). *Econometric Models, Techniques and Applications*. 2nd edition. Prentice Hall.

Izquierdo, L. R., Gotts, N. M. and Polhill, J. G. (2003). "Case-Based Reasoning and Social Dilemmas: An Agent-Based Simulation". *Online Proceedings* *of the First Conference of the European Social Simulation Association*, Groningen, The Netherlands, 18-21 September 2003.

Izquierdo, L. R., Gotts, N. M. and Polhill, J. G. (2004). "Case-based reasoning, social dilemmas, and a new equilibrium concept". *Journal of Artificial Societies and Social Simulation* 7(3), Article 1.

Izquierdo, L. R. and Polhill, J. G. (2006). "Is your model susceptible to floating-point errors?" *Journal of Artificial Societies and Social Simulation* 9(4), Article 4.

Janssen, J. and Manca, R. (2006). *Applied Semi-Markov Processes*. New York, NY, USA: Springer.

Joyce, D., Kennison, J., Densmore, O., Guerin, S., Barr, S., Charles, E. and Thompson, N. S. (2006). "My Way or the Highway: a More Naturalistic Model of Altruism Tested in an Iterative Prisoners' Dilemma". *Journal of Artificial Societies and Social Simulation* 9(2), Article 4.

Kahneman, D., Slovic, P. and Tversky, A. (1982). *Judgment under Uncertainty: Heuristics and Biases*. Cambridge University Press.

Kalai, E. and Lehrer, E. (1993a). "Rational Learning Leads to Nash Equilibrium". *Econometrica* 61(5), pp. 1019-1045.

Kalai, E. and Lehrer, E. (1993b). "Subjective Equilibrium in Repeated Games". *Econometrica* 61(5), pp. 1231-1240.

Karandikar, R., Mookherjee, D., Ray, D. and Vega-Redondo, F. (1998). "Evolving Aspirations and Cooperation". *Journal of Economic Theory* 80(2), pp. 292-331.

Kim, Y. (1999). "Satisficing and optimality in 2x2 common interest games". *Economic Theory* 13(2), pp. 365-375.

Kirchkamp, O. (1999). "Simultaneous evolution of learning rules and strategies". *Journal of Economic Behavior & Organization* 40(3), pp. 295-312.

Kirchkamp, O. (2000). "Evolution of learning rules in space". In *Tools and Techniques for Social Science Simulation*, Suleiman, R., Troitzsch, K. G. and Gilbert, G. N. (eds.), pp. 179-195. Berlin: Physica-Verlag.

Kleijnen, J. P. C. (1995). "Verification and validation of simulation models". *European Journal of Operational Research* 82(1), pp. 145-162.

Klein, G. A. and Calderwood, R. (1988). "How do people use analogues to make decisions?" *Proceedings of the DARPA Workshop on Case-based Reasoning*, Calif., USA. Morgan Kaufmann.

Kleindorfer, G. B., O'Neill, L. and Ganeshan, R. (1998). "Validation in simulation: Various positions in the philosophy of science". *Management Science* 44(8), pp. 1087-1099.

Kolodner, J. L. (1993). *Case-Based Reasoning*. San Mateo, USA: Morgan Kaufman Publishers.

Kreps, D., Milgrom, P., Roberts, J. and Wilson, R. (1982). "Rational cooperation in the finitely repeated prisoner's dilemma". *Journal of Economic Theory* 27(2), pp. 245-252.

Kuhn, S. (2001). "Prisoner's dilemma". In *The Stanford Encyclopedia of Philosophy*, Zalta, E. N. (ed.). Winter 2001 Edition.

Kulkarni, V. G. (1995). *Modeling and Analysis of Stochastic Systems*. Chapman & Hall/CRC.

Kushner, H. J. and Yin, G. G. (1997). *Stochastic Approximation Algorithms and Applications*. New York: Springer-Verlag.

Laslier, J. F., Topol, R. and Walliser, B. (2001). "A Behavioral Learning Process in Games". *Games and Economic Behavior* 37(2), pp. 340-366.

Laslier, J. F. and Walliser, B. (2005). "A reinforcement learning process in extensive form games". *International Journal of Game Theory* 33(2), pp. 219-227.

Ledyard, J. O. (1995). "Public goods: A survey of experimental research". In *The Handbook of Experimental Economics*, Kagel, J. H. and Roth, A. E. (eds.), pp. 111-194. Princeton University Press.

Lewontin, R. C. (1961). "Evolution and the Theory of Games". *Journal of Theoretical Biology* 1, pp. 382-403.

Linster, B. G. (1992). "Evolutionary stability in the infinitely repeated prisoner's dilemma played by two-state Moore machines". *Southern Economic Journal* 58(4), pp. 880-903.

Ljung, L. (1999). *System Identification. Theory for the User*. 2nd edition. Englewood Cliffs, NJ: Prentice-Hall.

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

Luce, R. D. and Raiffa, H. (1957). *Games and Decisions: Introduction and Critical Survey*. New York Wiley.

Macy, M. and Flache, A. (2007). "Reply: Collective Action and the Empirical Content of Stochastic Learning Models". *The American Journal of Sociology* 112(5), pp. 1546-1554.

Macy, M. W. (1995). "PAVLOV and the Evolution of Cooperation. An Experimental Test". *Social Psychology Quarterly* 58(2), pp. 74-87.

Macy, M. W. (1998). "Social Order in Artificial Worlds". *Journal of Artificial Societies and Social Simulation* 1(1), Article 4.

Macy, M. W. and Flache, A. (2002). "Learning dynamics in social dilemmas". *Proceedings of the National Academy of Sciences of the United States of America* 99, pp. 7229-7236.

Maier, N. R. F. and Schneirla, T. C. (1964). *Principles of Animal Psychology*. New York: Dover Publications.

Maynard Smith, J. and Price, G. R. (1973). "The logic of animal conflict". *Nature* 246(5427), pp. 15-18.

McAllister, P. (1991). "Adaptive approaches to stochastic programming". *Annals of Operations Research* 30(1), pp. 45-62.

Mohler, R. R. (1991). *Nonlinear Systems, Volume I: Dynamics and Control*. Englewood Cliffs: Prentice Hall.

Mookherjee, D. and Sopher, B. (1994). "Learning Behavior in an Experimental Matching Pennies Game". *Games and Economic Behavior* 7(1), pp. 62-91.

Mookherjee, D. and Sopher, B. (1997). "Learning and Decision Costs in Experimental Constant Sum Games". *Games and Economic Behavior* 19(1), pp. 97-132.

Nash, J. F. (1951). "Non-cooperative games". *Annals of Mathematics* 54(2), pp. 286-295.

Németh, A. and Takács, K. (2007). "The Evolution of Altruism in Spatially Structured Populations". *Journal of Artificial Societies and Social Simulation* 10(3), Article 4.

Nicolov, N. (1997). "Case-based Reasoning". In *Artificial Intelligence Techniques: A Comprehensive Catalogue*, Bundy, A. (ed.), pp. 13-14. Springer-Verlag.

Norman, M. F. (1968). "Some convergence theorems for stochastic learning models with distance diminishing operators". *Journal of Mathematical Psychology* 5(1), pp. 61-101.

Norman, M. F. (1972). *Markov Processes and Learning Models*. New York: Academic Press.

Nowak, M. (1990). "Stochastic strategies in the prisoner's dilemma". *Theor. Pop. Biol.* 38, pp. 93-112.

Nowak, M. and Sigmund, K. (1990). "The evolution of stochastic strategies in the Prisoner's Dilemma". *Acta Applicandae Mathematicae* 20(3), pp. 247-265.

Nowak, M. A. and May, R. M. (1992). "Evolutionary games and spatial chaos". *Nature* 359(6398), pp. 826-829.

Nowak, M. A. and May, R. M. (1993). "The spatial dilemmas of evolution". *International Journal of Bifurcation and Chaos* 3(1), pp. 35-78.

Nowak, M. A., Sasaki, A., Taylor, C. and Fudenherg, D. (2004). "Emergence of cooperation and evolutionary stability in finite populations". *Nature* 428(6983), pp. 646-650.

Nowak, M. A. and Sigmund, K. (1992). "Tit for tat in heterogeneous populations". *Nature* 355(6357), pp. 250-253.

Nowak, M. A. and Sigmund, K. (1995). "Invasion dynamics of the finitely repeated prisoner's dilemma". *Games and Economic Behavior* 11, pp. 364-390.

Nowak, M. A. and Sigmund, K. (2004). "Evolutionary Dynamics of Biological Games". 303(5659), pp. 793-799.

Palomino, F. and Vega-Redondo, F. (1999). "Convergence of aspirations and (partial) cooperation in the prisoner's dilemma". *International Journal of Game Theory* 28(4), pp. 465-488.

Pazgal, A. (1997). "Satisficing leads to cooperation in mutual interests games". *International Journal of Game Theory* 26(4), pp. 439-453.

Pearce, D. G. (1984). "Rationalizable Strategic Behavior and the Problem of Perfection". *Econometrica* 52(4), pp. 1029-1050.

Phansalkar, V. V., Sastry, P. S. and Thathachar, M. A. L. (1994). "Absolutely expedient algorithms for learning Nash equilibria". *Proceedings of the Indian Academy of Sciences Mathematical Sciences* 104(1), pp. 279-294.

Polhill, G. and Izquierdo, L. (2005). "Lessons learned from converting the artificial stock market to interval Arithmetic". *Journal of Artificial Societies and Social Simulation* 8(2), Article 2.

Polhill, J. G. and Edmonds, B. (2007). "Open Access for Social Simulation". *Journal of Artificial Societies and Social Simulation* 10(3), Article 10.

Polhill, J. G., Izquierdo, L. R. and Gotts, N. M. (2005). "The ghost in the model (and other effects of floating point arithmetic)". *Journal of Artificial Societies and Social Simulation* 8(1).

Polhill, J. G., Izquierdo, L. R. and Gotts, N. M. (2006). "What every agent-based modeller should know about floating point arithmetic". *Environmental Modelling and Software* 21(3), pp. 283-309.

Posch, M. (1997). "Cycling in a stochastic learning algorithm for normal form games". *Journal of Evolutionary Economics* 7(2), pp. 193-207.

Probst, D. (1996). *On evolution and learning in games*. PhD thesis. University of Bonn.

Probst, D. (1999). "Review of "Evolutionary Game Theory", by Jörgen Weibull." *Journal of Artificial Societies and Social Simulation* 2(1).

Raub, W. (1988). "An Analysis of the Finitely Repeated Prisoners' Dilemma". *European Journal of Political Economy* 4(3), pp. 367-380.

Raub, W. and Voss, T. (1986). "Conditions for Cooperation in Problematic Social Situations". In *Paradoxical Effects of Social Behavior: Essays in Honor of Anatol Rapoport*, Diekmann, A. and Mitter, P. (eds.), pp. 85-103. Heidelberg and Vienna: Physica-Verlag.

Reisberg, D. (1999). "Learning". In *The MIT Encyclopedia of the Cognitive Sciences*, Wilson, R. A. and Keil, F. C. (eds.), pp. 460-461.

Riolo, R. L. (1997). "The Effects and Evolution of Tag-Mediated Selection of Partners in Populations Playing the Iterated Prisoner's Dilemma". *Proceedings of the Seventh International Conference on Genetic Algorithms (ICGA97)*. Morgan Kaufmann: San Francisco.

Riolo, R. L., Cohen, M. D. and Axelrod, R. (2001). "Evolution of cooperation without reciprocity". *Nature* 414(6862), pp. 441-443.

Rissanen, J. (1978). "Modeling by shortest data description". *Automatica* 14, pp. 465-471.

Roberts, G. and Sherratt, T. N. (2002). "Behavioural evolution (Communication arising): Does similarity breed cooperation?" *Nature* 418(6897), pp. 499-500.

Ross, B. H. (1989). "Some psychological results on case-based reasoning". *Case-Based Reasoning Workshop, DARPA*. Morgan Kaufmann, Inc.

Ross, D. (2006). "Game Theory". In *The Stanford Encyclopedia of Philosophy*, Zalta, E. N. (ed.). Spring 2006 Edition.

Roth, A. E. (1995). "Introduction to Experimental Economics". In *Handbook of Experimental Economics*, Kagel, J. H. and Roth, A. E. (eds.), pp. 3-109. Princeton University Press.

Roth, A. E. and Erev, I. (1995). "Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term". *Games and Economic Behavior* 8(1), pp. 164-212.

Rustichini, A. (1999). "Optimal Properties of Stimulus - Response Learning Models". *Games and Economic Behavior* 29(1-2), pp. 244-273.

Santos, F. C., Pacheco, J. M. and Lenaerts, T. (2006). "Evolutionary dynamics of social dilemmas in structured heterogeneous populations". *Proceedings of the National Academy of Sciences of the United States of America* 103(9), pp. 3490-3494.

Sastry, P. S., Phansalkar, V. V. and Thathachar, M. A. L. (1994). "Decentralized learning of Nash equilibria in multi-person stochastic games with incomplete information". *IEEE Transactions on Systems, Man and Cybernetics* 24(5), pp. 769-777.

Schank, R. C. (1982). *Dynamic Memory: A Theory of Reminding and Learning in Computers and People*. Cambridge, UK: Cambridge University Press

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

Selten, R. (1975). "Reexamination of the perfectness concept for equilibrium points in extensive games". *International Journal of Game Theory* 4(1), pp. 25-55.

Sethi, R. and Somanathan, E. (1996). "The Evolution of Social Norms in Common Property Resource Use". *The American Economic Review* 86(4), pp. 766-788.

Simon, H. A. (1957). *Models of Man: Social and Rational; Mathematical Essays on Rational Human Behavior in a Social Setting*. New York: John Wiley and Sons.

Simon, H. A. (1982). *Models of Bounded Rationality*. Cambridge, MA: The MIT Press.

Söderström, T. and Stoica, R. (1989). *System Identification*. Hemel Hempstead, UK: Prentice Hall International.

Taylor, C., Fudenberg, D., Sasaki, A. and Nowak, M. A. (2004). "Evolutionary game dynamics in finite populations". *Bulletin of Mathematical Biology* 66(6), pp. 1621-1644.

Taylor, P. D. and Jonker, L. B. (1978). "Evolutionarily stable strategies and game dynamics". *Mathematical Biosciences* 40(1-2), pp. 145-156.

Thorndike, E. L. (1898). *Animal intelligence: An experimental study of the associative processes in animals*. New York MacMillan.

Thorndike, E. L. (1911). *Animal Intelligence: Experimental Studies*. New York: The Macmillan Company.

Traulsen, A., Nowak, M. A. and Pacheco, J. M. (2006). "Stochastic dynamics of invasion and fixation". *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 74(1), Article 011909.

Van Damme, E. (1987). *Stability and Perfection of Nash Equilibria*. 2nd edition. Berlin Springer Verlag.

Vanderschraaf, P. and Sillari, G. (2007). "Common Knowledge". In *The Stanford Encyclopedia of Philosophy* Zalta, E. N. (ed.). Fall 2007 Edition.

Vega-Redondo, F. (2003). *Economics and the Theory of Games*. Cambridge University Press.

Von Neumann, J. and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton University Press.

Watson, I. (1997). *Applying case-based reasoning: Techniques for enterprise systems*. Morgan Kaufman Publishers.

Weibull, J. W. (1995). *Evolutionary Game Theory*. Cambridge, MA: MIT Press.

Weibull, J. W. (1998). "Evolution, rationality and equilibrium in games". *European Economic Review* 42(3-5), pp. 641-649.

Weibull, J. W. (2002). "What have we learned from evolutionary game theory so far?" *Stockholm School of Economics and the Research Institute of Industrial Economics. Working Paper*.

Wilensky, U. (1999). *NetLogo*. Evanston, IL Center for Connected Learning and Computer-Based Modeling, Northwestern University.

Wustmann, G., Rein, K., Wolf, R. and Heisenberg, M. (1996). "A new paradigm for operant conditioning of Drosophila melanogaster". *Journal of Comparative Physiology - A Sensory, Neural, and Behavioral Physiology* 179(3), pp. 429-436.

Young, H. P. (1993). "The Evolution of Conventions". *Econometrica* 61(1), pp. 57-84.