Evolutionary Models of the Ultimatum Game

by

Melt van Schoor

Dissertation presented for the degree of Doctor of Philosophy in Economics in the Faculty of Economic and Management Sciences at Stellenbosch University

Supervisor: Prof. Rulof Burger

March 2025

Declaration

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: 2025/01/15

Declaration by the candidate

This dissertation includes one original paper published in the Economic Research Southern Africa (ERSA) working paper series (chapter 2) and two unpublished chapters (chapters 3 and 4). I am the sole author (100% contribution) of the working paper and also the sole author (100% contribution) of the two unpublished chapters.

Signature: ______________________________

Date: January 2025

[Declaration with signature in possession of candidate and supervisor.]

Abstract

Evolutionary Models of the Ultimatum Game

M. van Schoor

Department of Economics,
Stellenbosch University,
Private Bag X1, Matieland 7602, South Africa.

Dissertation: PhD (Economics)

March 2025

This dissertation uses evolutionary models to improve our understanding of the structure of the ultimatum game and to provide and critically interpret explanations from evolutionary models for empirical results.

Existing evolutionary models have established that stable evolutionary equilibria exist in which proposers make offers of relatively equal bargaining outcomes, while responders reject low but positive offers. These results are at odds with conventional game theoretic predictions, but consistent in some respects with observed human behaviour in experimental studies. The evolutionary dynamics leading to this result is not well understood for the full ultimatum game, as comprehensive analysis of dynamics have only been provided for a simplified minigame version of the game. Moreover, it has not been established that a minigame can constitute an adequate approximation for the full game. To address these deficiencies, conditional frequency dynamic analysis is used to formally link minigames to the full game, allowing the former to explain key results in the latter, particularly the existence and stability of non-subgame-perfect equilibria.

The relevance of these results for experimental data is addressed next. Evolutionary models can be interpreted as frameworks for examining the cultural evolution of behavioural norms and preferences. This requires constructing microfoundations for the aggregate dynamics model that describes the stochastic process of individual-level strategy revision. It is shown that certain assumptions made in the aggregate-level model, which are necessary to predict relatively equal divisions, lead to unreasonable discrepancies in learning rates between individual agents in the proposer and responder populations. Furthermore, the results are not robust to stochastic disturbances in finite populations, limiting their applicability to empirical data and suggesting that additional model features are required to account for experimental findings of relatively equal divisions and rejections of low offers.

The final research question investigated is whether allowing responders to build reputations can create the necessary incentives for higher offers and higher acceptance thresholds. However, a reputation for rejecting low offers can only be established if there is a sufficient frequency of low offers to reject, thus information must be treated as endogenous. A general endogenous-information framework is developed to calculate endogenous information equilibria in two-player population games, where the available information is consistent with the pattern of action profiles induced by it. The framework is used to explore different types of reputations, including negative reputations that harm responders when observed and positive reputations that benefit them when observed. It is shown that the different reputation mechanisms are complementary and can lead to relatively equal divisions and rejections of low offers, consistent with observed behaviour in experimental studies. These reputation-based models also offer a plausible explanation for the evolution of relatively equal divisions in societies where such norms were initially absent – something that the baseline and earlier reputation-based models fail to explain.

Acknowledgments

I thank my supervisor, Prof. Rulof Burger for keeping me on course and for dedicated and thoughtful guidance, advice and support. I also acknowledge the critical support, time and resources afforded by the Department of Economics at Stellenbosch University, and the roles played by Departmental Heads, Profs. Andrie Schoombee, Ada Jansen and Sophia du Plessis, and other colleagues in this regard. I thank my three examiners for their excellent feedback and advice, not only for improving this dissertation, but also for the journey onwards.

This work is dedicated to the memory of Prof. Philip Black, for friendship, encouragement and many stimulating discussions.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

During the 1970s and early 1980s, there were substantial interest and steady progress in the strategic modelling of bargaining, with some of the most important theoretical advances, including the alternating-offers models of Ståhl (1972) and Rubinstein (1982), originating in this period. At the same time, there was (and still remains today) a debate about the adequacy of standard methodological assumptions used in game theory to explain and predict real human behaviour, which lead to calls for empirical testing (Güth and Kocher, 2014, p. 397). The complexities of general bargaining theory, combined with the challenge of developing sound experimental methods, which were new in economics at the time, eventually led Werner Güth and his students (1982) to focus on one of the simplest possible bargaining interactions, which Harsanyi (1961) had earlier called the ultimatum game (UG).

Given a fixed amount of money to divide, one player makes a take-it-or-leave-it offer and a second player gives a simple accept/reject response. The payoffs are the first player's offered division, if the second player agreed, or zero for both players, if the second player refused. Apart from being simple, the UG is useful in applications as it captures the strategic essence of situations where one player can make a firm commitment to a specific offer (Harsanyi, 1961) or where institutions enable one party in a bargaining interaction, such as an employer (Mago et al., 2024, p. 529) or a landlord (Young and Burke, 2001, p. 566), to propose terms of an agreement unilaterally.

Figure 1.1 represents a simplified minigame version that illustrates the essential strategic features of the game. In the minigame, only two possible proposals are allowed, $(50, 50)$ ("Fair") and $(99, 1)$ ("Unfair"). However, regardless of which offer is made, the responder's best response is always to accept, since rejecting results in the lowest payoff: nothing. Applying backwards induction (indicated by thick lines) results in the proposer choosing the unfair offer. The game therefore has a unique subgame-perfect Nash equilibrium (SPNE), which allocates all the surplus to the proposer. Similarly in the full game, in which the proposer has many different possible offers, only the one allocating the max-

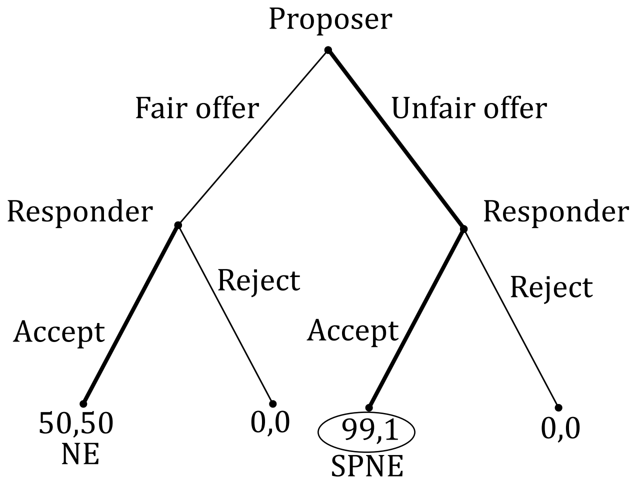


Figure 1.1: The ultimatum minigame (extensive form)

imum surplus to himself while still leaving a positive payoff to the responder leads to a SPNE. Note, however, that if the responder commits to a strategy of rejecting unfair offers, the proposer's best response is to make a fair offer, resulting in a second, imperfect, Nash equilibrium (NE) indicated in figure 1.1. In the full game, a commitment by the responder to reject offers lower than any specified amount makes an offer of such an amount a best response for the proposer, hence any division of the surplus is a NE outcome. These imperfect equilibria are conventionally regarded as non-credible, because if a low offer were actually made, the responder would have a clear incentive to accept rather than reject.

In experimental work, the theoretical game's payoffs are made concrete money amounts that are paid out to experimental subjects upon conclusion of the experiment. If the players' preferences are completely given by the current game's monetary payoffs, the only rational behaviour is given by the game's unique SPNE which predicts that proposers would make the lowest possible positive offers and responders would accept any positive amounts offered. But Güth et al. (1982)'s initial experimental study found results that were at odds with these "extreme" predictions. Numerous subsequent experimental studies have confirmed their results: human proposers typically offer between 30 and 50 per cent of the total money amount and responders often reject offers below 20 per cent (Camerer and Thaler, 1995, p. 210, Oosterbeek et al., 2004, Güth and Kocher, 2014, p. 398).

The stark discrepancy between theory and empirical results prompted numerous compelling research questions, generating considerable interest among scholars. Consequently, a steady stream of research appeared that sought to do further experimental tests under varying conditions, on the one hand, and to provide explanations for the results, on the other. More than four decades later, academic interest in the ultimatum game shows no signs of slowing down.[1]

From the outset, the need for empirical testing of bargaining theory sprung from doubts about two key assumptions made in conventional theory, namely whether people are perfectly rational utility calculators and whether people have narrowly selfish preferences (Güth and Kocher, 2014, p. 397). Over the years, interest in the latter question dominated the former, and the ultimatum game became a standard tool used by researchers interested in broad-ranging questions of cooperation, fairness, social norms, altruism, egalitarianism, the role of emotions in decision-making and brain research. The dominant explanation for experimental results in the standard ultimatum game is that responders experience positive utility from rejecting what they regard as inferior or unfair offers (Fehr and Schmidt, 2006, p. 630), and proposers are motivated by the need to avoid rejection and possibly also by other-regarding preferences including altruism, egalitarianism or social norms (Cooper and Kagel, 2016). In these endeavours, experimental researchers have largely taken rational behaviour for granted,[2] particularly in a game as simple as the UG (e.g. Camerer and Thaler, 1995, p. 210, Fehr and Schmidt, 2006, p. 617,628).

Nevertheless, the question of rationality has not been forgotten. Learning and evolution remain the most compelling justifications for the assumption of rational behaviour in economics and game theory (Selten, 1990; Mailath, 1998). Real people do not behave like sophisticated utility calculators when facing unfamiliar decision problems – instead, they combine limited reasoning with heuristics, rules of thumb, mimicry and social norms. Evolutionary models, in which a population of agents using different strategies evolve in such a way that more successful strategies' frequencies increase relative to less successful ones, are often used as a convenient and general proxy for bounded rationality and learning (Mailath, 1998, p. 1355). Thus, evolutionary models may be able to explain what happens when players play the game repeatedly and adapt their behaviour over time in response to their experience of good and bad outcomes. In addition, evolutionary models can provide explanations for how the norms and preferences exhibited by laboratory subjects may have originated. Conventions that arose in specific contexts, e.g. sharecropping agreements, acquire salience and become difficult to dislodge, and function to coordinate economic relations and reduce transaction costs and conflicts (Young and Burke, 2001; Burke and Young, 2011). More generally, these evolutionary processes are thought to have occurred in the past within a broader socioeconomic context, where people en-

[1] Güth and Kocher (2014, p. 396) mention that a Google Scholar search for "ultimatum bargaining" delivered more than 26 500 results in 2013. Today, the same search gives "about 51 100" results, while "ultimatum game" gives "about 104 000". Limiting results (on the latter) to publications since 2020 gives "about 16 800" results. Most of the results are experimental studies.

[2] Some have attempted to test for it empirically, e.g. Andreoni and Miller (2002), with positive results.

gaged in a variety of daily interactions (Gale et al., 1995, p. 70, Mailath, 1998, p. 1350, Güth and Napel, 2006, p. 1038, Skyrms, 2014). In these models, fitness is typically equated to material payoffs, so the models explain the evolution of behaviour without any specific assumptions of social preferences. If behaviour emerges that is compatible with social preferences, however, the outcome of the evolutionary model may be interpreted as reflecting such preferences.[3]

The research in this dissertation is aimed at using evolutionary models to advance understanding of the structure of the ultimatum game and to provide and critically interpret explanations from evolutionary models for empirical results. Evolution is used in the two senses mentioned above, firstly as a metaphor for interactive learning as players gain experience and secondly as an ultimate explanation for the possible origins of ingrained behavioural norms and preferences.

In this introductory chapter, I will provide some background, including a brief review of literature that uses evolutionary models to explain behaviour in the ultimatum game, with a specific focus on the model by Gale, Binmore, and Samuelson (1995) (henceforth GBS) that is used as the basis for the research in chapters 2 and 3 of this dissertation. This will be followed by summaries of the specific issues investigated in the respective chapters and their main findings, highlighting contributions and advances made. I conclude with a brief discussion of methodology.

1.1 Evolutionary models of the ultimatum game

Applications of evolutionary modelling to the UG roughly mirrors the experience of evolutionary game theory in economics, with a period of intense initial activity in the the 1990s, after which specific interest subsided, though the techniques had by then entered mainstream thinking and new studies continued to appear with some regularity. Many recent contributions have an interdisciplinary flavour. Two surveys on evolutionary modelling of the UG have been published (Debove et al., 2016; Akdeniz and Van Veelen, 2023), so the following literature review will be brief and focused on specific models of interest.

Many of the published articles seek to "explain fairness", i.e. relatively equal outcomes, in the ultimatum game, using some kind of treatment or device included in an evolutionary model that causes equal outcomes to result. In contrast, some of the early models were standard evolutionary models in which agents play the UG without any embellishments. I refer to these as baseline models as they are a useful benchmark for comparison to other models and they are also a good starting point for exploring other factors. They are also

[3]Similarly, rationality is not assumed in evolutionary models, but there are strong links between evolutionary equilibria and strategic equilibria, so outcomes may be assessed as reflecting rational behaviour.

the logical starting place when seeking to explain baseline empirical results in standard UG experiments, where no treatment is applied. The fact that there is no "treatment" effect in these evolutionary models does not mean they do not have their own remarkable results. In particular, they do not necessarily lead to the UG's SPNE.

1.1.1 Baseline models

The first baseline model, the GBS model, will be a primary focus of this dissertation, thus deserving of a more thorough review here. The authors apply standard replicator dynamics to two populations, proposers and responders respectively. A low rate of mutation is added, reflecting realistic mistakes and/or innovations. In Binmore and Samuelson (1994) a related model with a different version of the replicator dynamics, the so-called discrete or "adjusted" replicator dynamics, gives similar results. Differences between the two versions and how they may be interpreted are discussed in chapter 3. Mutation ensures that there is always a small positive frequency of each possible strategy. One would expect that the presence of at least a small frequency of low offers would create an incentive for responders to learn to accept all offers, so it is surprising to find that the model's final rest points in many cases do not reflect the SPNE where responders have learned to accept all positive offers.

Instead, proposers offer a substantial share of the total amount, such as 7 or 9 out of 40, and a substantial share of responders reject lower amounts. This result, i.e. evolutionary stable but imperfect equilibria, hinges on parameter values, and tends to appear when responders have *relatively* high mutation rates and proposers *relatively* high selection rates. Proposers then quickly learn to make acceptable offers, i.e. not too low, given some distribution of responder strategies. Consequently, responders have a very limited incentive to learn to accept lower positive offers, so strategies that reject lower offers can persist with the help of mutation. Even an arbitrarily small rate of responder mutation can achieve this result (depending on the other parameters), so it should not be thought that mutation forcefully "pushes" responders to reject low offers. Instead, the result is better explained by the very small size of the payoff advantage of the optimal SPNE strategy (accepting all positive offers) over other responder strategies.

In a series of models based on a single population with different types of mutation and strategy spaces, Harms (1997) confirms that the result that weakly dominated strategies and more equal bargains survive holds in many cases. A well-known paper by Roth and Erev (1995) uses a reinforcement learning model to derive broadly similar results, which are also shown to be relevant to experimental results that indicate learning taking place in a similar fashion as in the model. The similarity in these results to GBS supports GBS's (p. 83) assessment that similar effects should occur under a wide range of evolutionary dynamics specifications. Explaining, interpreting and critically assessing the relevance of this result, which I call the "GBS result", form a large part of this dissertation's objectives, so the specific aspects I investigate are discussed in more detail in the chapter summaries below.

Baseline models may provide partial explanations for interactive learning taking place as players gain experience over multiple rounds of play (with different partners to avoid repeated-game effects). But it is clear that there are many things that these models cannot explain on their own, so they are not a substitute for other-regarding preference models when conducting empirical work (Fehr and Schmidt, 2006, p. 628). However, evolutionary models may also be able to offer explanations for social norms, preferences and behaviour that experimental subjects bring to the laboratory.

Such explanations rely on the notion that the interactions that people have engaged in in their everyday lives do not resemble the once-off interactions in an experimental laboratory (Baumard and Sperber, 2010). Everyday interactions may allow for repeated interactions and relationships, reputations, different outside options, different bargaining rules, considerations of social status and many other factors that make them different from once-off ultimatum games. If preferences and behavioural norms were formed in complex real-world socioeconomic environments, there is an impetus to study evolution in such settings. This is a daunting task, but it is sensible to focus attention on interactions that are at least superficially similar to the one-shot UG, so that a plausible case can be made that a norm developed while playing the former can be applicable to the latter.

1.1.2 Information, reputation and commitment

It is plausible that humans have developed a strong and ingrained concern for protecting their reputations, and that the behaviour of a responder who rejects a low offer could be interpreted as serving this purpose. A person with a reputation for accepting low offers will tend to receive more low offers. There is strong experimental evidence that when the possibility of developing a reputation is made explicit, responders raise their acceptance thresholds (Fehr and Fischbacher, 2003; Poulsen and Tan, 2007), as this directly leads to getting higher offers from proposers when the proposers are informed about their past behaviour. It may be hard to completely switch off the inclination to reject low offers if it has become deeply ingrained, even in experiments where the framing is of anonymous one-shot interactions.

Nowak et al. (2000) present two evolutionary models in which responders are able to build reputations for toughness. Proposer lower their offers to a responder if they have information that the responder has accepted low offers in the past. This can be called a negative reputation model, as the reputation has a negative effect on the agent to which it is attached. Responders, now having a clear incentive to avoid negative reputations, are incentivized to rather reject low offers, which in turn gives proposers an incentive to make high offers rather than unconditional low offers. This simple mechanism provides a possible explanation for observed behaviour when experimental subjects engage in ultimatum bargaining interactions.

The notion of reputation is linked to the notion of commitment, which Akdeniz and Van Veelen (2023) argue is a more general concept that might explain fair outcomes in the UG. The key to either approach is that the proposer must have a way of obtaining information about the likely reaction of the responder to possible offers. Several other papers featuring evolutionary models of the UG with reputation mechanisms have been published (Poulsen, 2007; Debove et al., 2016; Zhang et al., 2023; Akdeniz and Van Veelen, 2023), but a number of important questions remain unexplored. Chapter 4, which is summarized below, explores a general approach to reputations that considers not only the effects of various kinds of reputations in the UG but also the feasibility of building them.

1.2 Rationale

The GBS model and result forms the basis for a large part of the research in this dissertation. There are several reasons why the GBS model is particularly interesting and important.[4] Firstly, the paper was an important advance in evolutionary game theory, where it has had a considerable impact, particularly in understanding the role of noise (mutation) and the stability of imperfect equilibria. The result is counterintuitive because intuition suggests that imperfect rest points in a noisy environment should be unstable, yet GBS show that they can be asymptotically stable. Secondly, the GBS model suggests that we should not expect to see SPNE result in experimental data, which is indeed the case. Even though the UG model is not able to explain every aspect of experimental data correctly, this observation remains important. Thirdly, experimental research specifically focused on learning behaviour have found trends that agree with what evolutionary models predict (Slonim and Roth, 1998, List and Cherry, 2000, Cooper and Dutcher, 2011), notably that proposers learn much faster than responders. Finally, the GBS and model and relatives may serve as an explanation for cultural evolution, or evolutionary processes outside the laboratory that could provide an ultimate explanation for norms and preferences which are relevant to experimental data.

[4]A more detailed discussion of the relevance of this paper can be found in section 3.3.

1.2.1 The dynamics for the full UG have only been explained using minigames

Despite its importance, it remains unclear from many of the articles citing GBS, as well as from GBS themselves, whether the evolutionary dynamics of the UG, particularly when using noisy replicator dynamics, are fully understood. The lack of clarity stems from the fact that the only detailed evolutionary analysis provided is for a minigame version of the UG, where the proposer has only two possible offer amounts. The main result comes from computer simulations of the full game. A satisfactory analysis of the dynamics of the full game has never been provided, nor has it been clearly established that a minigame can constitute an adequate approximation for the full game. As I will show in chapter 2, this is not a trivial problem, since there are dynamic complexities in the full UG model's evolutionary dynamic system that are absent from the minigame version, and a naive approach gives results that are incompatible with the minigame analysis.

1.2.2 The need for suitable microfoundations

I have noted above that evolutionary models can be used both as a proxy for boundedly rational interactive learning and to provide explanations for traits that have evolved in a different context, particularly cultural evolution of social norms. GBS (p. 70) argue that the social norms triggered in the short term by laboratory experiments can be presumed to have evolved in "real-life bargaining situations that are superficially similar to the ultimatum game in some respects" and that "we must therefore examine long-run behavior in these external situations for the origin of the norms that guide short-run behavior in laboratory experiments on the ultimatum game". This raises the possibility of reinterpreting the GBS model as a model of cultural evolution that can explain initial behaviour, norms and preferences.

However, this reinterpretation requires more than simply a assertion, as it needs to be clearly established that the model can be suitable for such an interpretation. The replicator dynamics only describes aggregate dynamics of strategy frequencies. In chapter 3, I argue that a cogent and defensible interpretation of such a model requires a more detailed account of what occurs at the level of individual actors, i.e. how and why they hold and revise their strategies over time in response to their experiences.

To fulfil this requirement, there have been a number of efforts by evolutionary game theorists to build more explicit "microfoundations" models for aggregate-level evolutionary models (Sandholm, 2010). GBS (p. 85) justify their use of replicator dynamics by presenting an individual-level learning model in which individuals adopt a randomly chosen alternative strategy when their payoffs fall below their aspiration levels. These aspiration levels are drawn from a uniform random distribution. However, this learning model is hard to reconcile with the notion of cultural evolution, and I argue that a model based on imitation of successful peers would be more suitable. Even though the different microfoundations models aggregate to the same systemic dynamics, there are important implications – elaborated upon in the chapter 3 summary below – for the interpretation of the model that calls into question the reasonableness of the GBS result.

1.2.3 How can fairness norms evolve where they do not already exist?

The conclusions in chapter 3 lead to a negative assessment of the relevance of the GBS result as a cultural evolution explanation for norms and preferences that support relatively equal divisions in UG experimental data. Harms (1997) and Akdeniz and Van Veelen (2023), in different ways, support the view that the GBS result is sensitive to the way mutation and state space is set up. In addition, part of my argument (section 3.7) is that the GBS result may be too fragile to survive realistic stochastic shocks. Baseline models may therefore provide persuasive arguments for why relatively equal divisions in the UG might survive for an extended period of time, but the result is not strong enough to explain why this should necessarily be the expected endpoint of very long-run evolutionary processes. In addition, these models do not provide a reason why such behaviour should be there in the first place. I agree with Akdeniz and Van Veelen (2023, p. 591) that a model that can explain empirical UG results should include some factors that can, at least sometimes, make rejections have a positive effect on responders' payoffs.

I have argued above (section 1.1.2) that concern for reputation is a plausible reason why responders reject low offers in the UG. This hypothesis does not require subjects to wrongly think they are playing a completely different game (Akdeniz and Van Veelen, 2023, p. 571) or that they are playing a repeated game (Fehr and Schmidt, 2006, p. 629). It merely requires subjects to behave as if they think their actions may become known, even with only a small probability, to future proposers that they might interact with. There is empirical support for this: Hoffman et al. (1996) find that subjects offer less in dictator games with unusually strict anonymity treatments[5] compared to treatments with "ordinary" anonymity. In real-world social groups, anonymity may have been quite limited, and consequences of bad reputations severe, leading to an ingrained reluctance to take risks in this regard.

The literature in this area, evolutionary models of the UG with reputational

[5]Their strictest treatment use a double blind design, with substitute blank slips that subjects can put in envelopes to make the envelopes they hand in appear equally thick regardless of their choices.

mechanisms, is not well developed, despite a number of valuable contributions. Four shortcomings can be identified. Firstly, existing models tend to make arbitrary assumptions about the availability of information to proposers on responders, rather than linking such knowledge to histories of past interactions that could conceivably have taken place. Secondly, there is a limited notion of what knowledge of past interactions can be useful to proposers; in particular only negative information tends to be considered. Thirdly, there is too heavy reliance on computer simulations without explicit analytical results to help understand how and why outcomes arise from model assumptions. Finally, there are open questions on e.g. the effects of initial conditions and existence and stability of equilibria, as these are not always systematically investigated and explicitly described.

The most widely cited paper on this topic, Nowak et al. (2000), reports that their model[6] is bistable, meaning that a fair (relatively equal) outcome can be sustained, but a highly unequal outcome as in the UG's SPNE is *also* stable.[7] There is nothing wrong with multiple equilibria, but the problem here is similar to that of the GBS model – the model cannot explain how norms for an equal division can arise from a state where such norms are initially absent. As I will show in chapter 4, by simply adopting a more holistic view of reputations – allowing both negative and positive reputations – it becomes possible to give such an explanation.

1.2.4 General concerns

Debove et al. (2016, p. 249) express a number of valid general concerns regarding existing literature on evolutionary models of the ultimatum game. Some of the issues not already mentioned include loose usage of technical terms (sometimes, it seems, due to researchers from different academic disciplines not sharing a common language), unjustified constraints on behaviour (e.g. some models where agents can play both proposers and responder roles simply assume that proposers will be empathetic, i.e. not offer less than they demand as responders), and lack of clear reporting on initial conditions. The last concern in particular seems to be connected to a general lack of adequate analysis of dynamics and over-reliance on computer simulation results, particularly agent-based models, to "explain fairness" without sufficient theoretical elucidation of the results.

[6] I am referring here to their first model, a minigame model, which is the only one for which a relatively comprehensive (though also brief) analysis of dynamics is provided.

[7] They do argue that the latter may have a smaller basin of attraction.

1.3 Research objective and questions

The overall objective of this dissertation is to use evolutionary models to advance understanding of the structure of the ultimatum game and to provide and critically interpret explanations from evolutionary models for empirical results.

The aim is not simply to build yet more models that explain how fairness could have evolved, but also to critically interpret and evaluate existing models, with a particular focus on GBS.

More specific research questions include:

1. Can minigames provide an adequate account of the full UG's dynamics?

2. Can the GBS model and the GBS result (i.e. evolutionary stable but imperfect equilibria) provide a reasonable account of cultural evolution of behavioural norms?

3. Can responders' concern with their reputations provide an explanation of behavioural norms exhibited in UG empirical results?

1.4 Chapter summaries and contributions

This section provides brief summaries of the main research chapters of this dissertation. At the end of each summary, I highlight advances and contributions made.

1.4.1 Chapter 2: Using minigames to explain imperfect outcomes in the ultimatum game

In this chapter, the first specific research question listed above is addressed: Can minigames provide an adequate account of the full UG's dynamics?

As with any game where players choose an amount of money, the UG has many pure strategies, which necessitates a large number of variables in evolutionary models to track the frequency of each pure strategy. This, along with a large number of weak Nash equilibria, leads to interesting and complicated dynamics in which the system can linger for an extended time period close to a Nash equilibrium, only to be followed by a sudden escape and a comparatively rapid transition to another such state, a pattern that can repeat a number of times before the model comes to an eventual final rest. However, in the literature, it is common to perform detailed analysis of dynamics only on simplified minigames in which proposers are each limited to only two possible offer amounts. It has not been established that minigames can provide a full understanding of the full game's evolutionary dynamics, since the simplified models lack the multiplicity of equilibria and complicated dynamics of the full UG.

I demonstrate in this chapter that the naive minigame analysis fails in a critical respect to account for a well known and important result from computer simulations of the full game, which I have called the GBS result above, i.e. that evolutionary dynamics in the full game does not generally lead to the SPNE result. It is clear from GBS's two-dimensional analysis that the role of mutation (noise) is to keep the frequency of the responder strategy that rejects low offers at a sufficient level to make the high offer a best response for proposers, which is necessary to maintain an imperfect equilibrium. But in the full game, in which there are many strategies, mutation appears to work in the wrong direction, by dissipating frequency among a large number of strategies, thus lowering rather than raising the required responder frequency in order to stabilize an imperfect equilibrium.

A more rigorous analysis is developed that does account for the full game's dynamics in a satisfactory way, by identifying a suitable subsets of the full game's strategy space and deriving *conditional frequency dynamics* for the subsets. Each of the game's Nash equilibria then has an associated conditional game in which all *higher* offers and demands are excluded from the strategy sets. I show that these conditional games are structurally similar to a full UG game and the conditional dynamics applicable to them are of the same form as the full game's noisy replicator dynamics. Since the conditional dynamics are derived directly from the full game's dynamics, an equilibrium in one of the conditional games is also an equilibrium of the full game. In addition, a two-dimensional minigame is shown to provide a good approximation for the conditional games' dynamics, so conclusions derived from stability analysis on the minigame can be mapped back to the full game.

Specific advances in this chapter include the development of a technique by which the analysis of dynamics in a minigame can be made applicable to a more complicated, larger game. The analysis also provides greater clarity on why some imperfect equilibria in the full game can be stabilized more easily than others, which has previously been the subject of educated guesswork based on simulation results rather than rigorous analysis. The chapter also develops techniques that are likely to have wider applicability. The analysis of conditional frequency dynamics allows clear focus on the dynamics within a game with reduced strategy sets. Graphical selection-mutation equilibrium loci analysis can help to understand dynamic forces in a two-dimensional model, particularly when there are multiple forces (e.g. selection and mutation) balanced against each other at the equilibria.

1.4.2 Chapter 3: Stochastic learning and emulation in the ultimatum game

This chapter addresses the second specific research question, namely, can the GBS model and the GBS result (i.e. evolutionary stable but imperfect equilibria) provide a reasonable account of cultural evolution of behavioural norms?

The second direction of research based on the GBS model, in chapter 3, addresses its interpretation and relevance to explaining empirical results. The GBS model's replicator dynamics specifies aggregate dynamics of strategy frequencies, which does not make clear what is occurring at the individual level. In their paper, GBS sketches an individual-level learning model based on agents changing their strategies when payoffs fail to meet random aspiration levels. I argue that this model is not appropriate for exploring the historical context of cultural evolution, where entrenched behavioural norms and preferences may have evolved. In this paper, I develop and match the aggregate deterministic GBS model with a simple stochastic individual-level model, similar to a model in Weibull (1995, p. 158), which is based on imitation of social peers. Here, agents have a degree of sensitivity to payoffs of others when deciding who to imitate, unlike the random-aspiration-level model where this decision is made blindly. I argue that this model is suited to an interpretation of cultural evolution of social norms. The model can be specialized so that it aggregates either to the standard continuous replicator dynamics used in GBS or to the discrete or "adjusted" replicator dynamics used in Binmore and Samuelson (1994).

The model suggests two implications that reflect negatively on the relevance of the GBS result to empirical data. Firstly, the asymmetry GBS assume between proposer and responder selection and mutation rate parameters to obtain their result may have the unreasonable implication at individual level that responder mutation rates are doubly boosted when the standard continuous replicator dynamics are used, and secondly their aggregate dynamic equations may be relying unreasonably heavily on the law of large numbers to suppress individual-level stochastic effects. GBS's modelling correctly points out that learning by responders may be very slow compared to proposers, and overall adjustment towards the subgame-perfect equilibrium may be slow enough to be practically imperceptible, so one should not necessarily expect to observe subgame-perfect outcomes in empirical data. However, their main result, the asymptotic stability of non-subgame-perfect equilibria, may require unreasonably biased parameter values, and the stability may be too fragile to survive realistic real-world stochastic effects, so we should not expect to observe such a result either.

Specific advances in this chapter include the development and interpretation of an individual-level learning model that aggregates to either the standard continuous replicator dynamics or to the adjusted discrete replicator dynamics. This model could be applied to other games. In addition, a negative assessment is reached about the applicability of the GBS model's ability to explain cultural evolution of behavioural norms and/or preferences that might explain empirical results in UG experiments.

1.4.3 Chapter 4: An evolutionary perspective on good and bad reputations in the ultimatum game

This chapter addresses the third specific research question: Can responders' concern with their reputations provide an explanation of behavioural norms exhibited in UG empirical results?

This chapter proceeds from the notion that building up a credible reputation as a tough bargainer requires costly demonstrations that one is willing to reject low offers. There is, however, a paradox in this: if responders successfully deter low offers, the conditions needed to generate and communicate a track record as a tough bargainer is thereby undermined. If the availability of a certain kind of information results in a pattern of behaviour that in itself has an impact on the availability of the information, then information must be treated as endogenous for the theory to be complete.

In this chapter, I therefore develop a general endogenous information model framework for two-player sequential-move population games in which the first player's strategies are mappings from sets of information, which I call *signal sets*, about the second player's past behaviour. A signal set consists of zero or more *signals*, where a signal corresponds to an action profile. For example, in the ultimatum game, a signal could be $(4, 1)$, indicating that the responder was observed accepting an offered amount of 4, and a signal set could be $\{(4, 1), (2, 0)\}$, indicating that the responder was observed, in separate interactions, to accept 4 and reject 2. The first player (proposer) is assumed to have a certain probability to observe each of the action profiles of the second player that occurred in a sample of some fixed size of that player's interactions. The framework is flexible and general enough to accommodate any number of strategies by either player, and any number of signals that can follow from interactions.

The framework defines an endogenous information equilibrium, in which the system's information state, which is a probability distribution over possible signal sets per responder type, induces a distribution of action profiles per responder type, which in turn generates the same information state. Strategies are fixed during the rounds of repeated play required for the information state to stabilize, as this is assumed to be a short-run process. Reputations are therefore generated instantaneously at every point in evolutionary time. The framework calculates, for the general case, expected utilities for each player given strategy distributions.

Applied to the UG, this general framework enables a more holistic view of reputations than have been considered previously. Even in the UG minigame, the framework identifies not only negative reputations, as considered by Nowak et al. (2000), but also positive reputations: proposers may want to *increase* their offers to responders they know to have rejected a low offer. As the results section shows, an evolutionary model based on positive rather than negative reputations in the ultimatum game gives entirely different dynamics and results.

A series of minigame models generated using the general framework is explored in detail using a combination of explicit solutions and deterministic simulations based on the standard replicator dynamics. Being deterministic, these simulations allow a clear and comprehensive analysis of dynamics using trajectory plots. The issue of endogenous information is shown to be crucial. The first model is a negative-reputation model, and the dynamic system is bistable for a large range of parameter values, with both low- and high-offer equilibria, reflecting the UG's SPNE and relatively equal divisions respectively. Paradoxically, in the low-offer equilibrium proposers have strong information about responders, but in the high-offer equilibria they have very weak information. Despite this, the payoff structure of the game prevents a low-offer equilibrium from being escaped, while it is shown that a high-offer equilibrium can be sustained by very little information.

In contrast to negative reputations, positive-reputation models can evolve out of a state where predominantly low offers are made (and accepted), thus providing a viable theory for how behavioural norms resulting in equal divisions could have emerged where they did not exist before. However, positive-reputation models have only mixed-strategy equilibria, and the equilibria are often unstable, leading to dynamic systems with limit cycles and oscillations. They also feature a significant positive rate of rejections. Finally, I combine both reputation types in a single model, and find that they have complementary effects: positive reputations can bootstrap the system out of a low-offer state, while negative reputations are then effective to stabilize a high-offer equilibrium and improve efficiency. The final part of the results section briefly explores larger models with more complex information structures, showing that sophisticated proposer strategies can evolve, though the computational intensity of simulating such models is a limiting factor that will require further work to resolve.

The chapter's specific advances include the general endogenous information framework, which is flexible and can be used to analyse many different kinds of games where the first player reacts to imperfectly observed past play of the second player, including other types of bargaining interactions and traditional signalling applications. As far as I am aware, this is the first general theoretical treatment of endogenous information in this class of population games.[8]

[8] Nowak et al. (2000)'s second model includes endogenous information but this is an agent-

The full analysis of dynamics in the reputation models, in the light of endogenous information, also contributes towards a comprehensive understanding of the role of the the reputation mechanisms in the UG.

1.5 Methodological notes

The methods used in this dissertation are a mixture of theoretical models and computer simulations. As my research is focused on a specific, concrete model, i.e. the ultimatum game, I do not generally prove statements and draw conclusions for abstract classes of games. In some cases where models and techniques are developed that have general applicability, such as the endogenous information framework in chapter 4, this is pointed out. Similarly, all three of the main research chapters make use of the replicator dynamics rather than abstract evolutionary dynamics. Although this limits generality, I provide justification for this choice throughout the text, and there is no reason to believe that the results will not apply to a more general class of dynamics. In Chapter 3, I develop an individual-level learning model that directly supports the replicator dynamics.

Computer simulations are used in conjunction with theoretical models to avoid arbitrary specifications and inexplicable "black-box" results. A particular emphasis of my research is building theoretical explanations where others have reported only simulation results. Explicit mathematical solutions are derived in cases where helpful, but this is not a specific objective.

In analysing dynamic systems, I attempted to provide defensible and comprehensive descriptions of dynamics across the entire state space, taking care to identify all equilibria and stability conditions as far as possible. In particular, I attempted to take into account and identify the effects of initial conditions, and provide a full account where multiple equilibria exist. For the analysis of dynamics, I generally use computer simulations of deterministic systems to trace out trajectories on graphs, where dimensionality allows. This often provides a clearer intuitive understanding of the dynamic systems than overly abstract general proofs can provide.

I have attempted to provide careful and justified interpretation of theoretical and simulation results for real-world relevance. Chapter 3's objective was

---

based computer simulation that simply tracks specific histories of play, not probability distributions of information, and it cannot easily be used for a comprehensive analysis of dynamics. After having independently completed the endogenous information framework in chapter 4, I discovered that Zhang et al. (2023)'s model, based on a Markov chain to represent transitions between information states, features endogenous information and a similar notion of endogenous information equilibrium to mine. However, they calculate their information equilibrium manually using an exhaustive process and they present no general theory on the information system. In addition, their model only accommodates two reputational statuses (each agent is "good" or "bad") and, only two strategies at any point in time, which are significant limitations. I note further differences in footnote 18 in chapter 4.

largely to illustrate the importance of giving careful attention to this aspect.

Chapter 2

Using Minigames to Explain Imperfect Outcomes in the Ultimatum Game

2.1 Introduction

Evolutionary models of apparently simple games can exhibit surprisingly complicated dynamics. The ultimatum game (UG) presents a particularly interesting case. In this game, a proposer is asked to divide a fixed amount of money between herself and a responder, and the responder can either accept or reject his offered share. Money is paid out according to the proposal if accepted, but both players receiving nothing if rejected. If the players' preferences are completely given by the current game's monetary payoffs, the only rational behaviour for the responder is to accept all positive offers. The proposer, in turn, rationally proposes the smallest positive amount possible. The game's unique subgame-perfect Nash equilibrium (SPNE) therefore allocates all the surplus to the proposer, even though this outcome does not enjoy empirical support in experimental settings.[1]

As with any game where players choose an amount of money, the game has many pure strategies, which necessitates a large number of variables in evolutionary models to track the frequency of each pure strategy. This, along with a large number of weak Nash equilibria, leads to interesting and complicated

[1] Human proposers typically offer between 30 and 50 per cent of the money and responders often reject offers below 20 per cent (Camerer and Thaler, 1995, p. 210, Oosterbeek et al., 2004, Güth and Kocher, 2014, p. 398). Many experimentalists have suggested that the subjects are motivated by social preferences (e.g. Charness and Rabin, 2002; Fehr and Schmidt, 2006; Cox et al., 2007; Blanco et al., 2011). Young and Burke (2001, p. 567) emphasize that the most common modal splits in experimental data are simple fractions, particularly 50%, and notions of fairness appear to be endogenous, with variability between cultures. They argue that these features lend support to a theory of evolutionary origins of bargaining behaviour.

dynamics in which the system can linger for an extended time period close to a Nash equilibrium, only to be followed by a sudden escape and a comparatively rapid transition to another such state, a pattern that can repeat a number of times before the model comes to an eventual final rest (c.f. 'puctuated equilibrium effect' in Young, 1996, p. 112). A thorough appreciation and interpretation of the strategic structure of the UG arguably requires studying the game in an evolutionary setting, because evolutionary models can show which kinds of game-theoretic solutions can be reached and the manner in which they are reached under simple behaviour-updating mechanisms that can loosely proxy for human learning and imitation processes (Mailath, 1998; Friedman, 1998).[2]

Despite a steady flow of literature on the UG including a number of evolutionary models and simulation results,[3] clear theoretical explanations of its evolutionary dynamics have been limited to simplified versions of the game, so-called "minigames", in which proposers are each limited to only two possible offer amounts (e.g. Gale et al., 1995; Mailath, 1998; Huck and Oechssler, 1999; Nowak et al., 2000; Sigmund et al., 2001; Napel, 2003; Uriarte, 2007; Shirata, 2012; Skyrms, 2014; Forber and Smead, 2014; Zhang et al., 2023). Minigames are also often used in experimental work (e.g. Abbink et al., 2001; Smith and Wilson, 2018; Aina et al., 2020).

In the minigame, as in the full game, there is always a NE for each possible offer amount, but only the lowest offer can be a SPNE, so from a classical perspective, the minigame analysis appears to give a reasonably complete analysis of the UG's essential strategic features. But it is not so clear that minigames can give us a full understanding of the dynamics in evolutionary models, since the simplified models lack the multiplicity of equilibria and complicated dynamics of the full UG.[4] I will demonstrate below that the naive minigame analysis fails in a critical respect to account for a well known and important result from computer simulations, namely that evolutionary dynamics in the full game do not generally lead to the SPNE result. The aim of this chapter is therefore to develop a rigorous analysis of the relationship between the minigame evolutionary analysis and the full game, and through this to derive a defensible explanation for the mentioned result in the full game.

For concreteness, this chapter will make use of a particular evolutionary modelling framework, first applied to the UG by GBS. Proposers and responders

[2]This is not to suggest that an evolutionary model necessarily provides the most appropriate explanation for empirical results, since the evolutionary dynamics we study may not be quite the right description of human learning processes, or there may important omitted factors, e.g. other-regarding preferences.

[3]Notable references include Binmore and Samuelson (1994), Gale et al. (1995), Roth and Erev (1995), Vriend (1997), Harms (1997), Peters (2000), Abbink et al. (2001), Napel (2003), Brenner and Vriend (2006) and Poulsen (2007). See also the surveys Debove et al. (2016), Akdeniz and Van Veelen (2023) and the literature referenced therein.

[4]Debove et al. (2016, p. 249) raise a number of different concerns with minigames, which seem to relate more to their interpretation than to their adequacy for explaining dynamics.

are drawn from separate populations of infinite size and randomly matched, with standard continuous-time replicator dynamics determining how population-level strategy frequencies are updated.[5] A deterministic noise (or mutation) term is added that represents occasional errors or innovations. Apart from adding realism, this should prevent the system from "getting stuck" at states simply because some strategies have become extinct. Ordinarily, one would expect the addition of such noise to flush out strategies that are not subgame perfect – if proposers accidentally make low offers every now and then, responders, given enough time, should stop rejecting such offers, which would eventually make low offers rational for proposers.[6] It is therefore surprising that simulation results (reported by GBS) show that the system can come to a permanent rest at stable equilibria which are not SPNE, for example where the responder gets a share of 7 or 9 out of a total amount of 40 (depending on parameter choices).

The GBS paper provides a valuable analysis of a simplified minigame version of the game (reviewed in sections 2.2 and 2.3), in which each player has only two possible strategies, high (equal split) and low (all surplus to proposer).[7] This leads to the finding that the ratio of responder to proposer mutation rates is a crucial factor: a stable imperfect equilibrium exists if and only if this ratio is above a particular threshold. It is clear from GBS's two-dimensional analysis that the role of mutation (noise) is to keep the frequency of the responder strategy that rejects low offers at a sufficient level to make the high offer a best response for proposers, which is necessary to maintain an imperfect equilibrium. Even if responders who reject positive offers get a slightly lower expected payoff because occasional mutant proposers make low offers, the difference is small enough that it is possible for the weak mutation force to balance the inherently stronger selection (learning) force.

When we consider the full UG, this explanation appears to break down. Consider the following modelling set-up: Responder strategies are identified as *demands*, e.g. to demand 9 means to accept a share of 9 or better and reject lower

[5]The replicator dynamics is based on biological models of differential reproduction rates (Taylor and Jonker, 1978) and is not an explicit learning model, though there are close links with more explicit learning models (e.g. Börgers and Sarin, 1997; Schlag, 1998). Such a model is also developed in chapter 3. GBS (p. 83) argue that their results should be similar in "virtually any system in which growth rates of strategy proportions are smooth, increasing functions of expected payoff differences". Indeed, the results of Roth and Erev (1995) using a reinforcement learning model are broadly similar.

[6]The idea of low-probability "mistakes" undoing some Nash equilibria was also the original motivation for Selten's (1975) trembling hand perfect equilibrium: given that every choice in a game is made with some positive probability if mistakes can occur, agents must choose rational behaviour also at off-equilibrium decision nodes (Van Damme, 1991, pp. 10–11). See also Young (1996, p. 109–111).

[7]A responder with a "low demand" strategy would accept all offers, while a "high demand" strategy would accept high and reject low offers.

amounts (we exclude nonmonotonic responder strategies, e.g. rejecting 5 but accepting 3), so that there is exactly one possible responder strategy for each possible offer amount. Proposers have the same strategy set, the amounts here interpreted as *offers*, meaning amounts that the responder would receive if accepted. In the noisy replicator dynamics model, uniform mutation acts to dissipate frequency among many strategies, so that it tends to bring each strategy's frequency to a uniform, relatively small share of the total, e.g. $\frac{1}{40}$ if there are 40 strategies – this can be called the *mutation target* for convenience. The problem is that, to sustain an equilibrium, one particular responder strategy must maintain a frequency above a critical value, which is itself higher than this mutation target.

For example, if proposers are asked to divide an amount of 40, an equilibrium where proposers offer 9 would require that the frequency of rejection of the next lower offer of 8, i.e. the responder strategy "demand 9", be at least $\frac{1}{32}$.[8] This means that the "demand 9" strategy must have a frequency higher than $\frac{1}{32}$, but it seems impossible for mutation to be able to do this because mutation will tend to pull frequencies *down* towards the mutation target of $\frac{1}{40}$, rather than push them up to above $\frac{1}{32}$ as suggested by the minigame analysis. We therefore lack an easy explanation for the simulation results that indicate that there is in fact a stable dynamic equilibrium where proposers offer 9 in this setting (reported by GBS, p. 63, and verified by my own simulations). What prevents the frequency of responders demanding 9 from dropping below the critical $\frac{1}{32}$, and thus the equilibrium from unravelling? On the surface, it appears that responder mutation should be a destabilizing force, so how can simulations show that high responder mutation rates are needed for stability?

As will be explained in this chapter, questions like these can only be answered on the basis of a fuller understanding of the evolutionary dynamics of the model than is available from the minigame. It turns out that the critical responder strategy frequency referred to above is held above the critical threshold *indirectly* through a combination of mutation and selection flows. If proposers are mostly offering 9, there are 31 responder strategies with higher demands that get a substantially lower expected payoff, approximately zero. As a result, in equilibrium, there are very low frequencies of such high-demand responder strategies. But precisely because their frequencies are so low, mutational inflow to them from fitter low-demand strategies occur (which also explains why these suboptimal strategies do not simply die out completely). Then, because their payoffs are so low, the mutational inflow to these high-demand strategies drives a matching outflow due to selection pressure. Some of this

[8] It must not be profitable for the proposer to offer any lower amount, in particular 8, instead of 9. Assuming that an offer of 9 is always accepted, the proposer's expected payoff is 31 when offering 9 and 32 weighted by the probability of acceptance when offering 8.

counterflow helps to increase the frequency of the critical strategy required to stabilize the equilibrium.

What then of the minigame? The chapter will show that a minigame analysis can be very helpful to understand important structural features of the full game, provided that one considers the full game's dynamics in *conditional* form. This involves identifying a suitable subset of the full game's strategy space and deriving *conditional frequency dynamics* for the subset, which is the evolution of frequencies relative to the combined frequency of all the strategies in the subset. Each of the game's Nash equilibria has an associated conditional game in which all *higher* offers and demands are excluded from the strategy sets. I show in section 2.5 that these conditional games are structurally similar to a full UG and the conditional dynamics applicable to them are of the same form as the full game's noisy replicator dynamics.

Since the conditional dynamics are derived directly from the full game's dynamics, an equilibrium in one of the conditional games is also an equilibrium of the full game. This is useful as it helps to focus the analysis on the forces and factors that are relevant to the stability of the equilibrium in question, by removing complicating frequency flows to and from strategies outside the subset from consideration. The puzzle alluded to earlier, that mutation appears to have the "wrong" direction, is resolved as the mutation target for the critical *conditional* frequency is higher in inverse proportion to the number of strategies in the subset, thus providing an explanation for the role of mutation in maintaining the stability of an imperfect equilibrium that matches the situation in the two-dimensional minigame.

The argument continues in section 2.6 by noting that in these conditional games, only two variables really matter for understanding the system's dynamics in the vicinity of the equilibrium: the (conditional) frequencies with which its characteristic amount e is offered and demanded, respectively. This is because, in these conditional games, the best response for each player only depends on the conditional frequency of the corresponding strategy of the other player. This means that the dynamic behaviour of each conditional game near its characteristic Nash equilibrium can be understood in terms of only two variables, and moreover these variables behave, to a close approximation, like the two variables in the two-dimensional minigame. This is supported by a Monte Carlo simulation exercise. In this way, the dynamics of the full game can be related to a series of two-dimensional minigames, each linked to the corresponding conditional game.

Apart from showing that the minigame analysis is indeed applicable to the full game's dynamics, the more precise understanding of the relationship between the minigame and full game yields insights into the dynamics of the full game in the vicinity of the various Nash equilibria. In particular, it is possible to give clear explanations for why some equilibria can be stabilized more easily than others. To achieve this, a graphical analysis, based on partial or unilateral dynamic equilibrium conditions for one population at a time, called selection-mutation equilibrium loci, is developed (introduced in section 2.4). From this analysis the critical role of two different thresholds for a specific responder frequency for each conditional game is identified, namely the mutation target and the critical frequency that would make the corresponding offer a best response for proposers. As explained in section 2.7, these vary systematically with the offer amount and the game's payoffs. Clear explanations are gained for why certain equilibria are more difficult, or even impossible, to stabilize by adjusting the parameters of the evolutionary model, such as selection and mutation strengths.

The chapter therefore aims to advance understanding of the UG, building on the GBS analysis, which should interest theorists as well as experimental researchers. It aims to illustrate that minigames have their use but their connection to a larger game needs to be explicitly derived to avoid misleading explanations and to appreciate their explanatory value in full. The chapter also develops techniques that are likely to have wider applicability, *viz.* the analysis of conditional frequency dynamics, which allows focus on the dynamics within a game with reduced strategy sets, and the graphical selection-mutation equilibrium loci analysis, which helps to understand dynamic forces in a two-dimensional model, particularly when there are multiple forces (selection and mutation in this case).

2.2 The Gale, Binmore and Samuelson model

Following GBS, the UG is set up to allow the proposer a fixed number of possible strategies, each corresponding to an amount to offer to the responder, while the responder has the same number of strategies, each representing a minimum acceptable offer (a "demand").[9] Let \$ be the fixed total amount of money to be divided and $\mathcal{S}$ be the set of possible offers, assumed to contain evenly spaced, strictly positive elements. In the full GBS model, $\$ = 40$ and $\mathcal{S} = \{1, 2, ..., 40\}$. It can be pointed out immediately that there is a pure strategy Nash equilibrium for every element in $\mathcal{S}$, because if the proposer offers an amount of i it is a best response for the responder to demand i, and if i is demanded by the responder, it is a best response to offer i. Let the payoff functions for proposers and responders following strategy i, playing against an opponent following strategy j be $\pi^P(i,j) = \{\$ - i \text{ if } i \geq j, \text{ otherwise } 0\}$ and $\pi^R(i,j) = \{j \text{ if } j \geq i, \text{ otherwise } 0\}$ respectively. To conserve space and aid readability, I will drop P and R superscripts for statements that apply to both proposers and responders where possible, for example π can refer to either π^P or π^R depending on context.

[9] This assumes that a responder who accepts a certain offer will also accept all higher offers and a responder who rejects a certain offer will also reject all lower offers.

Assume there is one infinite population of proposers and one infinite population of responders, and that each agent plays a fixed strategy at any moment in time. Let x_i^P be the fraction of proposers making offers of i and x_i^R be the fraction of responders demanding i at the current point in time. The evolutionary fitness, or expected payoff, for each strategy at a point in time is its expected payoff if the game were played once against a randomly drawn player of the opposite population, given the current frequency distribution of the opposite population. We will also be interested in the expected payoff to an agent playing i conditional on the randomly selected opponent following a strategy in a particular subset $\mathcal{A} \subseteq \mathcal{S}$, written as $\pi(i, \mathcal{A})$, and the average expected payoff for players following strategies in $\mathcal{A} \subseteq \mathcal{S}$ conditional on the opponent following a strategy in $\mathcal{B} \subseteq \mathcal{S}$, written as $\pi(\mathcal{A}, \mathcal{B})$.[10] The (unconditional) current expected payoff to i is then $\pi(i, \mathcal{S})$ and the population's average expected payoff is $\pi(\mathcal{S}, \mathcal{S})$.

Evolutionary dynamics for each population are specified by a noisy version of the standard continuous-time replicator equation:

$$\frac{dx_i}{dt} = \Delta x_i(\pi(i, \mathcal{S}) - \pi(\mathcal{S}, \mathcal{S})) + \delta\left(\tfrac{1}{|\mathcal{S}|} - x_i\right), \tag{2.1}$$

where (following GBS notation) Δ and δ are scaling parameters controlling the rates of the two parts, namely selection and mutation. Selection embodies the core features of an evolutionary process (see Friedman, 1998, p. 16): there is adaptation, so that the frequencies of strategies whose payoff are currently higher (lower) than the population average increase (decrease), but also inertia, so that, for example, a new strategy with a higher payoff than existing strategies will only slowly break into and take over a population. Essentially, it is this inertia, which classical game theory lacks, that allows suboptimal strategies, such as responders rejecting positive offers, to persist for any length of time (let alone indefinitely).

The second term represents noise, or mutation, in the learning process: at a low rate δ, agents adopt random strategies selected from a uniform distribution over $\mathcal{S}$, either due to mistakes or experimentation.[11] Given an infinite

[10] Specifically,

$$\pi(i, \mathcal{A}) = \frac{\sum_{j \in \mathcal{A}} x'_j \pi(i, j)}{\sum_{j \in \mathcal{A}} x'_j}, \ \mathcal{A} \subseteq \mathcal{S} \quad \text{and} \quad \pi(\mathcal{A}, \mathcal{B}) = \frac{\sum_{i \in \mathcal{A}} x_i \pi(i, \mathcal{B})}{\sum_{i \in \mathcal{A}} x_i}, \ \mathcal{A} \subseteq \mathcal{S}$$

where x' refers to frequencies in the opposite population.

[11] GBS set $\Delta = 1 - \delta$ for most of their simulations, suggesting an interpretation that, when learning, an agent will make a mistake with probability δ *instead* of learning. There is no particular mathematical significance, however, of $\Delta + \delta = 1$, since the effects of selection and mutation on x_i are independent in this continuous-time specification. For a more general replicator-mutator equation explicitly linking selection and mutation see Page and Nowak (2002, p. 94).

population, the population-level effect of mutation is fully deterministic, tending to make the distribution more uniform over time, a direct consequence of the assumption of a uniform distribution from which mutating agents select their new strategy (appendix 2 considers non-uniform mutation). Notice that the mutation part of the above equation tends to bring the frequency of each strategy x_i towards $1/|\mathcal{S}|$, (e.g. $\frac{1}{40}$ if there are 40 possible strategies) which I refer to as its *mutation target*, and mutation's effect is proportionally stronger the further x_i is from this target.

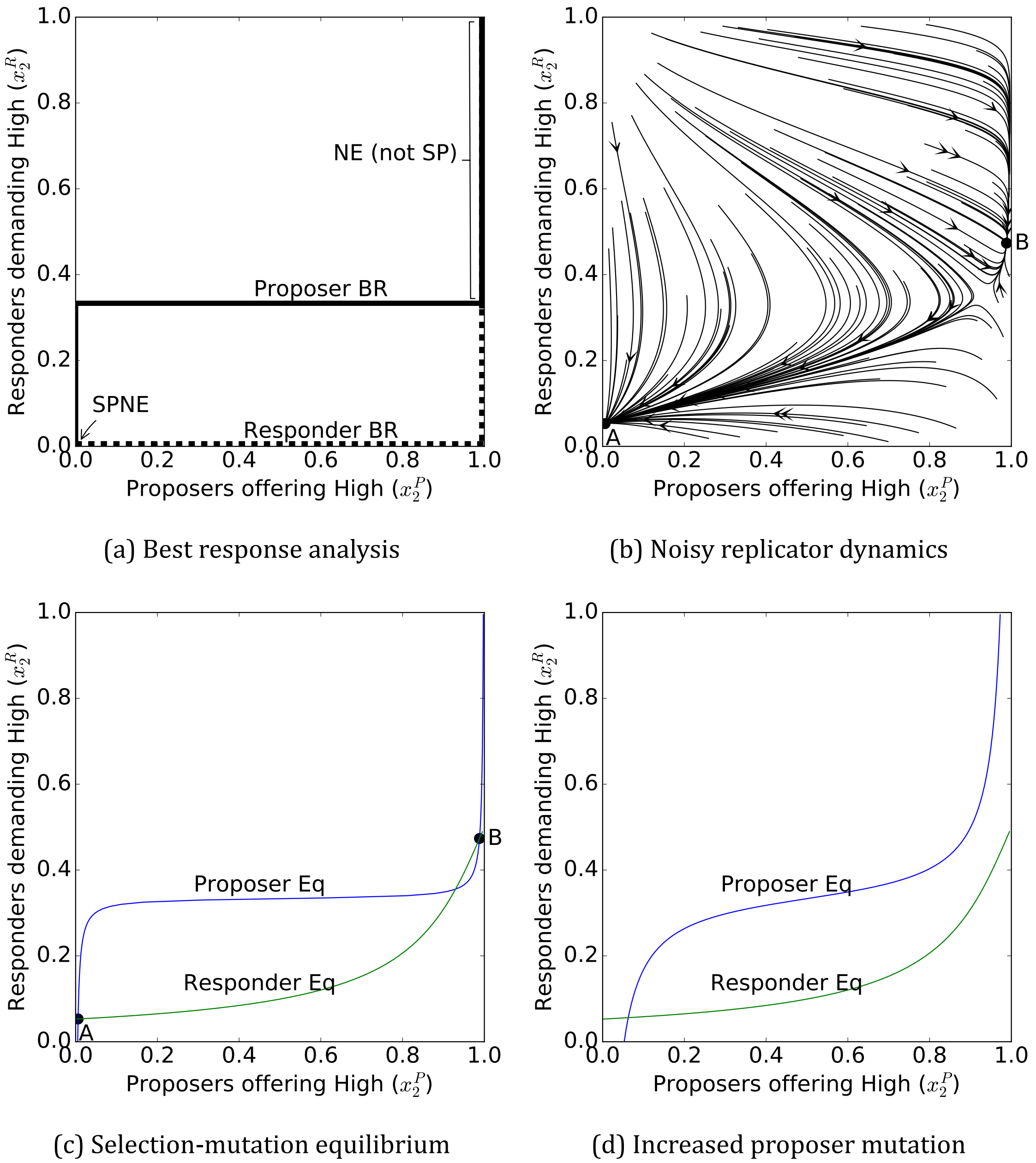
2.3 The minigame

In both the minigame and the full versions of the UG, a Nash equilibrium (NE) exists for each possible offer, and we are interested in which of these correspond to asymptotically stable dynamic equilibria in the evolutionary model based on the game. Consider a version of the minigame with $\$ = 4$ and $\mathcal{S} = \{1, 2\}$. The proposer can offer 1 (Low), thus aiming to keep 3 for themselves, or offer 2 (High) for a proposed equal split. The responder can likewise choose between two strategies, demand 2 (reject Low offers) and demand 1 (always accept). This minigame has a subgame-perfect Nash equilibrium (SPNE) where the proposer offers Low and it is accepted, and also a continuous set of mixed strategy NEs where the proposer offers High and the responder rejects Low offers with a probability of $\frac{1}{3}$ or more (figure 2.1a). The latter is not subgame perfect and would not be robust if the responder were perfectly rational and (even slightly) uncertain about what the proposer will do.

Trajectories of the noisy replicator dynamics (2.1), with $\delta^P = 0.01$, $\delta^R = 0.1$ and $\Delta = 1 - \delta$ (following GBS), are illustrated in figure 2.1b. The dynamics clearly reflect the incentive structure of the game: for proposers, when the rejection frequency is high, there is a rapid move towards High offers, when rejection frequencies are low, there is conversely a rapid move towards Low offers. For responders, there is strong pressure to accept Low offers if the frequency of Low offers is substantial, but there is no selection pressure either way when all offers are High. There are two asymptotically stable equilibria: A, corresponding to the SPNE, and B, the imperfect equilibrium.[12]

Mutation has the effect of pushing the system towards the centre of the graph, most strongly when the current distribution is most unequal, i.e. at the edges. Thus, there is always a small share of proposers making Low offers even when this is clearly suboptimal, i.e. when more than $\frac{1}{3}$ of responders reject Low offers. As long as the current frequency of the responder strategy to reject Low is below the mutation target $\frac{1}{|\mathcal{S}|} = \frac{1}{2}$, mutation will push it upwards.

[12] A and B do not correspond exactly to the underlying game's NEs due to mutation, which always keeps a small share of suboptimal strategies alive – see below.



(a) Best response analysis

(b) Noisy replicator dynamics

(c) Selection-mutation equilibrium

(d) Increased proposer mutation

Figure 2.1: The ultimatum minigame
Notes: $\$ = 4$ and $\mathcal{S} = \{1, 2\}$. (a) indicates multiple Nash Equilibriums, (b) indicates the noisy replicator dynamics (with $\delta^P = 0.01$, $\delta^R = 0.1$), stable equilibria are at A and B (the latter is imperfect), (c) shows Proposer Eq curve where $dx_i^P/dt = 0 \ \forall \ i$ and Responder Eq where $dx_i^R/dt = 0 \ \forall \ i$. (d) is the same except $\delta^P = 0.1$, thus changing the Proposer Eq curve and eliminating the rest point at B.

This explains the main GBS result, namely the existence of a stable imperfect equilibrium (point B). The key to understanding why B is stable is to see that mutation is relatively strong when selection is weak, which is the case for responders when the frequency of High offers by proposers is close to one (because the expected payoffs to the responder's strategies are then almost equal), so that mutation can overcome selection even though δ^R is low, and there is a net upwards movement in a critical region just below B. We can also see why a higher mutation rate for responders and a lower mutation rate for proposers would tend to produce a stable imperfect equilibrium more easily. In the vicinity of B, responder mutation pushes upwards, away from A's basin of attraction, while proposer mutation pushes leftwards, towards A's basin of attraction. GBS (pp. 77–80) formally show that the asymptotic attractor B exists if the ratio of responder and proposer mutation rates divided by the respective ratio of selection rates, $\phi = (\delta^R \Delta^P)/(\delta^P \Delta^R)$ exceeds a critical threshold, in this case $3+2\sqrt{2}$. This threshold derives from the structure and payoffs of the minigame in question and can be regarded as a measure of how difficult it is to create a stable imperfect equilibrium by adjusting selection and mutation rates. Unfortunately the GBS analysis provides little intuition for why the threshold is at this particular level – a gap this chapter seeks to fill by identifying the factors affecting it.

2.4 Selection-mutation equilibrium loci

The stable imperfect equilibrium at B, if it exists, must satisfy the basic requirement for all rest points of dynamic systems, namely that $dx_i/dt = 0$ for all strategies for both players. It is clear from (2.1) that this implies $\Delta x_i(\pi(i,\mathcal{S}) - \pi(\mathcal{S},\mathcal{S})) = -\delta(1/|\mathcal{S}| - x_i)$, i.e. that the forces of selection and mutation have opposite signs and equal magnitudes. It is useful to consider this condition separately for proposers and responders, as is done in figure 2.1c, which shows the loci where selection and mutation balance each other for proposers ($dx_i^P/dt = 0 \quad \forall \quad i$) and responders ($dx_i^R/dt = 0 \quad \forall \quad i$) respectively, given the frequency distribution of the opposite population.[13] These selection-mutation equilibrium curves are useful because they link the system's dynamics (figure 2.1b) to the best-response analysis of the underlying UG. The curves are essentially "smoothed out" versions of the best response (BR) curves (c.f. figure 2.1a), the smoothing due to mutation, and can be interpreted similarly in that proposers will adjust horizontally until they are on their curve and responders adjust vertically until they are on theirs. For example, if the frequency of responders rejecting Low is currently above the Responder Eq curve (but below $\frac{1}{2}$), it means that

[13] The curves are *isoclines* of (2.1), because they represent points where a variable's rate of change is constant, specifically zero. The proposer (responder) eq curve connects all the points in figure 2.1b where the trajectories are vertical (horizontal).

downwards selection pressure is stronger than the upwards mutation force, so the frequency will tend to decline. For proposers and responders alike, where selection pressure is weaker due to a low payoff gradient, mutation asserts itself more strongly so that the point where selection-mutation equilibrium is reached is closer to the mutation target (i.e. the centre of the graph).[14] For proposers, this occurs when the frequency with which Low offers are rejected is closer to $\frac{1}{3}$ and for responders when the frequency of High offers are closer to one.

Where the two curves intersect, $dx_i^P/dt = dx_i^R/dt = 0$ and the system as a whole is at rest. In figure 2.1c, there are three such points, namely A, B and one more which is an unstable saddle point. If we increase the proposer mutation rate (from 0.01 to 0.1) while keeping the responder mutation rate the same (at 0.1), as in figure 2.1d, the proposer curve is visibly more smoothed. The rest point B now disappears: in the vicinity of the missing B, the increased proposer mutation rate implies that selection-mutation equilibrium lies further to the left, but this region lies above the responder equilibrium curve, so the system moves leftwards and downwards in the area between the two curves, until it eventually reaches the low offer equilibrium. Another way to eliminate the rest point B would be to decrease the responder mutation rate, which would have the effect of pushing the reponder eq curve downwards while still anchored on the right-hand side at $\frac{1}{2}$, becoming relatively more "square" like the BR curve in figure 2.1a, so that the responder eq curve passes to the right of the proposer eq curve near B in figure 2.1c without crossing it. We can conclude that a stable imperfect equilibrium exists whenever the two curves cross three times because the basic shapes are due to the game's type.[15]

In addition to mutation rates, other factors that affect the shapes and positions of these curves, and therefore also the existence of a stable imperfect equilibrium, can now be identified. A critical role is played by two important thresholds for the frequency with which Low offers are rejected. The first threshold is $\frac{1}{2}$, the mutation target for responders, which (as previously discussed) derives from the assumption of uniform mutation and the fact that there are two strategies available to responders in this minigame. If the mutation target were lower, it would be more difficult to stabilize an imperfect equilibrium. As will be explained later, this becomes important in the analysis of the full game because

---

[14] Setting $dx_i/dt = 0$ in (2.1), we can also write the rest value for x_i as the implicit function,

$$x_i^* = \frac{\frac{1}{|\mathcal{S}|}}{1 - \frac{\Delta}{\delta}(\pi(i,\mathcal{S}) - \pi(\mathcal{S},\mathcal{S}))},$$

which shows that as $\pi(i,\mathcal{S}) - \pi(\mathcal{S},\mathcal{S}) \to 0$, $x_i^* \to \frac{1}{|\mathcal{S}|}$ (the mutation target).

[15] This gives a graphical and intuitive explanation for the GBS condition for ϕ mentioned in the previous section.

then there are indeed more than two strategies. The second critical threshold is the frequency of rejections of Low offers that would make it worthwhile for proposers to offer High, which is $\frac{1}{3}$ in this game. This threshold could be made lower by increasing the expected loss the proposer would suffer if her lower offer were rejected relative to the gains she would obtain from a lower offer that is accepted. For example, if \$ were increased from 4 to 5, then the potential gain from offering 1 instead of 2 remains the same, but the potential loss due to rejection would be 3 instead of 2 and therefore a rational proposer would offer 1 instead of 2 only if the probability of rejection were $\frac{1}{4}$ or lower (assuming an offer of 2 is always accepted). This threshold corresponds to the height of the horizontal part of the proposer eq curve in figure 2.1c and it would clearly make it easier to stabilize an imperfect equilibrium near B if this part of the curve were lower.[16]

2.5 Conditional frequency dynamics

The full game has two similar thresholds for *each* imperfect equilibrium: the minimum frequency for a specific responder strategy required to make the corresponding offer a best response for proposers and the mutation target for that responder strategy. As explained in the introduction, however, there are imperfect equilibria known to be evolutionarily stable where the mutation target is much lower than the minimum level required for an equilibrium, thus seeming to defy the minigame analysis's logic.[17]

[16]If the second threshold is *very* low, and the responder to proposer mutation ratio is high enough, it is possible that the responder eq curve lies entirely above the proposer eq curve in the vicinity of A, so that there is in fact no longer a stable equilibrium near the subgame-perfect solution of the game. In this case, the two curves would cross only once at B, which would then be a stable imperfect equilibrium. Except for this case, the stable imperfect equilibrium would exist if and only if the curves crossed each other exactly three times. What if the curves crossed once near A but there were also a point of tangency near B? The point of tangency would be immediately adjacent to a region that does not converge to it and the point would therefore not be asymptotically stable.

[17]GBS (p. 63) report simulation results for their full model (with $\$ = 40$ and $\mathcal{S} = \{1, 2, ..., 40\}$). From uniform starting frequencies, with $\delta^P = 0.01$, $\delta^R = 0.1$ and $\Delta = 1 - \delta$, the system eventually settles at a stable equilibrium where proposers offer 9 and responders play a mix of strategies demanding 9 or less. There are also low frequencies of other strategies due to mutation (which we can ignore for the moment as their effect is insignificant). For the situation to reflect a stable equilibrium it must be a best response for proposers to offer 9 instead of a lower amount, say 8. This requires x_9^R, the frequency of responders that would reject an offer of 8 but accept 9, to be above a certain threshold, which can be calculated as approximately $\frac{1}{32}$ (the payoff to a proposer offering 9 is 31, while the expected payoff to a proposer offering 8 is $(1 - x_9^R)32$. If $31 \geq (1 - x_9^R)32$ then $x_9^R \geq \frac{1}{32}$). Indeed, x_9^R eventually comes to rest at approximately 0.1047 (my own simulation of the same model), which is comfortably above the required threshold. Yet if $\frac{1}{|\mathcal{S}|} = \frac{1}{40}$ and $x_9^R \approx 0.1047$, the mutation part of equation (2.1) will

A different viewpoint is necessary. Suppose we want to understand the dynamics of the system in the vicinity of an equilibrium in which proposers offer an amount e. Responders with demands higher than e will get a payoff of approximately zero so the strong selection force will tend to diminish the frequency of such demands quickly. Given that the frequency of demands higher than e will be very low – approximately zero if the rate of mutation is much lower than that of selection – proposers making offers higher than e will face strong negative selection pressure as well. The frequency of strategies higher than e for both populations will therefore tend to drop close to zero (see also Lemma 1 in Peters, 2000), and, in the absence of large coordinated mass mutations, will stay close to zero *forever*, because strategies higher than e are never best responses to e or below for either player in the ultimatum game (excepting the case where proposers offer zero so that all responder strategies would be weak best responses, but this is not important).

The evolutionary trajectory has effectively lead to a truncated game, so that from a certain point in time onwards, both populations place significant frequency only on offers/demands of e and lower values, and approximately zero on all strategies higher than e – only the weak mutation force keeps such strategies' frequencies slightly positive. If we completely remove all offers/demands higher than e from both players' strategy sets, we have a reduced game which is itself a full ultimatum game, with the special feature that e is the highest possible offer that can be made. This can also be called a conditional game, as it reflects the full game's strategic structure conditional on both players choosing strategies in the reduced strategy sets (offers/demands of e or lower). As already noted, offers/demands higher than e will not be best responses to demands/offers of e or lower, so a best response in the conditional game is a conditional best response in the full game. Thus, every NE in a conditional game maps directly to a NE in the full game.

The evolutionary dynamics of the full game can be related to the conditional game, through the use of *conditional frequency dynamics*, which considers the evolution of frequencies relative to the combined frequency of all strategies in the relevant subset. In what follows, I will assume that the game has been normalized so that the set $\mathcal{S}$ consists of successive integers starting at 1. Let $\mathcal{E} := [1, e]$ be the subset of strategies in $\mathcal{S}$ smaller than or equal to e, let $X_{\mathcal{E}}$ be the combined frequency with which all strategies in $\mathcal{E}$ are played and let ξ_i be the frequency of i being played conditional on a strategy in $\mathcal{E}$ being played ("conditional frequency" for short):

$$X_{\mathcal{E}} := \sum_{i \in \mathcal{E}} x_i \quad \text{and} \quad \xi_i := \frac{x_i}{X_{\mathcal{E}}}$$

clearly tend to decrease, not increase x_9^R. It appears as if mutation cannot play the same role as in the minigame to stabilize a subgame-imperfect equilibrium.

The dynamics of ξ_i can be derived from (2.1) by a simple application of the quotient rule:

$$\frac{d\xi_i}{dt} = \frac{X_\mathcal{E}\frac{dx_i}{dt} - x_i\frac{dX_\mathcal{E}}{dt}}{X_\mathcal{E}^2} \tag{2.2}$$

$$= \frac{X_\mathcal{E}\left[\Delta x_i(\pi(i,\mathcal{S}) - \pi(\mathcal{S},\mathcal{S})) + \delta\left(\frac{1}{|\mathcal{S}|} - x_i\right)\right]}{X_\mathcal{E}^2}$$

$$- \frac{x_i\left[\Delta X_\mathcal{E}(\pi(\mathcal{E},\mathcal{S}) - \pi(\mathcal{S},\mathcal{S})) + \delta\left(\frac{|\mathcal{E}|}{|\mathcal{S}|} - X_\mathcal{E}\right)\right]}{X_\mathcal{E}^2}$$

$$= \Delta\xi_i(\pi(i,\mathcal{S}) - \pi(\mathcal{E},\mathcal{S})) + \delta\frac{|\mathcal{E}|}{|\mathcal{S}|}\frac{1}{X_\mathcal{E}}\left(\frac{1}{|\mathcal{E}|} - \xi_i\right) \tag{2.3}$$

This shows that it is possible to consider the evolution of conditional frequencies within a subset of strategies in isolation, with reference to the subset's average payoff $\pi(\mathcal{E},\mathcal{S})$, but not the population average payoff $\pi(\mathcal{S},\mathcal{S})$. This is a general property of the replicator dynamics, and is not dependent on the type of game nor the choice of subset $\mathcal{E}$. The above equation references the payoffs to strategies in $\mathcal{E}$ as played against the entire set of strategies in the opposite population $\pi(\bullet,\mathcal{S})$, but it is possible to convert these to expected payoffs conditional on *both* players' strategies being in $\mathcal{E}$, by making use of the ultimatum game's payoff structure. Write the payoff to a strategy as a weighted average of payoffs against opposite population strategies respectively within and outside of $\mathcal{E}$:

$$\pi^P(i,\mathcal{S}) = X_\mathcal{E}^R\pi^P(i,\mathcal{E}) + (1 - X_\mathcal{E}^R)\pi^P(i,\mathcal{S}\setminus\mathcal{E})$$
$$\pi^R(i,\mathcal{S}) = X_\mathcal{E}^P\pi^R(i,\mathcal{E}) + (1 - X_\mathcal{E}^P)\pi^R(i,\mathcal{S}\setminus\mathcal{E})$$

where $\setminus$ is the set difference operator. The UG's payoff functions assign $\pi^P(i,\mathcal{S}\setminus\mathcal{E}) = 0$ and $\pi^R(i,\mathcal{S}\setminus\mathcal{E}) = \pi^R(e,\mathcal{S}\setminus\mathcal{E})$ for all $i \in \mathcal{E}$, so (2.3) can be written as,

$$\frac{d\xi_i^P}{dt} = \Delta^P X_\mathcal{E}^R\xi_i^P\left[\pi^P(i,\mathcal{E}) - \pi^P(\mathcal{E},\mathcal{E})\right] + \delta^P\frac{|\mathcal{E}|}{|\mathcal{S}|}\frac{1}{X_\mathcal{E}^P}\left(\frac{1}{|\mathcal{E}|} - \xi_i^P\right)$$
$$\frac{d\xi_i^R}{dt} = \Delta^R X_\mathcal{E}^P\xi_i^R\left[\pi^R(i,\mathcal{E}) - \pi^R(\mathcal{E},\mathcal{E})\right] + \delta^R\frac{|\mathcal{E}|}{|\mathcal{S}|}\frac{1}{X_\mathcal{E}^R}\left(\frac{1}{|\mathcal{E}|} - \xi_i^R\right) \tag{2.4}$$

We see that these equations are a form of the noisy replicator equations for the reduced game; they only reference payoffs of strategies within $\mathcal{E}$ played against opponent strategies also in $\mathcal{E}$. Compared to the dynamics in equation 2.1, selection is scaled by $X_\mathcal{E}^R$ for proposers and by $X_\mathcal{E}^P$ for responders: If the entire opposite population followed strategies in $\mathcal{E}$, then selection would be "full strength", but otherwise effective selection would be proportionally weaker. This is a direct result of the fact that, for both proposers and responders, all strategies in $\mathcal{E}$ give the same payoffs when the opposite side is following a strategy outside $\mathcal{E}$. Mutation is scaled by $\frac{|\mathcal{E}|}{|\mathcal{S}|}\frac{1}{X_{\mathcal{E}}}$. The second factor represents the inverse of the share of the current population following frequencies in $\mathcal{E}$, so that mutation effectively becomes infinitely strong relative to selection as $X_{\mathcal{E}} \to 0$.[18] The first factor shows that mutation's effect on conditional frequencies will be diluted by the presence of more strategies outside $\mathcal{E}$ – thus mutation is weak when e is low.

If $X^P_{\mathcal{E}}$ and $X^R_{\mathcal{E}}$ were constants, the conditional frequencies would move in exactly the same way as the corresponding frequencies in the reduced game, with modified selection and mutation rates as indicated. But I have argued above that, in the vicinity of an e-offer equilibrium, these variables will tend to move to approximately one and stay there forever, thus suggesting that $X^P_{\mathcal{E}}$ and $X^R_{\mathcal{E}}$ will be *almost* constant. We can therefore hypothesize that it is possible to analyse the full game's dynamic properties near the e-offer equilibrium by reference to the dynamics of a reduced game – further support for this from simulation results is presented below.

Let us now briefly return to the puzzle sketched above, where it was pointed out that mutation does not appear to play the required role of keeping the unconditional frequency x^R_9 high in the GBS model. Now consider the effect of mutation on the *conditional* frequency with which 9 is demanded: this variable comes to rest at approximately 0.1057 in the simulation, but (2.4) reveals that the mutation target for ξ^R_9 is $\frac{1}{|\mathcal{E}|} = \frac{1}{9}$ when $e = 9$, which is *higher* than 0.1057, indicating that mutation does tend to push ξ^R_9 *upwards* at the rest point (balancing the downwards selection force). Responder mutation is then responsible for maintaining a high enough (conditional) frequency of a high demand strategy, thus stabilizing the imperfect equilibrium according to the existing explanation we had for the (stand-alone) minigame. In short, mutation's direction for the critical variable reverses and the puzzle is resolved when we work with conditional rather than unconditional frequency dynamics.

The source of the difficulty with the unconditional dynamics is revealed by decomposing it into two parts using (2.2),

$$\frac{dx_i}{dt} = X_{\mathcal{E}}\frac{d\xi_i}{dt} + \xi_i\frac{dX_{\mathcal{E}}}{dt},$$

the first describing selection and mutation flows within the subset $\mathcal{E}$, and the second describing selection and mutation flows between $\mathcal{E}$ as a whole and the rest of $\mathcal{S}$. If mutation's effect on dx_e/dt is negative, but mutation's effect on $d\xi_e/dt$ is positive, then that simply shows that mutation's effect on $dX_{\mathcal{E}}/dt$

[18] To understand this, consider that when $X_{\mathcal{E}}$ is very small, inflows to $\mathcal{E}$ will be predominantly due to mutation from strategies outside $\mathcal{E}$.

must be negative and comparatively large. This makes sense because $X_{\mathcal{E}} \approx 1$ and the mutation target for $X_{\mathcal{E}}$ is $|\mathcal{E}|/|\mathcal{S}| < 1$. Since strategies outside $\mathcal{E}$ are substantively suboptimal, this mutational outflow from $\mathcal{E}$ creates a corresponding inflow due to selection, some of which accrues to x_e. One could therefore say that the mechanism that creates a stable imperfect equilibrium at e is the same as in the minigame, but it is obscured by flows to and from the subset $\mathcal{E}$ that have no counterpart in the minigame.

2.6 The minigame approximation

While the conditional dynamics described above already allow certain insights, the conditional games they refer to are obviously not *minigames*, because generally they have more than two strategies available to each player. Nevertheless, in this section, I argue that the two-dimensional minigame dynamic analysis represents a good approximation of the dynamics applicable to the conditional games because the UG minigame's two state variables behave much like two particular variables of interest in the conditional game, especially near a relevant imperfect equilibrium.

Suppose we are interested in understanding the full game's dynamics in the vicinity an equilibrium (typically to answer questions related to its stability) in which proposers offer e and responders accept e, where e is higher than the lowest available positive offer. Then, as argued earlier, we may consider the conditional game characterized by having e as its highest available offer/demand and we can determine the evolution of conditional frequencies ξ_e^P and ξ_e^R. We know from the previous section that these two variables evolve approximately like the corresponding frequencies in the conditional (reduced) game. It can be claimed that it is only necessary to consider the dynamics of ξ_e^P and ξ_e^R to gain a reasonably complete understanding of the dynamic forces of the system near the equilibrium in question, and that these variables do indeed behave like the (offer High, demand High) strategy frequencies in the UG minigame.

The argument for both claims proceeds in three parts: Firstly, I will show that for both proposers and responders, the condition for e to be a best response references no variable except the conditional frequency with which e is played by the opposite population. Secondly, these best response conditions will be decisive not only to describe static equilibria, but also for the dynamics of ξ_e^P and ξ_e^R, because the dynamics for ξ_i depend largely on whether or not i is a best response, or more accurately on how close its best-response condition is from being met, so that the effect of all other variables on the evolution of ξ_e^P and ξ_e^R is approximately zero. Thirdly, simple reasoning along these lines suggests that the dynamics of ξ_e^P and ξ_e^R, as summarized by selection-mutation equilibrium loci in conditional strategy space, should be the same as in the UG minigame (section 2.4). This will be confirmed by a Monte Carlo sensitivity analysis showing that the selection-mutation equilibrium loci are as expected irrespective of random variations in variables other than ξ_e^P and ξ_e^R.

It is trivial that for responders, e is a best response in the conditional game (and consequently in the full game) only when $\xi_e^P = 1$. Appendix 1 contains a proof that e is a best response for proposers if and only if

$$\xi_e^R \geq T(e) := \frac{1}{\$ + 1 - e}. \tag{2.5}$$

As in the minigame analysis this reflects a weighing up of the risk and reward of offering less than e when some responders might reject a lower offer. The condition follows specifically from a comparison of $\pi^P(e, \mathcal{E})$ against $\pi^P(e-1, \mathcal{E})$. This single comparison turns out to be sufficient for establishing whether e is a best response offer, because even lower offers imply progressively higher probabilities of rejection, and the expected losses grow faster than the expected gains as the offer is lowered. This result depends on responder frequencies x_i^R being nonincreasing in i, a condition that the system has an inherent tendency to establish and maintain because the expected payoff for responders is nonincreasing in i (see appendix 1 for details). The upshot is that we only need to consider ξ_e^P and ξ_e^R to determine whether e is a best response for responders and proposers respectively.

Notice that these best response conditions, taken as equalities, are indifference conditions: if $\xi_e^P = 1$ then all responder strategies in $\mathcal{E}$ deliver the same payoff, and if $\xi_e^R = T(e)$, then $\pi^P(e, \mathcal{E}) = \pi^P(e-1, \mathcal{E})$. As mentioned previously (section 2.4), and as is evident from (2.3), when the payoff gradient within a group of strategies is low, selection will be effectively weak and mutation will assert itself more strongly. The closer the indifference conditions are to being met, i.e. the closer ξ_e^P is to 1 for responders, and the closer ξ_e^R to $T(e)$ for proposers, the smaller the payoff gradient between e and its best alternative in $\mathcal{E}$,[19] which should be reflected in a dynamic equilibrium by a rest value of ξ_e closer to its mutation target (i.e. a more uniform distribution).

It will again be useful to consider a partial dynamic equilibrium, where the rest condition ($d\xi_i/dt = 0$) only needs to be true for one of the populations at a time, while the other is systematically varied across its state space, resulting in selection-mutation equilibrium loci similar to those in figures 2.1c and 2.1d. Irrespective of other details of the proposer distribution, if $\xi_e^P = 0$ (and held fixed), no responders in $\mathcal{E}$ would want to demand e, so ξ_e^R will rapidly move to and eventually rest at a very low level with only mutation keeping it slightly away from zero. If $\xi_e^P = 1$, all responders in $\mathcal{E}$ get the same payoff, so selection will have no effect and mutation will result in responders moving slowly

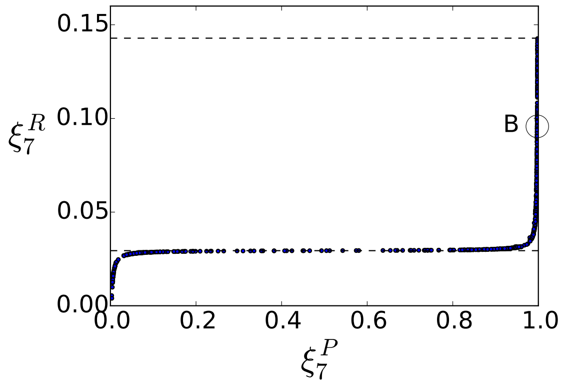
[19] Recall that $\pi(i, \mathcal{E})$ is a linear combination of payoffs of i against strategies in $\mathcal{E}$, the weights being the latter's frequencies, so the payoff difference between two strategies will vary linearly with opposite population frequencies.

towards a uniform distribution. If $\xi_e^R = 1$, so that all responders in $\mathcal{E}$ demand e, we can expect ξ_e^P to come to rest very close to one (mutation will keep it slightly away from one) and if $\xi_e^R = 0$, then ξ_e^P will come to rest close to zero. A transition therefore has to take place at some point as ξ_e^R moves from one to zero, and following earlier reasoning, this transition should be at $\xi_e^R = T(e)$. Close to this threshold, selection will be weak and mutation should push proposers towards uniformity, inducing a "smooth" transition similar to the minigame. These statements do not give precise values for the extremal rest points and we do not know the precise curvature of the loci, but we have sufficient information to deduce that the shapes of the selection-mutation equilibrium loci for ξ_e^P and ξ_e^R must qualitatively resemble those in the standalone minigame.

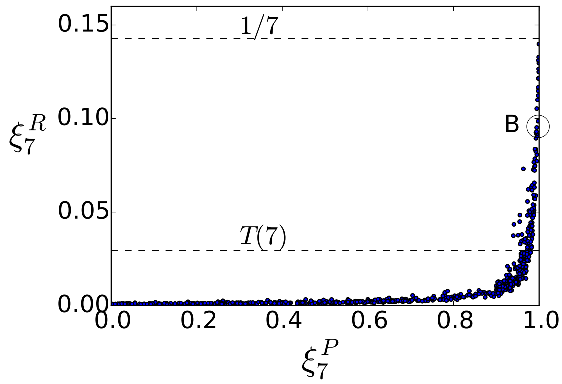
We should therefore expect evolutionary dynamics in a conditional game to broadly follow the minigame's dynamic analysis. It may be argued that, mathematically, variations in conditional frequencies of strategies in $\mathcal{E}$ other than e may affect the dynamics of ξ_e^P and ξ_e^R, through their effects on $\pi(\mathcal{E},\mathcal{E})$ and $\pi(e,\mathcal{E})$, especially when ξ_e's are low, and stability questions may hinge on subtle variations in weak forces, so even very small effects may matter. To test the claim that the minigame analysis can be usefully applied to conditional games, I conducted a Monte Carlo exercise using simulations of the full model. The objective was to generate a large number of selection-mutation equilibrium points for one population at a time given random distributions for the other population, thus tracing out "numerical" selection-mutation loci.[20] Apart from

[20] Here is a short description of the algorithm, starting with the construction of the proposers' selection-mutation equilibrium locus. Assume we know e. A random value is drawn from a uniform distribution between 0 and 0.1, which will be the combined frequency of responder strategies *outside* $\mathcal{E}$. The frequencies for strategies outside $\mathcal{E}$ are assigned by drawing values from a Pareto distribution (with characteristic parameter randomly selected from a uniform distribution between 0 and 10 for each simulation), and scaling them so that their combined frequency matches the desired value. There is no particular justification for this procedure other than that it often produces "extreme" distributions highly skewed to one or a few of the strategies and also sometimes fairly uniform distributions, so that the effects, if any, of variations in the distribution should become apparent. Now, another random variable is drawn which will be the conditional frequency of e, or ξ_e^R. The latter is drawn from a uniform distribution between zero and $1/e$ (for some simulations I also used a narrower distribution designed to get more data points in regions of interest). The value cannot exceed $1/e$ since then it would then be impossible to assign frequencies to demands lower than e without breaking the requirement that lower demands must have equal or higher frequency than that of e (see Appendix 1). The frequencies of the remaining strategies, for demands lower than e, are then determined in sequence. For $e-1$, an allowable interval is determined, with the lower bound equal to x_e^R and the upper bound determined by the amount of unallocated frequency in the unit interval (again keeping in mind the restriction that lower demands must have at least equal frequency). A uniform random value is drawn from the allowable interval and assigned to x_{e-1}^R, and the process is repeated for lower demands until the entire distribution has been determined. Having determined the responder distribution, it is held fixed while letting proposer frequencies evolve normally according to the original full model specification (2.1). (This is easy to do in the full model by setting responder selection and mutation rates to zero.) When proposer frequencies

graphically illustrating how ξ_e^P determines the dynamics of ξ_e^R, and vice versa, this exercise is an effective sensitivity analysis: if there were significant effects of variations in the opposite population conditional frequencies other than ξ_e, it should show up as noise in the results.



(a) Proposers



(b) Responders

Figure 2.2: Selection-mutation equilibrium for conditional frequencies
Notes: Results from Monte Carlo simulation exercise. (a) shows final rest values for ξ_7^P given various random strategy distributions for responders, (b) shows final rest values for ξ_7^R given various random strategy distributions for proposers. $\delta^P = \delta^R = 0.1$, $\$ = 40$ and $e = 7$.

Figure 2.2 show the results for (a) 650 simulations for proposers, each time holding the responder frequency distribution constant, and (b) 650 simulations for responders, each time holding the proposer frequency distribution constant, with $\delta^P = \delta^R = 0.1$, $\$ = 40$ and $e = 7$. Both graphs show clear relationships, confirming that the main factor determining the value at which a conditional frequency comes to rest is the corresponding conditional frequency in the opposite population. There is very little noise in (a), indicating that, given ξ_7^R and the restriction of decreasing frequencies in offers in $\mathcal{E}$, the remaining details of

stop evolving (I deem this to have occurred when the absolute difference in frequency from one round to the next for no frequency in the model exceeds $1/10^8$), we know that selection and mutation is balanced and we can calculate the rest value of ξ_e^P. This procedure is repeated many times for different values of ξ_e^R, thus tracing out the selection-mutation equilibrium locus for proposers.

The whole exercise is then repeated for responders, this time fixing ξ_e^P at various values and letting only responder frequencies evolve. The only difference is that since there is no restriction on the frequencies of proposer strategies below e, they are simply drawn from a Pareto distribution and scaled.

the responder distribution hardly matters.[21] The relationship for responders in (b) is noisier, indicating that the details of the proposers' conditional distribution apart from 7 matters to a somewhat greater extent, but the rest value for ξ_7^R is nonetheless overwhelmingly driven by ξ_7^P.

We can also easily verify that the shapes of the curves in figure 2.2 are what we expected from our analysis above, particularly the locations of the extremal points and the transition at $T(7)$. The conditional frequencies behave approximately as if they were in a standalone UG minigame, and our insights from the minigame analysis may be applied to an equilibrium of the full game. Mutation plays the expected role of moving the relative distributions towards uniformity when selection is weak, thus smoothing the proposer's locus close to $\xi_7^R = T(7)$ and the responder's locus close to $\xi_7^P = 1$. However, compared to figure 2.1d, which has the same mutation and selection parameter values as these simulations, mutation appears much weaker. This is exactly what we would expect from (2.4) given the presence of the factor $|\mathcal{E}|/|\mathcal{S}|$, which indicates that mutation in conditional frequencies is weaker than in the full model. By reference to the GBS result that only the ratio of responder and proposer mutation rates matter, the factor $|\mathcal{E}|/|\mathcal{S}|$ should not affect the existence of an asymptotic attractor in the equilibrium because it affects proposers and responders alike.

2.7 How easily can an imperfect equilibrium be stabilized?

For the parameter values used in the exercise above, the full model has a stable imperfect equilibrium where an offer of 7 is made by proposers, indicated on the graphs by the midpoint of the circles labelled B. At this point, where both ξ_7^P and ξ_7^R are at rest, the roles of selection and mutation to stabilize the equilibrium are the same as in the standalone UG minigame: mutation pushes responders upwards and proposers leftwards, while selection pushes in the opposite directions. Responder mutation must be strong enough to keep proposers from being overly tempted to make lower offers, while proposer mutation must be weak enough to not punish responders too harshly for high demands. If, for example, the proposer mutation rate were increased, ξ_7^P will be pushed lower, which is leftwards on the graph, which strengthens downwards responder selection, moving the system to the SPNE's basin of attraction (refer to figure 2.1b). We know from the minigame analysis that if the stable imperfect equi-

[21] There is in fact a small amount of noise that is visible when zooming in (not shown) – the relationship is not perfectly monotonic. Recall that a small fraction of proposers will play each strategy due to mutation, so the details of the responders' relative distribution below 7 will slightly affect the average payoff to proposers, even if not affecting $\pi^P(e, \mathcal{E})$. This affects the strength of selection and therefore the rest value of ξ_7^P, but the effect is small because the mutation rate is low.

librium exists, there should be three rest points in the space of the graphs, indicating that the "true" selection-mutation equilibrium loci should cross three times.[22] Why, given the same selection and mutation rate parameters, or the same ϕ's (see p. 27), is there a non-SPNE asymptotic attractor in the full game where $e = 7$, but not for the minigame illustrated in figure 2.1d?

Consider the two critical ξ_7^R thresholds indicated on figure 2.2. The first, $T(7) \equiv 1/(\$ + 1 - e) = \frac{1}{34}$, is determined by the gains and losses to proposers of adjusting their offer downwards, as explained previously. In the vicinity of $T(e)$, proposer selection becomes relaxed, and the rightwards force is reduced. Since this force is needed to counteract leftwards mutation in order to maintain the equilibrium, it helps if this threshold is low, so that effective selection is strong near the (potential) asymptotic attractor. In general terms, it is easier to maintain an equilibrium in which proposers have much to lose by lowering their offers. In the minigame, where $\$ = 4$, and the high offer is 2, the potential loss to offering lower is not so great as in the full game when the offer is 7 (and $\$ = 40$), which explains why in the minigame the high offer equilibrium is not as easily stabilized as the equilibrium in the full game.

The second critical threshold is $\frac{1}{7}$, which is the mutation target for ξ_7^R. It is determined by the shape of the mutational noise (assumed uniform here) and by the number of strategies in the subgroup $\mathcal{E}$ – the fewer there are, the higher this threshold will be, since uniform mutation tends to push conditional frequencies towards equal shares. The further ξ_7^R is below this target, the stronger the effective upwards mutation force will be, so to maintain an equilibrium, it helps if there are fewer strategies in $\mathcal{E}$. Graphically, the lower the $T(e)$ horizontal "bar" and the more vertically stretched the responder locus (by a high mutation target), the more likely it is that the two loci will cross three times instead of once, and therefore a non-SPNE asymptotic attractor will exist. For successively lower-offer equilibria, the mutation target $\frac{1}{e}$ moves upwards, while the $T(e)$ threshold moves downwards, thus both effects conspire to make it easier for the selection-mutation loci of the proposers and the responders to cross thrice and therefore to stabilize a subgame imperfect equilibrium.

Selection and mutation rates are still relevant; for example it is possible to stabilize an equilibrium with a modal offer of 9 by boosting the rate of mutation for responders or reducing the rate of mutation for proposers sufficiently. However, this will not always be possible: if the $T(e)$ threshold were actually higher than the mutation target $\frac{1}{e}$ threshold, no amount of responder mutation

[22]This is it not to suggest that the full game has only three rest points, since variation in dimensions not shown in these graphs – specifically the relative distributions of frequencies below e – can lead to multiple other rest points. The vicinity of the bottom-left rest point in this space can be thought of as a "portal" leading to other dimensions containing their own rest points. The "multiple valleys" metaphor of Binmore and Samuelson (1999) is also useful in this regard.

would make the loci cross three times, so an asymptotic attractor can only exist if,

$$
\begin{aligned}
\frac{1}{\$ + 1 - e} &< \frac{1}{e} \\
e &< \frac{\$ + 1}{2}
\end{aligned}
\tag{2.6}
$$

which, if $\$ = 40$, means that the highest offer equilibrium that can be stabilized by changing mutation and/or selection rates is 20 (which would require *extremely* noisy responders relative to proposers).

2.8 Conclusion

Can a two-dimensional minigame analysis of the ultimatum game's noisy evolutionary dynamics account for crucial aspects of the full game with a large number of pure strategies? A naive approach that simply ignores the fact that the full game has more than two strategies provides a loose analogy at best. In particular, it is difficult to explain why there are stable imperfect equilibria in the full game, because in the minigame, an essential part of the explanation is that mutation keeps responder rejection of low offers at a high enough frequency for the equilibrium to be maintained, but in the full game, the direct effect of mutation on a strategy that would reject a lower offer was shown to *lower* its frequency. A more careful approach, taking explicit account of the relationship between the two-dimensional minigame and the full game, however, gives the more positive answer that the minigame does have much relevance for the full game, though the relationship is not quite straightforward.

It was illustrated that, using an appropriate conditional frequency evolutionary analysis, in which mutation and selection flows within a subgroup of strategies are studied in isolation, the simple minigame explanation can be restored. A conditional frequency is the frequency with which an amount is selected (as offer or as demand) conditional on all higher amounts not having been selected. The particular structure of the ultimatum game causes dynamics in this conditional strategy space to be approximately independent of higher amount strategy frequencies when the full system is near an associated equilibrium, and the dynamics have the same shape as an ultimatum game with a reduced strategy set. Finally, it is established that the dynamics of any particular conditional strategy are almost entirely determined by the conditional frequency of the corresponding strategy in the opposite population, and thus that a two dimensional analysis based on the ultimatum minigame is feasible and appropriate. The Monte Carlo sensitivity analysis in section 2.6 is particularly encouraging, strongly suggesting that the pair of conditional frequencies behaves the same as the two variables in a standalone minigame.[23]

In practical terms, a particular equilibrium offer can be maintained at a relative frequency near 100% by the evolutionary system as long as the frequency of the responder strategy that demands that same amount is kept above a critical threshold (by mutation, directly and/or indirectly), but if the threshold is breached proposers will lower their offers and the system will typically reach a new equilibrium with a modal offer which is the next lower amount in the strategy set. Given that the new offer is a lower amount, the new equilibrium will be intrinsically easier to stabilize so may therefore be maintained for longer, or even permanently, depending on evolutionary parameters. The minigame analysis accurately describes the evolution of the two conditional frequencies in this account.

By explicitly considering the relationship between the full ultimatum game, with its many moving parts and seemingly complicated dynamics, and a simplified two-dimensional minigame analysis, a reasonable understanding is gained of the interplay of critical forces that lead to the stabilization of imperfect equilibria in the full game. A better understanding is also gained of the factors affecting the difficulty (i.e. the minimum required ratio of responder to proposer mutation rates) of stabilizing any particular imperfect equilibrium. For this, the graphical analysis based on selection-mutation equilibrium loci proved useful. For example, lower-offer equilibria are easier to stabilize because this has the effects of pulling down $T(e)$ and pushing up the mutation target for responders, both of which will tend to make it easier for the two loci to cross three times rather than only once.

The analysis was based specifically on the ultimatum game - would the techniques developed in this chapter, particularly the use of conditional frequency dynamics apply to other games? It seems that the essential feature of the ultimatum game that made this fruitful was that, for both players, a particular strategy i treats whole sets of opponent strategies, e.g. offers below i for responders or demands above i for proposes, alike. Many types of bargaining and other games with many strategies share such properties, so it may be worth exploring the use of conditional frequency analysis more generally. It would also be worthwhile to establish whether this type of analysis could generalize to evolutionary dynamics other than the replicator dynamics.

[23] An important caveat is that my interest was restricted to the question of the existence of stable imperfect equilibria, thus I do not claim that the minigame analysis can account for all dynamic features of the full ultimatum, particularly when the system is not near the characteristic equilibrium of the conditional game in question.

Appendix

2.A Proposer's best response

An important part of the justification for a two-dimensional analysis of the dynamics of the full UG model relies on the fact that whether a particular strategy is a best response or not depends only on a single conditional strategy frequency in the opposite population, i.e. ξ_e^P determines whether e is a best response for responders and ξ_e^R determines whether e is a best response for proposers. This appendix sets out to prove the latter proposition (i.e. for proposers).

The main theorem below therefore proves that e is a conditional best response (i.e. best response within $\mathcal{E}$) if ξ_e^R equals or exceeds a particular threshold $T(e) \equiv 1/(\$ + 1 - e)$, which means e is an overall best response for proposers if $X_{\mathcal{E}}^R \approx 1$. But the theorem depends on a a prerequisite condition regarding the distribution of responder frequencies, namely that they are nonincreasing in the demanded amount; so before the main theorem is presented, it is necessary to consider the reasonableness of the prerequisite condition, which follows by means of a pair of lemmas. The prerequisite condition is specified as,

$$x_i^R \geq x_j^R \quad \forall \;\; i < j \tag{2.7}$$

Lemma 1. *If condition (2.7) holds, it will be maintained forever by the full model's dynamics (**??**).*

Proof. Given any two demands, the expected payoff to a responder making the higher demand can be lower or equal, but never higher than the expected payoff to the lower demand, because a higher demand could potentially lead to rejection and a payoff of zero:

$$\pi^R(i, \mathcal{S}) \geq \pi^R(j, \mathcal{S}) \quad \forall \;\; i < j \tag{2.8}$$

From (2.1),

$$\frac{d(x_i - x_j)}{dt} = \Delta \left[x_i(\pi(i, \mathcal{S}) - \pi(\mathcal{S}, \mathcal{S})) - x_j(\pi(j, \mathcal{S}) - \pi(\mathcal{S}, \mathcal{S}))\right] + \delta(x_j - x_i). \tag{2.9}$$

For any i < j, condition (2.7) requires that either $x_i = x_j$ or $x_i > x_j$. In the case that $x_i = x_j$, it can be verified that the first term (selection) of the above equation will be nonnegative given (2.8), and the second term (mutation) will be zero, so $d(x_i - x_j)/dt \geq 0$ and (2.7) will be maintained. In the case that $x_i > x_j$, it is possible that $d(x_i - x_j)/dt < 0$, in which case dx_j/dt will exceed dx_i/dt. However, given continuous dynamics, this cannot result in a state where $x_i < x_j$, because for this to occur, the system would have to pass through a state where $x_i = x_j$, at which point we would again have $d(x_i - x_j)/dt \geq 0$. □

It is clear that Lemma 1 always applies when the starting distribution of responder frequencies is uniform ($x_i = x_j$), so it is possible to justify condition (2.7) by assuming a uniform starting distribution. However, this is unnecessarily restrictive, and the condition may have relevance regardless of the starting distribution. Given our interest in the existence of stable imperfect equilibria of the model, I prefer an alternative justification, expressed in the following lemma.

Lemma 2. *Condition (2.7) holds at any rest point of the full model's dynamics (2.1).*

Proof. Assume condition (2.7) does not hold, so that there is some i, j in $\mathcal{S}$ such that $i < j$ and $x_i^R < x_j^R$. Using a similar procedure as was used to derive (2.3), we obtain,

$$\frac{d\left[\dfrac{x_i}{x_i + x_j}\right]}{dt} = \Delta \frac{x_i}{x_i + x_j}\left(\pi(i, \mathcal{S}) - \pi(\{i, j\}, \mathcal{S})\right) + \delta \frac{2}{|\mathcal{S}|} \frac{1}{x_i + x_j}\left(\frac{1}{2} - \frac{x_i}{x_i + x_j}\right), \tag{2.10}$$

where $\pi(\{i, j\}, \mathcal{S})$ is the weighted average payoff to i and j. The first term (selection) will be nonnegative because (2.8) implies $\pi(i, \mathcal{S}) \geq \pi(\{i, j\}, \mathcal{S})$, while the second term (mutation) will be strictly positive if $x_i < x_j$ (assuming $\delta > 0$). Therefore, if condition (2.7) does not hold, x_i^R's share of $x_i^R + x_j^R$ will be strictly increasing over time, which implies that the system cannot be at rest unless condition (2.7) is true. □

This effectively shows that any deviations from condition (2.7) can only arise from selecting a starting position in which it does not apply, and that such deviation will be temporary.[24] Nevertheless, such a temporary deviation could

[24]The tendency to restore the condition is actually somewhat understated by these lemmas, since in practice proposer mutation means the inequality in (2.8) will be strict, so that if the condition does not hold, both selection and mutation's effects in (2.10) will be strictly positive, and $x_i^R = x_j^R \implies d(x_i^R - x_j^R)/dt > 0$, suggesting that condition (2.7) must apply not only "at" but also "near" rest points. In simulations, the condition tends to be established quickly when it does not hold and it is maintained forever when it holds.

mean that the theorem below is temporarily inapplicable, which could conceivably cause the system to skip out of the general vicinity of some dynamic equilibrium towards another, or otherwise affect the system's trajectory, *but it does not affect the existence of any dynamic equilibrium*.

Theorem 1. *Assume (2.7) is true. If and only if $\xi_e^R \geq T(e) \equiv 1/(\$ + 1 - e)$, then e is a conditional best response for proposers within $\mathcal{E}$, meaning there is no alternative in $\mathcal{E}$ that would give proposers a higher expected payoff.*

Proof. Let q_i be the conditional frequency of responders (i.e. the proportion of responders following strategies in $\mathcal{E}$) that would reject an offer of i or smaller. In general, $q_i = q_{i+1} + \xi_{i+1}^R$, and in particular, $q_{e-1} = \xi_e^R$ since $q_e = 0$. For e to be a conditional best response for proposers, it is necessary for e to get the same or better payoff as any other strategy in the subset $\mathcal{E}$:

$$\begin{aligned} \pi^P(e, \mathcal{S}) &\geq \pi^P(i, \mathcal{S}) \\ X_{\mathcal{E}}^R[\$ - e] &\geq X_{\mathcal{E}}^R[(1 - q_i)(\$ - i)] \\ e - i &\leq q_i(\$ - i) \quad \forall \ i < e \end{aligned} \tag{2.11}$$

where $G(i) \equiv e - i$ can be interpreted as the gain in offering i rather than e if i is accepted and $L(i) \equiv q_i(\$ - i)$ is the expected loss due to rejection when i is offered. Notice that $X_{\mathcal{E}}^R$ cancels out, because responders that demand higher than e will reject any offer in $\mathcal{E}$ and therefore reduce payoffs to all proposers playing strategies in $\mathcal{E}$ equiproportionally. A necessary condition for (2.11) to be satisfied is that it be satisfied for $i = e - 1$, which requires that

$$\xi_e^R \geq T(e) \tag{2.12}$$

However, (2.12) is in fact a *sufficient* condition for (2.11) to be satisfied for all $i < e$. To show this, it will be sufficient to demonstrate that $L(i)$ increases faster than $G(i)$ as i is decreased, so that (2.11) will continue to hold for all $i < e - 1$ if it holds for $i = e - 1$:

$$\begin{aligned} G(i-1) - G(i) &= 1 \\ L(i-1) - L(i) &= q_{i-1}(\$ + 1 - i) - q_i(\$ - i) \\ &= q_i(\$ + 1 - i) + \xi_i^R(\$ + 1 - i) - q_i(\$ - i) \\ &= q_i + \xi_i^R(\$ + 1 - i) \end{aligned}$$

From (2.12) we have $\xi_e^R(\$ + 1 - e) \geq 1$, but since $e > i$ and $x_i^R \geq x_e^R$ according to (2.7),

$$1 \leq \xi_e^R(\$ + 1 - e) < \xi_i^R(\$ + 1 - i) < q_i + \xi_i^R(\$ + 1 - i)$$

Therefore (2.12) implies $L(i-1) - L(i) > G(i-1) - G(i)$ for all $i < e$ and therefore (2.12) implies (2.11) as well.[25] □

[25] Condition (2.7) is not strictly necessary for this conclusion. The conclusion would still hold if $(\$ + 1 - i)/(\$ + 1 - e) > \xi_e^R/\xi_i^R \ \forall \ i < e$, or even failing that, if q_i is large enough.

Using the theorem, if $\xi_e^R \geq T(e)$ and it is also given that $X_{\mathcal{E}}^R \approx 1$, then e must be an overall best response for proposers.

2.B Directed Mutation

Throughout the chapter, I have assumed that mutation was uniform in the sense that an individual changing her strategy due to mutation, either due to "trying something new" or "making a mistake", selects her new strategy so that the probability that each particular new strategy is chosen is independent of her old strategy and also independent of the new strategy. In different ways, Harms (1997), Zhang (2013) and Akdeniz and Van Veelen (2023) show that the central GBS result of stable imperfect equilibria can be sensitive to the precise way in which mutation is modelled. Suppose, following GBS (p. 68), that the probability of a particular new strategy j being chosen is ψ_j. They report the following simulation results: a) putting relatively more weight on offers and responses near zero leads to outcomes in which responders get less than 20% of the pie (i.e. less than for proportional mutation), b) putting more weight on "somewhat higher" offers leads to outcomes in which the responders get more than 20% of the pie; however, c) changing the values of ψ_j for proposers and responders that are attached to "relatively high offers has virtually no effect on the outcome", for example putting high weight on ψ_{20} with the remaining values of ψ_j remaining equal to one another had "almost no impact" on the results. Their explanation is that such high-amount strategies earn such a low payoff that insufficient probability accumulates on such offers to affect the final results.

The conditional frequency model in this chapter can help shed light on these results by showing that it matters whether weight is shifted within $\mathcal{E}$ or between $\mathcal{E}$ and the rest of $\mathcal{S}$. If we replace $\frac{1}{|\mathcal{S}|}$ in (2.1) with ψ_i, then (2.3) becomes,

$$\frac{d\xi_i}{dt} = \Delta\xi_i(\pi(i,\mathcal{S}) - \pi(\mathcal{E},\mathcal{S})) + \delta\psi_{\mathcal{E}}\frac{1}{X_{\mathcal{E}}}\left(\frac{\psi_i}{\psi_{\mathcal{E}}} - \xi_i\right) \tag{2.13}$$

where $\psi_{\mathcal{E}} = \sum_{i\in\mathcal{E}} \psi_i$. This shows that changing the share of total mutation going to strategies in $\mathcal{E}$ will modify the effective rate of mutation on the conditional frequency, which will tend to make the locus of selection-mutation equilibrium curve more smoothly. But changing the relative share within $\mathcal{E}$ going to e will instead change the mutation target, which will have the effect of vertically stretching the locus of selection-mutation equilibrium in the case of responders (refer to figure 2.2).

Consider what would happen in the case of a "norm" strategy that would be picked with a greater probability than others when an individual mutates. Say the norm is outside the current equilibrium, e.g. an offer of 20 when $e < 10$. Assign a value of $\frac{1}{2} + \frac{1}{80}$ to ψ_{20} and $\frac{1}{80}$ to all other ψ_j's. Compared to uniform mutation, this does not change the mutation target $\psi_e/\psi_{\mathcal{E}}$, but it does lower the effective rate of mutation. If this norm (a) only applied to proposers, we would expect the equilibria $e < 10$ to be more easily stabilized, while if (b) the norm applied only to responders, we would expect these equilibria to be less easily stabilized, and if (c) the norm applied to both proposers and responders, we would not expect any change as the effective ratio of proposer and responder mutation rates would be unchanged. Simulations of the model with these mutation probabilities implemented, with $\delta^P = \delta^R = 0.1$ and uniform starting frequencies as before, come to a final rest at (a) an offer of 9, (b) an offer of 6 and (c) an offer of 7, respectively, confirming the predictions (recall that the original model stabilizes at an offer of 7 for these parameter values). To interpret these results, it is best to regard the "norm" of 20 as no more than a distraction that makes the mutation that really matters, i.e. of relative frequencies within $\mathcal{E}$, effectively slower, but since relative mutation rates matter, distracting one population more than the other can change the result. This analysis suggests that the role of "bias" in mutation to stabilize high-offer equilibria is not quite as simple as increasing the average offer and demand above the lowest possible values, as suggested by Akdeniz and Van Veelen (2023, p. 579).

Next, we can adjust $\psi_i/\psi_{\mathcal{E}}$ while leaving $\psi_{\mathcal{E}}$ unchanged. For $e = 8$, let $\psi_8 = \frac{1}{10}$, let $\psi_i = \frac{1}{70}$ for $i < 8$ and let $\psi_i = \frac{1}{40}$ for $i > 8$. When applied to either proposers or responders, this stabilizes the equilibrium with modal offer of 8 (which is not stable with uniform mutation) easily in a simulation. For responders, this vertically stretches the selection-mutation equilibrium locus as explained previously. For proposers, raising the mutation target will effectively weaken mutation, given that $\xi_e^P \approx 1$ in an equilibrium characterized by offers of e, so the target comes closer to the actual value.

Chapter 3

Stochastic Learning and Emulation in the Ultimatum Game

3.1 Introduction

Evolutionary game theory can be understood as the application of insights from the study of the dynamic processes of biological reproduction and adaptation to human behaviour. When humans do not have perfect information or understanding of the decisions they face, their behaviour is thought to derive, at least in part, from habit and rule-of-thumb (Mailath, 1998, p. 1349), thus they exhibit myopia and inertia while slowly learning to behave more optimally. A difficult question is whether the evolutionary models that have been employed are appropriate to model human decision-making processes. Models are always imperfect and simplified representations, but the link between evolution and decision-making can be somewhat loose and there is a natural tension between evolution, which applies to populations in aggregate and the standard methodology of economics which is focused on individual decisions in a particular context. There have therefore been initiatives to build more explicit "microfoundations" for aggregate-level evolutionary models, which this chapter aims to contribute to.

The need for a concrete and defensible interpretation of a model at individual actor level is especially important when the outcomes are interesting, counterintuitive or sensitive to modelling parameters or variation in specification. A prime example is the ultimatum game, where GBS find that, in a model based on the replicator dynamics with noise/mutation, the system can settle at evolutionary stable states far from the classical subgame-perfect Nash equilibrium (SPNE).[1] Their results show that responders would irrationally reject

[1]The same basic result is obtained in Binmore and Samuelson (1994) under a different version of the replicator dynamics, the "discrete" or "adjusted" replicator dynamics introduced by Maynard Smith (1982). I will discuss this version along the way, but my main concern will be

positive offers from proposers if they are too low, and proposers consequently make relatively fair offers (which are then accepted). Evolution fails to push the responders towards fully rational behaviour, which would be to accept all positive offers. While the general result of stable non-subgame-perfect equilibria holds for a range of model parameters, the stability of a particular equilibrium may require responders to be more "noisy" than proposers, and a particular equilibrium may be stable under the standard continuous-form replicator dynamics but not under an alternative "discrete" version of the replicator dynamics. A satisfactory evaluation of the significance of these factors has not yet been provided in the literature.

What is the significance of the GBS result? Any answer to this question, whether based on empirical validation or other criteria, must start from a full understanding of what the theory implies, or can imply, at the level of the individual actor. The replicator dynamics is a mathematical representation of Darwin's theory of natural selection where game payoffs are equated directly with the number of surviving offspring, i.e. biological *fitness*. The replicator dynamics does not directly purport to describe learning behaviour of active decision makers (Sandholm, 2020, p. 574), but, over the years, evolutionary game theorists have discovered that the replicator dynamics equations can follow mathematically from a variety of alternative models more suited to economists' focus on behaviour as directed, individual choices. Indeed, GBS provide such a model in their 1995 paper (p. 85), explicitly motivated by the need to justify interest in the replicator dynamics. Their model assumes that agents change their strategies to a random alternative when payoffs fail to meet aspiration levels, which is drawn from a uniform random distribution.

This interpretation is well-suited to studying interactive learning by boundedly rational agents, but I will argue that it is less appropriate if one wishes to explore the broader historical context of cultural evolution in which entrenched behavioural norms and social preferences could have developed. In this chapter, I match the aggregate deterministic GBS model with a simple stochastic individual-level model, similar to a model in Weibull (1995, p. 158), which is based on imitation of social peers. Here, agents have a degree of sensitivity to payoffs of others when deciding who to imitate, unlike the random-aspiration-level model where this decision is made blindly. In my model, agents choose to imitate at random times according to a Poisson process at a rate common to all agents in the population. When it is time to choose a new strategy, agents still choose random individuals from the population to imitate, but the choice is now weighted by expected payoffs of the target strategies, so better strategies are more likely to be imitated. This is justified by the notion that agents are motivated to choose strategies with higher payoffs, but are constrained by noisy information about payoffs, and possibly also by social constraints.

specifically with the result as reported in the GBS paper.

The model can be specialized either to the standard ("continuous") or the adjusted (also called "Maynard Smith" or "discrete") versions of the replicator dynamics by adjusting time step and revision rate parameters. An important result shown by my analysis is that the two versions of the replicator dynamics have quite different implications for individual selection and mutation rates – in the continuous version, selection rates are proportional to mean payoffs for the population while for the discrete version the selection rate is completely exogenous. This result recalls Maynard Smith (1982, p. 183), who, upon introducing the discrete version of the replicator dynamics, immediately shows that the continuous version can be obtained from it simply by multiplying by mean payoff. The implication in the UG based on the standard replicator dynamics is that individuals in the population of proposers "naturally" learn at a faster rate than individuals in the population of responders, even with the same selection rate parameters in both equations, due to the differences in mean payoff between the two populations.

This challenges (as far as I am aware, for the first time) a justification made by GBS for an assumption of a larger selection rate parameter for proposers relative to responders, namely that proposers have potentially more to gain from switching than responders. In light of the interpretation provided by my imitation model, such an asymmetric treatment of the selection rates in the two populations appear superfluous – learning for proposers is in fact now doubly boosted relative to responders. The asymmetric selection rate parameters are not essential for obtaining stable non-subgame-perfect evolutionary equilibria, but they do sharpen the results and stabilize certain equilibria (those with relatively more equal divisions between proposers and responders) than would otherwise be the case. The analysis furthermore provides a clear understanding for why the GBS result does not emerge as strongly in the adjusted / discrete version of the replicator dynamics, which Binmore and Samuelson (1994) apply to the ultimatum game in an earlier paper.

The stochastic individual learning model leads to a further important implication that challenges the relevance of the GBS result. Deterministic aggregate dynamics models are invariably based on the idea of *large* populations. But if real populations are not infinite, stochastic effects at individual level can have significant effects at aggregate level, instead of cancelling out as they would in a truly infinite population. In this case, aggregate model cannot be regarded as an adequate representation of the situation envisaged as there may be stochastic trends and reversals that give very different results compared to the mean aggregate dynamics, especially where stability of a fragile equilibrium is concerned. Most researchers apply the idea of a law of large numbers informally or implicitly, while Boylan (1995) and Sandholm (2010, chapter 10) prove that the procedure is indeed mathematically valid in the limit for large enough finite populations under certain conditions. However, it remains difficult to understand the practical significance of the assumption in a specific application like the ultimatum game. I therefore conduct an agent-based simulation of the imitation learning model which shows concretely that, in the case of a rather sensitive result like that of GBS, population sizes well in excess of a million agents per population are required before the results reliably resemble the predictions of the deterministic model. This suggests that for reasonable population sizes applicable to an interpretation of cultural evolution, the GBS result will not hold – instead, the model will lead to the SPNE or a result close to it.

The remainder of this chapter is organized as follows. Sections 3.2 and 3.3 review the GBS model and result. Section 3.4 develops the stochastic-individual-level learning/imitation model. Section 3.5 considers the question of selection rates and interpretational differences between the continuous and discrete versions of the replicator dynamics. Section 3.6 details the agent-based simulation of the imitation model and presents results. The conclusion offers an assessment that the GBS result remains theoretically important, but that it is unlikely to be directly discernible in experimental data.

3.2 The Gale, Binmore and Samuelson result

In the ultimatum game, there are two players, the proposer and the responder. The proposer is asked to divide a certain fixed amount of money between herself and the responder. The responder can then choose to either accept or reject the proposal. If the responder accepts, both participants receive the proposed amounts, but if the responder rejects, neither party receives anything. The game ends after one such interaction. GBS model a specific scenario where the two players must split an amount of 40, with only integer offers in $\mathcal{S} = [1, 40] = \{1, 2, ..., 40\}$ allowed (higher offers representing a higher amount offered to the responder). The proposer's strategy is simply the offer amount, and the responder's strategy is specified as the minimum acceptable offer (henceforth called "demand").[2] This results in forty Nash equilibria in pure strategies,[3] of which only one is subgame perfect, namely $(1, 1)$, i.e. the lowest possible offer and demands.

Following GBS, I will ignore all stochastic effects related to player matching, so players are effectively "playing the field" throughout this chapter. The payoff for each strategy at a point in time, π_i, is thus defined as the expected payoff a player using strategy i would obtain if paired with a randomly selected player from the opposite population, given the current strategy frequency dis-

[2] This assumes that a responder who accepts a certain offer will also accept all higher offers and a responder who rejects a certain offer will also reject all lower offers.

[3] Each combination of offer and demand of the same amount represents a Nash equilibrium, because given that i is the minimum acceptable offer, i is the optimal proposal and given that i is proposed, the responder can do no better than accepting a minimum of i.

tribution.[4]

A standard continuous replicator equation is applied to each of the two populations: the rate of change over time t in the fraction of the population adopting strategy i is equal to $dx_i/dt = x_i(\pi_i - \overline{\pi})$ where x_i is the current fraction of the population adopting strategy i and $\overline{\pi} = \sum_{i \in \mathcal{S}} x_i \pi_i$ is the current average expected payoff in the population. It can be seen that strategies that earn an expected payoff higher than the average payoff will tend to increase their share in the population and strategies earning expected payoffs below the average will be adopted by a decreasing fraction of the population.

Given this set up, every one of the forty pure strategy Nash equilibria is a rest point of the dynamic system, and GBS show that many of them are also asymptotically stable local attractors. A simulation of this system with a uniform initial distribution of strategies converges to a state where all proposers make an offer of 9, and all responders accept this offer (some also accept lower offers). But, they argue, this in itself does not make a strong argument against the subgame-perfect prediction, because in such a state, evolutionary pressure against weakly dominated strategies, i.e. responders who would reject low offers, has been "artificially" removed (GBS, p. 61). They therefore introduce noise, or mutation, into the dynamics, to "test the rationality" of weakly dominated strategies and prevent the system from getting stuck simply because some strategies have been effectively removed from the population.[5] Mutation entails a small fraction of agents, δ, erroneously adopting the wrong strategy (or a strategy adapted to a different game) instead of the one mandated by the simple replicator dynamics. The authors choose to let each agent committing an error choose from a uniform distribution of strategies in $\mathcal{S} = [1, 40]$ so that the probability that a particular erroneous strategy is chosen is $\frac{1}{|\mathcal{S}|} = \frac{1}{40}$. The "noisy replicator dynamic" equations making up their model are therefore,[6]

$$\frac{dx_i}{dt} = \Delta x_i(\pi_i - \overline{\pi}) + \delta\left(\frac{1}{40} - x_i\right), \tag{3.1}$$

[4]To save space, I do not show population indicators and write only one equation instead of two whenever possible.

[5]This would also in principle remove the possibility of the system reaching a state where some strategies that would at this point be a best response have become extinct through low payoffs in the past. The idea of low-probability "mistakes" undoing some Nash equilibria was also the original motivation for Selten's (1975) trembling hand perfect equilibrium: given that every choice in a game is made with some positive probability if mistakes can occur, agents must choose rational behaviour also at off-equilibrium decision nodes (Van Damme, 1991, p. 10–11). See also Young (1996, p. 109–111) who explains how allowing for errors in actions can cause a dynamic system not to have any absorbing states. In GBS, this does not occur due to the deterministic nature of the added noise, representing expected aggregate effects, in which combinations of highly unlikely errors is not possible. This issue is addressed in section 3.6.

[6]Proposer and responder (sub-)populations are modelled separately, each population with its own δ (δ^P and δ^R below), giving a total of eighty differential equations for the eighty frequencies in the model.

where Δ is the selection rate, equal to $(1-\delta)$ in the GBS model.[7]

Simulation of these equations[8] then show that, starting from uniform initial distributions of strategies, the system converges to some outcome that depends on the mutation parameter δ, also called the mutation rate, for each population. If the mutation rate for the responders is sufficiently small relative to the mutation rate for the proposers, the subgame-perfect equilibrium appears. But if responders are at least as noisy as proposers, the system *never* reaches the subgame-perfect outcome. Instead, a state is reached where almost all proposers (apart from a small fraction due to noise/mutation) offer some modal amount, between 7 and 9, depending on parameter values. Responders follow a mix of strategies but approximately all (apart from a small fraction due to noise/mutations) demands are at or below the modal offer so approximately all offers are accepted.

GBS (p. 75) make use of a simplified minigame analysis to explain why their result is obtained. The main part of the explanation is that when proposers learn quickly to make acceptable offers, due to strong incentives, responders have little incentive to learn to accept low offers, because they see very few low offers so suffer only a slight negative consequence from choosing strategies that would reject low offers. The role of mutation is not to hammer the system towards a non-subgame-perfect equilibrium,[9] but to apply a subtle bias to a knife-edge balance scenario. In fact, stable non-subgame-perfect equilibria appear even when mutation rates are vanishingly small in the GBS model.

[7] Notice that mutation always tends to bring the frequency of a strategy towards $1/40$, its "equitable share" given a uniform mutation distribution. Note also that while setting $\Delta = (1-\delta)$ seemingly provides a neat interpretation, that a fraction δ of agents misreads the game and the remaining $(1-\delta)$ selects according to the replicator equation, there is no necessary link between the selection and mutation rates. They can also be regarded as independent continuous processes: over a small time interval dt, a fraction $dt\Delta$ changes according to selection dynamics and an independent fraction $dt\delta$ changes according to mutation dynamics. Setting Δ to some other value, e.g. 1, would not mean that the value of $\sum_i x_i$ would deviate from unity over time. For a more general replicator-mutator equation explicitly linking replication and mutation see Page and Nowak (2002, p. 94).

[8] To simulate the model, GBS discretize the continuous noisy replicator dynamics as follows: $x_i(t+\tau) - x_i(t) = \tau\left[\Delta x_i(\pi(i,\mathcal{S}) - \pi(\mathcal{S},\mathcal{S})) + \delta(\frac{1}{40} - x_i)\right]$, where $\tau = \frac{1}{100}$. Experimentation with smaller values confirmed that this provides a close approximation to the differential equation.

[9] By contrast, Rand et al. (2013) obtain relatively equal money divisions from an evolutionary model that depends on a large mutation rate and a small selection rate. This seems uninteresting to economists working from the premise that human behaviour is generally goal-oriented.

3.3 Why the GBS result is interesting

The GBS result is interesting, firstly, because it is counterintuitive from a theoretical perspective. One would expect every rest point in the noisy evolutionary game that is not subgame perfect to be unstable. This is because the presence of all possible offers by proposers, even at low frequencies, should offer an incentive for responders with relatively high demands to adopt lower demands. The GBS result is important because it presents a challenge to the relevance of one of the most important solution concepts in classical game theory, the subgame-perfect Nash equilibrium. Underlying this is the broad notion that evolutionary models can serve as justification for the central assumption of rational behaviour in economic models (Mailath, 1998).

A second reason the GBS result is interesting is because it seems to provide an explanation for the voluminous experimental data showing that laboratory subjects do not play the SPNE (Camerer and Thaler, 1995, p. 210). Experimental researchers have not always been enthusiastic about evolutionary or learning explanations for UG results.[10] There are two difficulties. The first is that there is an apparent conflict between the bounded rationality implicit in evolutionary models and the dominant approach in experimental economics, which is based on the assumption of orthodox utility maximisation, allowing laboratory results to be interpreted as reflecting social, or other-regarding, preferences.[11] The modern view, supported by numerous empirical studies, is that responders in the UG experience positive utility when they punish proposers who violate fairness norms (Fehr and Schmidt, 2006, p. 630). A common argument is that the assumption of rational behaviour is justified because the the situation facing responders in the UG is just too simple for them to reject positive offers merely due to lack of understanding the game (e.g. Camerer and Thaler, 1995, p. 210, Fehr and Schmidt, 2006, p. 617,628). Secondly, there are robust features of the empirical data that cannot easily be explained by models populated by boundedly-rational, adaptive agents who care only about income. A good example is that responders show much higher acceptance rates when they are informed a low offer they received was generated by a computer (e.g. Sanfey et al., 2003).

Nevertheless, a smaller number of experimental researchers have found some evidence of learning taking place over successive rounds of play by experimental subjects (e.g. Slonim and Roth, 1998, List and Cherry, 2000, Cooper and Dutcher, 2011). Proposer behaviour tends to change over time towards income-maximizing offers (Cooper and Dutcher, 2011, p. 540). Responders

[10]The GBS model is often discussed in conjunction with Roth and Erev (1995) who present a model of reinforcement learning that yields generally similar results and implications for experimental game theory. Given my research's aim, the discussion will be restricted to the specific relevance of the GBS model.

[11]Cooper and Kagel (2016) provide a literature review.

also adapt, but more slowly, and generally in the direction of higher acceptance rates (e.g. List and Cherry, 2000, Winter and Zamir, 2005). The mere fact that proposers learn faster than responders is a key feature of the GBS model. But Cooper and Dutcher (2011) raise a problem, namely that human responders appear to learn to reject very low offers with higher frequency in later rounds, which could not happen in a learning model if rejections can only reduce utility.[12] Cooper and Kagel (2016, p. 240) suggest there is a need for models that combine elements of learning and other-regarding preferences.

A third reason the GBS result is interesting is that it could give some insights about the cultural and/or biological evolutionary processes that may have given rise, over longer time-spans, to behavioural traits, social norms and even preferences relevant to real-life ultimatum and related bargaining and other social interactions. We want to know more than simply that people have other-regarding preferences – we want to understand why they do (Güth, 1995, p. 342).[13] GBS (p. 70) argue that the social norms triggered in the short term by laboratory experiments can be presumed to have evolved in "real-life bargaining situations that are superficially similar to the ultimatum game in some respects" and that "we must therefore examine long-run behavior in these external situations for the origin of the norms that guide short-run behavior in laboratory experiments on the Ultimatum Game". But such real-life bargaining situations would hardly ever be free of future consequences – repetition, reputation and social status all matter in everyday life, so the norms are unlikely to be adapted to the ultimatum game as played in a typical experimental setting. The work of Henrich et al. (2005), who studied UG behaviour among subjects drawn from 15 different small-scale societies with richly diverse social practices between them, gives strong support to the notion that norms brought into the laboratory developed as a result of experience in real-life social interactions applicable to the different societies. These authors (p. 812) also point to longer-term genetic and cultural evolutionary processes as the likely ultimate source

[12] The effect is quite small and statistical significance is only obtained through the use of combined data from numerous studies.

[13] In an application of the *indirect evolutionary approach*, according to which it is preferences that evolve rather than behaviour, Huck and Oechssler (1999) found that preferences in the UG can evolve that give rise to behaviour similar to those observed in experimental data. However, these results rely on the specific assumptions such as that the population is small and that agents can act in both roles, so that the negative payoff effect of rejecting a low offer can be outweighed by the even more negative payoff effect the rejection has on other strategies, thus turning a payoff loss into a *relative* evolutionary advantage for the rejecting strategy. Debove et al. (2016, p. 248) refer to this effect as 'spite". The indirect evolutionary approach is compelling because it may provide more direct explanations for the evolution of social preferences, which connects it more closely to experimental work, and can be regarded as complementary to the direct evolutionary approach, which aims to explain the evolution of behaviour in the absence of the rationality assumption.

of the preferences and behaviour displayed by their experimental subjects.[14]

Research aimed at discovering and validating these ultimate explanations for behaviour in the laboratory, and in real life, is a large and challenging project. Here, it is not possible to apply the kind of strict controls available in laboratories, and, as we have learned, norms and behaviour are not neatly tied to specific games but come from exposure to a fluid and rich mixture of complicated real-life games, and are driven by a mixture of evolutionary processes such as learning, imitation and even genetic evolution. As always, simple and tractable models are likely to be helpful, and credible stories for how behaviour could have developed, even in fictitious societies where the ultimatum game is the only interaction, and reputation effects are somehow irrelevant, can form a part of the answers we seek.[15] But this will be useful only if we have a proper understanding of what stories our model can tell us.

Unfortunately, evolutionary theorists have too often been vague about the interpretation of their theories. Debove et al. (2016, p. 251), surveying various models that seek to identify mechanisms according to which fair behaviour can evolve among ultimatum game players, suggest that this might have to do with the fact that the same evolutionary equations can often be applicable at different levels – learning, cultural and genetic evolution. However, they point out quite reasonably, this vagueness on interpretation may have negatively affected our understanding of the origins of human fairness. To this end, I aim to provide in this chapter a more concrete interpretation of the GBS model applicable to a society in which cultural evolution takes place.

3.4 A model of imitative learning

The setting for this model is a large well-mixed population of individuals who play the ultimatum game every day with random partners. They may have other kinds of interactions, but these do not affect their behaviour when playing the ultimatum game. Player behaviour is myopic and non-strategic; players do not believe that their actions can affect the behaviour of others, or that others adapt their behaviour (Mailath, 1998, p. 1355).[16] This may be due to the large size of

[14] See also Güth and Napel (2006, p. 1038).

[15] It is good to recognize explicitly the tension between acknowledgement that the UG is uncommon in historical real life, but then to proceed to model a world where the UG is the only game being played, hoping to understand evolved behaviour in the real world. Like GBS, Skyrms (2014, p. 29) is explicit on both these points. The question of how we can learn about the real world using unrealistic, very simple models, is not unfamiliar to economists. The thing to avoid is to relate results from the model back to the real world in roughshod fashion; for example when we find "anomalous" behaviour in an evolutionary model, we should be careful not to suggest that we have suddenly "explained" seemingly similar "anomalous" behaviour in the real world with a very different context, without further ado.

[16] Mailath adds, "and they do not look for patterns in historical data".

the population and bounded rationality (Friedman, 1998, p. 16). It may also be because behaviour is constrained by social norms and/or emotional responses elicited by social interactions.

The payoffs players receive from the ultimatum interactions can be thought of as a generalized notion of "resources" that are necessary for survival, security and happiness. We may call these resources "dollars" for convenience. The payoffs are not exactly "utility" values, because "utility" in the modern sense is what would be maximized by the agents' behaviour on the assumption that they have consistent preferences and behave fully rationally in that pursuit.[17] These agents are not really endowed with preferences – each agent at simply holds a "strategy", and rationality is not assumed either.[18]

The critical assumption about payoffs is that players will tend to replace strategies that obtain relatively low payoffs with strategies that obtain relatively high payoffs over time. This may be due to a mixture of learning (deliberate improvement) or gradual change in behavioural norms to accommodate more effective strategies. In this model (a version of Model 2 in Weibull, 1995, p. 158), I simply assume that all agents have a small exogenous probability of updating their strategies in a short time interval, and the new strategy is selected by imitating a random member of the population, weighted by their current expected payoffs. Since the target for imitation is an agent rather than a strategy, the probability of a strategy being chosen for imitation will naturally also be weighted by the current frequency of that strategy in the population. The result would be approximately the same if an agents who have decided to update consider sufficiently representative samples instead of the whole population and choose their new strategies with probability weights proportional to payoffs observed in the samples.[19] Note that agents can sometimes choose strategies worse than their current ones.

Assume there are N proposers and N responders, with the number of proposers and responders following each strategy i equal to z_i^P and z_i^R respectively

---

[17] In the indirect evolutionary approach, it is likewise the case that "evolutionary success" and "utility" are distinct concepts (Konigstein and Muller, 2000, p. 236). In this literature, utility is concrete and utility function parameters are subject to evolution, but evolutionary fitness depends on some more objective measure, e.g. material resources.

[18] It may be that evolution eventually results in behaviour that appears rational and that can reasonably be fitted to a utility function, but there is no guarantee that the utility function will be neatly tied to the game's dollar payoffs. It may be argued that the agents are unreasonably unsophisticated, but bear in mind that there may be unmodelled complications, e.g. complicated strategic structure, considerations involving social relations and behavioural norms, etc. See Mailath (1998, p. 1356).

[19] Simulation results (not shown) suggest that agents taking small samples will have a small bias in favour of strategies with higher frequencies in the population, as low-frequency strategies will often not be represented in a small sample, in which case their payoffs do not come into play. This effectively blunts the force of selection, but the effect appears to be relatively minor for reasonable sample sizes (e.g. 10) so it will not be pursued in detail in this chapter.

(population indicators will again be suppressed below). As this is a cultural learning/imitation model, and not a biological model (e.g. Taylor and Jonker, 1978), N remains constant and all changes in frequencies are due to individuals changing their strategies. Time proceeds in periods of length τ. Time can be either discrete, with $\tau = 1$, or continuous, with $\tau \to 0$.

In each period, or at each point in time, each agent interacts with other agents and gets assigned the expected payoff against the current strategy distribution of the opposite population. Next, agents enter a selection phase. Each agent has a probability of changing its strategy within the period due to selection equal to $\tau\Delta s$, where s is a scaling factor (discussed below). Notice that with $\tau \to 0$, this becomes a Poisson process with agents changing their strategies at a rate of Δs. Agents changing their strategies select a new strategy randomly, with the probability weight of strategy i being selected equal to

$$\frac{z_i}{N}\frac{\pi_i}{\overline{\pi}}$$

Following the selection phase, a mutation phase, reflecting low-frequency mistakes, innovation or other unmodelled shocks at individual-level, completes the round. Each agent's strategy is changed due to mutation with probability $\tau\delta$ during the period. Assume the new strategy is selected randomly from a uniform distribution. Putting selection and mutation together, the expected change in the number of agents following i during the current period is,

$$E(z_i(t+\tau) - z_i(t)) = N\tau\Delta s\frac{z_i(t)}{N}\frac{\pi_i}{\overline{\pi}} - z_i(t)\tau\Delta s + N\tau\delta\frac{1}{N} - z_i(t)\tau\delta$$

This can be written as,

$$E\left(x_i(t+\tau) - x_i(t)\right) = \tau\left[\Delta x_i\left(s\frac{\pi_i}{\overline{\pi}} - s\right) + \delta\left(\frac{1}{N} - x_i\right)\right] \tag{3.2}$$

where $x_i = z_i/N$ and time arguments (t) have been suppressed on the RHS. Given that the population is large and has no structure, and the effects of both selection and mutation are idiosyncratic, we can appeal to the law of large numbers to argue that the system's evolution should be closely approximated by its mean dynamic (Sandholm, 2010, p. 119). In section 3.6, I treat this claim critically, using simulations to show that extremely large populations could be needed for the deterministic approximation to be a good fit for the stochastic model, especially in relation to the GBS result of stable non-SPNE equilibria. But for the moment, assume an infinite population, which simply means we drop the expectation operator,

$$x_i(t+\tau) - x_i(t) = \tau\left[\Delta x_i\left(s\frac{\pi_i}{\overline{\pi}} - s\right) + \delta\left(\frac{1}{N} - x_i\right)\right] \tag{3.3}$$

The model can now be specialized by letting $\tau = 1$ and $s = 1$, which gives,

$$x_i(t+1) - x_i(t) = \Delta x_i \left(\frac{\pi_i - \overline{\pi}}{\overline{\pi}} \right) + \delta \left(\frac{1}{N} - x_i \right) \tag{3.4}$$

Setting $\Delta = 1$ and $\delta = 0$, it can be seen this is the discrete replicator dynamics equation of Maynard Smith (1982, appendix D), also called the adjusted replicator dynamics. It is based on the evolutionary biological model of non-overlapping generations, with each organism asexually producing a number of offspring proportional to its payoff (directly interpreted as fitness) to form the successive generation, i.e. $x_i(t+1) = x_i \pi_i / \overline{\pi}$. It is also (with mutation added) the model used by Binmore and Samuelson (1994, p. 57) to study the ultimatum game's evolutionary trajectories, with somewhat similar but different results to the GBS paper (to be discussed in the next section).

Alternatively, we can let $\tau \to 0$ and $s = \overline{\pi}$ in (3.3), which gives,

$$\frac{dx_i}{dt} = \Delta x_i \left(\pi_i - \overline{\pi} \right) + \delta \left(\tfrac{1}{N} - x_i \right) \tag{3.5}$$

This is identical to (2.1), the standard continuous noisy replicator dynamics used by GBS. Thus the imitation model can provide plausible microfoundations for the GBS aggregate deterministic model, with the critical assumption required that $s = \overline{\pi}$, which makes the rate at which individuals revise their strategies equal to $\Delta\overline{\pi}$, i.e. proportional to average population payoff.

The imitation dynamics of this model has been selected to give the GBS model a concrete interpretation as a process of cultural evolution, playing out over longer time-spans (perhaps generations), where most behaviour is socially acquired, with the addition of occasional individual innovations through mutations. The feature of the model that agents due for strategy revision choose a new strategy from a representative sample from the population, in a single step, distinguishes this model from a number of individual-level strategy adjustment models described in the literature, termed "revision protocols" by Sandholm (2010). Quite a number of such models exist that lead to the continuous replicator dynamics, which provide valuable alternative interpretations for the same aggregate dynamics; a brief review follows.

It is instructive to start with the model presented in GBS (p. 85). In it, the probability of revision in a given short time period is proportional to the difference between current payoff and an aspiration level drawn randomly for each agent from a uniform distribution. If a new strategy is chosen, it is chosen from a randomly selected member of the population. Models presented in Björnerstedt and Weibull (1995, p. 162), Binmore, Samuelson, and Vaughan (1995) and Sandholm (2010, p. 154) have the same working principle, i.e. that poor strategies tend to fade away because they get replaced more rapidly than good strategies. These models place only a slight gloss on top of the biological survival of the fittest principle, and agents who have decided to revise their strategies are totally insensitive to the payoffs of target strategies, which is not a reasonable implication in the context of socially aware humans undergoing a process of cultural evolution.[20]

Models in which agents tend to imitate those who are successful are more appealing. The *proportional imitation rule* model of Björnerstedt and Weibull (1995, p. 163)[21] is relatively sophisticated. Agents receive revision opportunities according to a Poisson process at a common rate.[22] A revising agent picks another agent from the population at random, then adopts that agent's strategy only if it delivers a better payoff and then with a probability proportional to the difference in payoffs between the current and potential target strategy. Weibull (1995, p. 155) attempts to motivate this model by supposing that agents observes both payoffs with error, but this only works when the observation error for own payoff is larger than the error for the other agent's payoff, so that agents can never mistakenly imitate lower-payoff strategies, which seems incongruous. Alternatively one can assume perfect observation but some random impediment to switching, e.g. switching costs.

A simpler rule leading to the same result, as long as payoffs are positive, is given by the "Imitation of Success" rule of Sandholm (2010, p. 155). Here, as in the previous model, a single individual from the population is sampled when a revision opportunity arrives, but the probability of then switching is simply equal to the potential target strategy's payoff.[23] This interpretation appears to require that all payoffs be in the interval [0..1] so that they are valid probabilities; a rescaling of payoffs to achieve this would also scale the speed of aggregate adjustment, which may be undesirable in certain cases. Nevertheless, the standard replicator dynamics follow quite directly from this interpretation as expected inflows to a strategy's frequency is $\Sigma_{j\in S} x_j x_i \pi_i = x_i \pi_i$ and expected outflows from the strategy is $x_i \Sigma_{j\in S} x_j \pi_j = x_i \overline{\pi}$. This rule is similar to my model in that the switching behaviour is independent of own payoff, but differ in the sample size of one and the possibility of not switching at all.

A variant, due to Hofbauer and Schlag (2000, p. 529), called "Imitation of

[20] A counterargument is that they would be sensitive *indirectly*, because even though they choose new strategies from random members of the population, they would drop poor-performing ones relatively quickly afterwards.

[21] Schlag (1998) shows that this rule has optimality properties under condition of limited memory. See also Model 1 in Weibull (1995, p. 155), "Imitation via Pairwise Comparisons" in Sandholm (2010, p. 154) and Hofbauer and Schlag (2000, p. 529)

[22] Within the population, the rate at which strategies are updated is now the same across individuals. This may seem unrealistic because those with poor payoffs may want to change their strategies quickly. It is however conceivable that all agents are equally keen or able to change their strategies at a rapid pace, driven by imperfect knowledge of where their payoffs fit into the rankings, practical constraints in the rate at which they are able to switch and/or an independent inclination to conform (i.e. switch strategies).

[23] Sandholm allows a constant to be added, but this is not necessary if payoffs are positive.

Success with Repeated Sampling" by Sandholm (2010, p. 155), lets the revising agent sample another agent, imitating with probability proportional to the potential target's payoff, and repeating the procedure *until the agent has imitated* someone. A little awkwardly, this "repeat until" loop procedure does not consume any time regardless of the number of pairwise comparisons required. This does not result in the standard replicator dynamics as the other models reviewed here, but in the so-called *Maynard Smith replicator dynamics* (Maynard Smith, 1982, Appendix D) or (payoff-)adjusted replicator dynamics, which can be changed into the standard replicator dynamics by scaling the strategy revision (selection) rate by current average payoff in the population.

This is mathematically equivalent to my model, including the rescaled selection rate (obtained in my model by setting $s = \overline{\pi}$), except that I interpret the procedure as drawing a single large sample (or even the entire population) and choosing, in a single step, one individual from the sample with payoff-weighted selection probabilities, obtaining the same result. This seems a less cumbersome interpretation, avoiding the repeated sampling procedure, and a better reflection of a cultural evolutionary process, where it is reasonable that individuals can observe multiple other individuals at the same time and choose one of them to imitate. I will discuss the reasonableness of the rescaling operation below.

It is worth pointing out that, while these models seem quite different from each other, they all deliver the same mathematical end result, namely the replicator dynamics. To these, one could also add individual-learning models that likewise lead to the replicator dynamics, such as Börgers and Sarin (1997), Posch (1997) and Easley and Rustichini (1999), which may be more relevant to interactive decision scenarios, e.g. laboratory behaviour. It may seem then, in a certain sense, that all of these models must really be "the same", so the distinctions of interpretation must be unimportant. Three responses can be offered in support of the contrary view. Firstly, each model is a special case in its class, and the general analysis between classes of models would not be equivalent, so when one decides to "move beyond" the replicator dynamics, specific interpretations will matter for results. For example, in my model, relaxing the assumption of payoff proportionality in selection probabilities would lead to a different analysis than relaxing the assumption of payoff-aspiration difference proportionality in GBS's model. Secondly, different interpretations can illuminate or obscure meaning and implications for the applicability of the mathematical model to different real-world scenarios – I explore a case in point in the next section. Thirdly, while the various models may have the same *expected* trajectories, their underlying stochastic models can give different results (Sandholm, 2010, p. 498). In section 3.6, I report on stochastic simulations of my underlying model, which clearly shows that stochastic effects can matter a great deal for non-infinite populations, so the details of the microfoundations are consequential.

3.5 Selection and mutation rates in the continuous and discrete replicator dynamics

The main GBS result, asymptotically stable evolutionary equilibria that are not subgame perfect, is sensitive to parameter choices for Δ and δ, the selection and mutation rates. In their table I, GBS (p. 63) report different modal offers that their simulations settle on for different values of δ^P and δ^R, the respective mutation rates for proposers and responders. These are all Nash equilibria: proposers offer this amount and responders accept it (using a variety of strategies). GBS always set, for each population, $\Delta = 1 - \delta$, so selection rates go up as mutation rates go down and vice versa. In fact, the same results could easily be obtained by adjusting only the mutation rate, or only the selection rate (in the opposite direction); only the ratios matter for the dynamical system's trajectories.[24] The results show, for example, that if $\delta^P > \delta^R$, the subgame-perfect result, or a result close to it, is obtained, but if $\delta^P < \delta^R$, modal offers of 9 are obtained, which is far from the subgame-perfect solution. When $\delta^P = \delta^R$, the end result is modal offers of 7, still far from the subgame-perfect solution, but seemingly on the edge of the region where the result "clearly" occurs.

There is a a similar table in Binmore and Samuelson (1994, p. 58), showing results for the same UG but with dynamics determined by the discrete-time adjusted replicator dynamics described by (3.4) instead of the standard replicator dynamics. Here, it seems to be more challenging to show the non-subgame-perfect result clearly – equal δ^P and δ^R result in a final a modal offer of 2, and $\delta^P < \delta^R$ is essential to show the result more forcefully. A comparison of (3.3) and (3.4) shows clearly enough that the only difference on the RHS is that the selection term in the adjusted replicator dynamics is divided by $\overline{\pi}$, but the one in the standard replicator dynamics is not (see footnote 12 in GBS).

Since $\overline{\pi} > 1$ in this application, for both populations, this has the effect of making the force of selection relatively weaker in the adjusted dynamics, especially for proposers, due to their higher average payoff compared to responders.[25] As mentioned, weaker selection has the same effect as stronger mutation – and stronger mutation for proposers mean more accidental low offers, hence greater incentives for responders to accept low offers, hence greater incentives for proposers to make such lower offers, thus explaining the greater tendency for higher-offer equilibria to be unstable under the adjusted replicator dynamics. It was known from the beginning that the two versions of the

[24] Adjusting both selection and mutation rates by a common factor only affects the speed of adjustment.

[25] This also suggests, somewhat uncomfortably, that results should be sensitive to the choice of currency units in which payoffs are measured. $\overline{\pi} < 1$ does not necessarily mean "small" money amounts are at stake, e.g. payoffs could be measured in pots of gold. Perhaps in such cases it would be reasonable to scale payoffs or to rather use the adjusted replicator dynamics (see below).

replicator dynamics are equivalent for single populations, apart from a change in the rate of time passage, but Maynard Smith (1982) warned that the equivalence does not extend to the case of multiple interacting populations, each with its own average payoff.

Which model is right? Armed with a defensible microfoundations model, from the previous section of this chapter, a view can be formed. It was seen that the rate at which individual agents revise their strategies must be $\Delta\overline{\pi}$ to obtain the standard replicator dynamics, and simply Δ to obtain the adjusted replicator dynamics. Under the former, agents revise their strategies faster the higher $\overline{\pi}$ is, while under the latter, the revision rate is independent of $\overline{\pi}$. Since $\overline{\pi}^P > \overline{\pi}^R$, because proposers are getting larger shares of the pie in all reported results, individual proposers will be revising their strategies at a faster rate than responders under the standard replicator dynamics, *even when* $\Delta^P = \Delta^R$ and $\delta^P = \delta^R$, i.e. the selection and mutation rate parameters are equal for the two populations. We could say that the use of the standard replicator dynamics tends to trigger the GBS result because it features proposers that adjust their strategies at a higher rate, thus weakening the effect of mutation for proposers, relative to responders.

In biological applications, payoffs are typically interpreted as biological fitness, i.e. the number of viable offspring, and the distinction between standard and adjusted replicator dynamics hinges on whether generations overlap (standard) or not (adjusted) (Taylor and Jonker, 1978, p. 149). In social science applications, this distinction does not apply so the choice is harder to motivate. One potential advantage of the adjusted dynamics is that they are independent of average payoff levels (Binmore and Samuelson, 1994, p. 57) – this should indicate that the adjusted dynamics should be preferred if payoffs are inherently meaningless numbers, and/or if there is some reason to think that revision frequency should be exogenous.

In the case of the UG, the payoff numbers are not entirely meaningless - they are money/resource amounts. The question is whether people should learn faster if payoff levels were higher, in which case the standard replicator dynamics (used by GBS) would in fact be more appropriate. It should be quite natural for economists to think that higher payoffs, or, more specifically, higher potential gains/losses from good/bad choices, could act as an incentive to invest more mental and other resources into the process of decision making. In laboratory UG experiments, higher stakes appear to drive faster learning (Slonim and Roth, 1998; List and Cherry, 2000). Mäs and Nax (2016) find that people make costlier mistakes less frequently in coordination games. Similar effects should apply in a setting of cultural evolution.

GBS themselves (p. 65) argue that agents have limited computational resources, so the higher potential gains for proposers should result in more diligence in their choices compared to responders, by which they justify applying lower mutation and higher selection rate parameters for proposers. The argument could as easily be applied to argue in favour of using the standard over the adjusted version of the replicator dynamics, because in it, selection becomes proportionally stronger as average payoffs rise. Average payoffs is a sensible proxy for potential gains of switching, especially as the switching that really matters for stability of an equilibrium would be for very poor compared to optimal or near-optimal strategies. Consider a situation near an equilibrium of the model, where almost all proposers make some modal offer and almost all responders demand this offer or lower amounts. UG proposers who make offers that get rejected get a payoff of zero, while those making the modal offer would get an expected payoff approximately equal to the average payoff. Similarly for responders, the difference that really matters is between strategies that reject the modal offer, getting a payoff of zero, and strategies that accept it, which get a payoff approximately equal to the average payoff for responders.

However, this presents a challenge: are GBS not perhaps overdoing it when they use the standard replicator dynamics, which feature inherently stronger learning for proposers due to their higher average payoffs, and *on top of that* also apply differential selection and mutation parameters favouring stronger learning by proposers?[26] The question of exactly how much boosting of proposer learning relative to responders is appropriate is probably unanswerable, but one can at least attempt to clarify the effects. With a modal offer of 9, proposers have average payoffs around three times that of responders, so even with equal selection and mutation rate parameters, proposers would be revising their strategies roughly three times as often as responders. At rest points of the system, this would have the same effect as a reduction by two thirds of the proposer's mutation rates, which we know would tend to support the stability of non-subgame-perfect equilibria. To see this, simply note that at rest, $dx_i/dt = 0$, so $\Delta x_i\,(\pi_i - \overline{\pi}) = -\delta\,(1/N - x_i)$. If the LHS were multiplied by k, it would have the same effect as multiplying δ by $1/k$, either to bring the two forces into balance or to disturb such a balance.[27] GBS report results for various parameter values, but it seems that a boost to responder mutation rates relative to proposer mutation rates by a factor of ten gives a clear result, so let us consider this ratio as an example. With equal mutation rates, the same effect as this particular ratio, by the aforegoing logic, would require the selection rate parameter of proposers to be ten times that of responders. But in

[26] Despite a fair number of references to the paper, GBS's assumption of higher responder mutation rates do not appear to have been challenged before. It also appears in Binmore and Samuelson (1994, p. 55), but in this case the adjusted replicator dynamics are used, which do not feature inherently stronger learning for proposers so it may be more appropriate to impose differential mutation rate parameters here. The argument is again slightly elaborated upon in Samuelson (1997, p. 153), mentioned in Fudenberg and Levine (1998, p. 84) and again (with endorsement) in Akdeniz and Van Veelen (2023, p.580).

[27] We see also that any equiproportional changes to selection and mutation parameters would leave the coordinates of rest points undisturbed, though trajectories away from rest points need not be unaffected, thus affecting basins of attraction, velocity, etc.

the imitation model, this implies individual proposers would be updating their strategies around *thirty* times as fast as responders[28] – perhaps too much?

The other individual-level learning/imitation models reviewed in the previous section do not all show as clearly that, under standard replicator dynamics, proposers learn at an effectively faster rate than responders due to their higher average payoffs. Nevertheless, the same effect is present in some form or another in all of them. For example, in the GBS random aspiration level model, and other models where revision is driven by payoffs relative to a standard, selection is driven by the fact that strategies with relatively low payoffs are relatively more likely to switch their strategies. However, if the average payoff in the population is low relative to the standard, average switching rates, including for the relatively better strategies in the population, will be high, thus diluting the differential effect which would have allowed relatively poor strategies to fade away; the net effect is more inertia and less effective adaptation towards relatively optimal strategies in low average payoff populations. By contrast, higher average payoff populations will more effectively be able to eliminate poor strategies because there is less switching among the bulk of the population, so the comparatively higher switching rates of a minority with poor strategies are more effective at changing the population's composition.[29]

Under both the Proportional Imitation and Imitation of Success rules discussed in the previous section, an agent picked for revision will revise or not with a probability that depends positively on the payoff of the potential target strategy and thus on average payoffs, thus higher-average-payoff populations will tend to have more frequent payoff-increasing switching. Under the Imitation of Success with Repeated Sampling rule, as well as my own model, agents always switch when granted a revision opportunity, so there is no link between switching rates and average payoffs, and selection is achieved by biasing switching towards high-payoff strategies, so a lower average payoff does not add inertia.

[28]To be unnecessarily exact, at a rest point featuring modal offer of 9, with $\delta^P = 0.01$ and $\delta^R = 0.1$, simulations show that the ratio of average proposer payoff to average responder payoff is 3.44. And we would also need to take into account that GBS always choose $\Delta = 1 - \delta$ so setting $\delta^P = 0.01$ and $\delta^R = 0.1$ also means $\Delta^P = 0.99$ and $\Delta^R = 0.9$. To achieve the same effect with equal mutation rates would then require $\Delta^P/\Delta^R = 11$. So, in the imitation model, proposers would be updating their strategies 37.85 times as fast as responders.

[29]GBS (p. 87) state that their procedure assigns identical learning rules to the two populations, because they draw aspiration levels from the same uniform distribution for the two populations. In practice, this may mean that responders, often frustrated at rarely achieving the kind of payoffs typically enjoyed by proposers, have a high rate of distributionally neutral switching that simply dilutes the selection effect. The effect is similar to assigning a higher mutation rate to the responder population. Perhaps, rather than drawing from a common distribution, it would be more fitting to align aspiration levels with achievable payoffs within the practical scope of the given role.

3.6 Stochastic effects in finite populations

In section 3.4, I appealed to the law of large numbers to set the differential motion of aggregate frequencies equal to their expected values in the stochastic learning/imitation model that was developed there. GBS (p. 86) and many others working with aggregate dynamics make such an assumption as a matter of standard practice (Sandholm, 2010, p. 119). Friedman (1998, p. 20) discusses several benefits of the assumption. It is understood that deterministic aggregate dynamics models serve as approximations to stochastic individual-level models in *large* populations over finite time horizons where stochastic effects are independent and idiosyncratic, so cancel out in aggregate; the underlying stochastic models are often implicit.

Stochastic effects can occur in matching of players in interactions that affect payoffs or it can occur in the learning / mutation processes. As my interest here is in the latter, I maintain the questionable assumption that each agent receives the payoff equal to the expected payoff for its strategy. But perhaps this is not so unreasonable: recall that payoffs matter in the stochastic imitation model because agents selected for strategy revision choose a new strategy at random, but with probability weights proportional to payoffs. The role of payoffs is thus limited to demonstrating the effectiveness of various strategies to others. The random choice of a new strategy can therefore be interpreted as imperfect information about payoffs, but it could just as well be interpreted as the payoffs themselves being stochastic.

What kind of stochastic effects related to the learning process do we have in mind? Consider, for example, that the stochastic model allows an individual agent to sometimes imitate a lower-performing strategy, and it is possible in finite populations, through a combination of such chance events, that enough agents do this at the same time for the aggregate frequency of a lower-performing strategy to increase, which could never happen in the deterministic approximation. Such accumulation of improbable effects can also occur due to the mutation process described, which allows any individual agent to switch, with some small probability, to any other strategy.

Consequently, in a finite population, there is a small probability of significant jumps in aggregate frequencies occurring that do not occur in the deterministic model. Such fluctuations can be regarded as insignificant if the system has a tendency to correct itself; for example in the vicinity of an asymptotically stable equilibrium.[30] Stochastic fluctuations outside of equilibrium can also be regarded as insignificant if they shift the system to a slightly different trajectory, but the system ends up at the same or very similar equilibrium along a qualitatively similar trajectory as the original over a comparable time span.

[30]The defining essence of an asymptotically stable equilibrium is that small stochastic jumps are counteracted by the aggregate dynamics, thus any small "errors" are automatically corrected.

But stochastic effects can also potentially take the system away from a stable equilibrium, move a system from the basin of attraction of one equilibrium to another or have large effects on the timing of evolutionary developments. In short, if such stochastic effects are significant, deterministic dynamics may no longer provide an accurate description of the system it is intended to model.

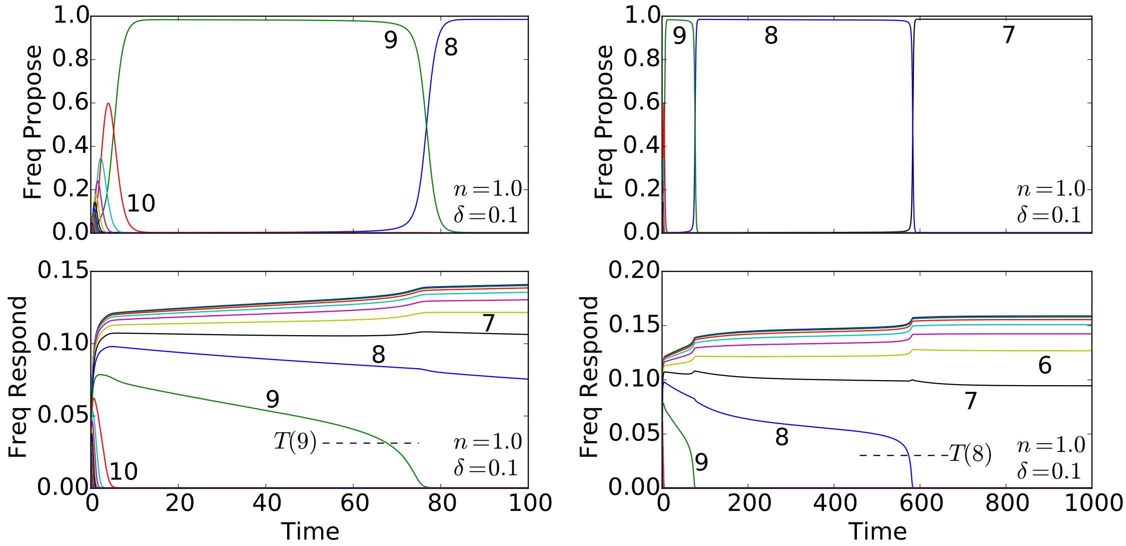
Can stochastic fluctuations of this nature affect the stability of an equilibrium in the Ultimatum Game? In principle, in a dynamical system with multiple asymptotic attractors, when agent-level stochastic elements are introduced, there is always the remote possibility that a combination of improbable stochastic movements can take the system out of the basin of attraction of one equilibrium and to the basin of attraction of another equilibrium (see Kandori et al., 1993; Young, 1993). In such a system, given infinite time, the system will cycle between all equilibria. In the UG, for example, suppose the current modal offer is 6, with almost all proposers making this offer and an acceptance rate close to 100%. There would be very strong selection pressure against any mutant responder demanding an offer like 8, as the mutant would reject the modal offer and get a payoff of zero. Only if, by coincidence, almost all proposers simultaneously started offering 8, would the strong selection pressure against these mutants disappear. Such simultaneous coincidences are simply too improbable to have relevance to our context of cultural evolution over longer time spans and involving large numbers of agents.

More interesting stochastic effects could occur when aggregate selection and mutation dynamics are "weak", meaning that following a stochastic shift, the system would ordinarily take a very long time to "correct", if at all. Under such circumstances, the system can "swim upstream" (Binmore and Samuelson, 1997, p. 248): stochastic shifts can accumulate and potentially overwhelm the aggregate dynamics with significant probability. Indeed, consider one of the non-subgame-perfect asymptotically stable equilibria identified by GBS. Almost all responders accept the modal offer associated with the equilibrium, but a significant share of responders follow suboptimal strategies that would reject positive offers below the modal offer. There is some selection pressure against these strategies, because of the very low-frequency presence of mutant proposer strategies that make such low offers (which are themselves quickly eliminated by strong selection pressure), but the force is weak. Mutation is always weak, by design. When two such weak forces are delicately balanced against each other, random disturbances can be consequential. The GBS result for the UG, which we know to be sensitive to selection and mutation parameter values, may be a such a case where stochastic effects can be consequential for finite populations of reasonable size.

A straightforward way to establish whether stochastic effects are significant is to run computer simulations of the underlying stochastic model and compare the results to simulations of the corresponding deterministic model. In the remainder of this section, I first present some graphs showing the evolution of frequencies of proposer and responder strategies according to the deterministic (standard) replicator dynamics equations. These serve both to clarify the evolutionary trajectories of the deterministic system, thus helping to understand the GBS result of stable imperfect equilibria, and as a baseline to compare to the trajectories of the stochastic model simulations that follow.

3.6.1 Deterministic model simulation

While I have argued in section 3.5 that unequal selection and mutation rate parameters for the two populations may not be justified under standard replicator dynamics, the GBS result occurs even when proposer and responder selection and mutation rates are equal (see GBS, p. 63, or simulation results below). I will investigate this case, i.e. $\delta^P = \delta^R = 0.1$ and $\Delta^P = \Delta^R = 0.9$.



(a) Up to t=100 (b) Up to t=1000

Figure 3.6.1: Deterministic UG simulation

I follow GBS in discretizing the continuous noisy replicator dynamics (2.1) as follows: $x_i(t+\tau) - x_i(t) = \tau\left[\Delta x_i(\pi(i,\mathcal{S}) - \pi(\mathcal{S},\mathcal{S})) + \delta\left(\frac{1}{40} - x_i\right)\right]$, where $\tau = \frac{1}{100}$.[31] Figure 3.6.1 shows the evolution of proposer and responder frequencies for each of the forty possible strategies over time, starting from uniform distributions. On the left-hand, two separate graphs with a common time scale show the evolution strategy frequencies for proposers (top) and responders

[31] Experimentation with smaller values confirmed that this provides a close approximation to the differential equation.

(bottom), while the right-hand-side graphs show the same simulation over a longer time span. The numbers indicate specific strategies, e.g. "9" on the top indicates the proposer strategy of offering 9 to the responder, while the on the bottom the numbers indicate minimum acceptable offer amounts for responders. It can be seen that at the beginning there are rapid developments, after which the system settles into a series of equilibria punctuated by occasional critical transitions to new equilibria. By the end of the simulation, all variables are stationary.

Initially, high demands by responders quickly die out, but some, such as 10 (call this strategy R10) linger long enough to make offers of 10 (call this P10) the optimal strategy for a short time, and the frequency of P10 can be seen to increase near the beginning.[32]

However, this is transient, because enough proposers are making offers below 10 for R10 to diminish rapidly (when other, even worse, strategies have died out), so that P9 becomes the optimal strategy for proposers before long. P9's frequency rises and by time $t = 20$ the system has reached an equilibrium where approximately all proposers are offering 9 and approximately all responders accept 9 using a mix of demands of 9 and lower (call this the the "P9 equilibrium"). This is indeed (very close to) a Nash equilibrium, because all strategies (barring random mutations at very low frequencies) are best responses to the mixed strategy represented by the opposite population's distribution of strategies. Any strategy that does not reject 9 is (near-)optimal for responders, because rejecting it entails a payoff of zero. For proposers, an offer of 9 is a best response given that it is accepted by (nearly) all responders and that the risk of rejection when making a lower offer outweighs the potential reward. The latter is the case when the frequency of responders that would reject a lower offer, closely approximated by the frequency of R9, is higher than a certain threshold, $\frac{1}{32}$, which is indicated as $T(9)$ in the figure.[33]

The system remains in the P9 equilibrium for a long time, but the composition of responder strategies changes during this time. Amongst others, it can be seen that the frequency of R9 declines in favour of other strategies that also

[32] Recall that the frequency of strategies will increase if their payoffs are higher than the current average payoff. At the very beginning, the average payoff is quite low, so strategies that are not optimal but still above average may rise, but the average quickly catches up so that only optimal strategies, or very nearly optimal strategies (i.e. strategies that would be optimal except for low-frequency noise strategies in the opposite population), can increase their frequencies through selection from about time t=5 onwards. The evolutionary dynamics are particularly efficient at eliminating behaviour that leads to rejections, and after the initial period, rejections occur only at very low frequencies (due to mutations).

[33] For a proposer, the frequency of R9 represents the approximate probability that P8 would be rejected, given that higher demands have frequency of approximately zero. Switching from P9 to P8 means proposers can gain an extra dollar for themselves with probability approx. $1 - x_9^R$, but there is a probability of approx. x_9^R that the lower offer would be rejected, which would entail a loss of 31 dollars.

accept P9. Eventually, the frequency of R9 approaches the $T(9)$ threshold, and an equilibrium transition is initiated that eventually results in a new equilibrium, P8. Even before R9 reaches the threshold, proposer strategy P9 starts to decline and P8 starts to rise, slowly at first, as payoff gradient between P9 and P8 flattens and selection's force weakens, allowing mutation's effect to be more assertive despite being an inherently weak force. Now, selection pushes P9 up while mutation pulls it down, but there is a tiny advantage to mutation, leading to the slow but accelerating downwards trend. Once the $T(9)$ threshold is actually crossed, the equilibrium is broken and proposer selection changes direction, now working together with mutation to pull P9 down rapidly to near zero, where it remains forever. Fast forwarding to around $t = 570$, another equilibrium transition takes place, the frequency of R8 having crept down in a similar fashion as R9 earlier, but over a longer period of time, to finally break through the threshold $T(8) = \frac{1}{33}$ (see figure 3.6.1b).

In summary, we see that there are extended periods during which the system is in a state of equilibrium that are punctuated by rapid transition periods, after which it settles into a different equilibrium. Furthermore, the system always moves from higher to lower-offer equilibria, since once a substantial portion of proposers make a particular offer, there is permanent and strong selection pressure on responders not to reject that offer. Successively lower offer equilibria appear to be increasingly difficult to escape, and more time is spent in relatively lower offer equilibria. Ultimately, an equilibrium, which is not necessarily the SPNE, is reached that is so difficult to escape that the system remains in it forever.

3.6.2 Stochastic model simulation

The stochastic model is an agent-based model, with each individual's selection and mutation modelled according to the imitation model in 3.4. As the model is continuous-time, I follow the same discretization procedure as above, i.e. time proceeds in small steps of $\tau = \frac{1}{100}$. In each small time step, each agent has a probability of $\tau\Delta$ of being selected for strategy revision, and if selected, choose a new strategy at random from the population with probability weights equal to current expected payoffs. Individual strategic interactions are not modelled (as explained above). The same parameter values as before are used, $\delta^P = \delta^R = 0.1$ and $\Delta^P = \Delta^R = 0.9$, as well as the same uniform initial frequencies.

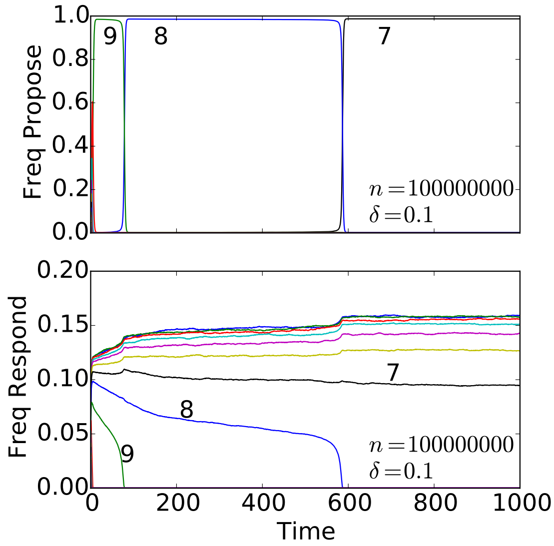
Agent-based modelling can be notoriously computationally intensive (Parry and Bithell, 2011), and one rarely sees simulations for large populations of over a few thousand agents or so. Fortunately, in the imitation model, a simple programming technique could be used to dramatically speed up simulations and allow very large populations to be modelled. Instead of determining, in each small time step, if each agent is to be selected for strategy revision, a single draw from a binomial distribution is made per strategy that gives the number of agents whose strategy should be revised. Then, new strategies can be selected for all revising agents in a population in a single step, using a single draw from a multinomial distribution,[34] with weights equal to $x_i\pi_i$. The weights only need to be calculated once per strategy per time period. The number of agents following each strategy can be adjusted each time period by subtracting the number revising agents and adding the relevant number from the multinomial draw. The computational intensity of this procedure is independent of N, so arbitrarily large number of agents can easily be modelled without adding to computational requirements or running time. As can be seen below, this was very useful, and necessary, to model a large enough number of agents that the infinite-population deterministic model's results could be approached convincingly.

Figure 3.6.2 shows the results of four simulations with the same model and parameters, only differing by the number of agents. For an extremely large population, with 100 million proposers and 100 million responders (a), the simulation proceeds as predicted by aggregate dynamics (2.1) (see figure 3.6.1b), but as N is reduced, responder frequencies in particular become visibly noisy. With a million agents in both populations (b), the end result is still essentially the same, as the the system progresses through the different equilibria and remains in an equilibrium with offers of 7 after $t = 1\ 000$, but notice that the timing of the second equilibrium transition is much earlier. Despite the significant stochastic drift in responder frequencies, R7 never drops near enough to the critical threshold $T(7) = 1/34$ for the system to transition to a lower-offer equilibrium in this simulation.[35] Different behaviour is obtained for smaller populations: if $N = 100\ 000$, stochastic drift can more easily pull R7's frequency below the critical level and the system transitions to a lower-offer equilibrium.[36] Finally, when $N = 1\ 000$, there is so much stochastic drift that no equilibrium is stable in practical terms – at $t = 1\ 000$, proposers are making offers of 2. Even here, though the effect of stochastic drift on proposer frequencies is clearly visible, it has no practical significance, because strong proposer selection quickly corrects any movement away from the optimal strategy. Instead, stochastic drift effects transition changes in smaller populations via its effect on non-rejecting responder strategies, and in particular the relative fre-

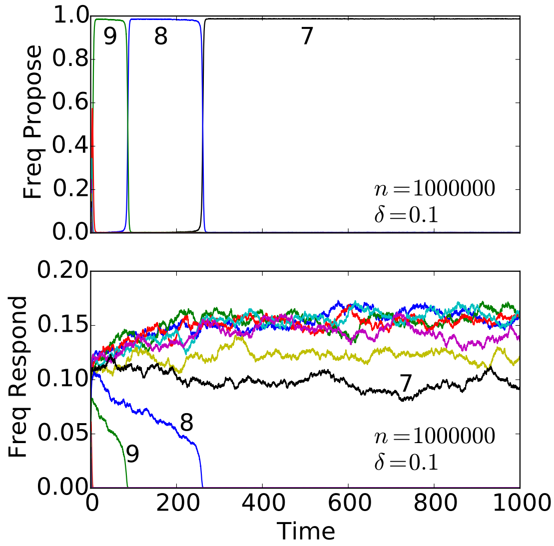
[34] My simulations were implemented in Python, using the NumPy library which has efficient functions for binomial and multinomial distributions.

[35] It seems probable that, if the simulation were allowed to run for long enough, an equilibrium transition would occur. However, I ran four additional simulations with $N = 1\ 000\ 000$ for an extended time (until $t = 80\ 000$), and in all cases the simulations remained in P7 equilibria. In the different runs, the timing of the transition to the P7 equilibrium is sometimes earlier and sometimes later than in the deterministic model.

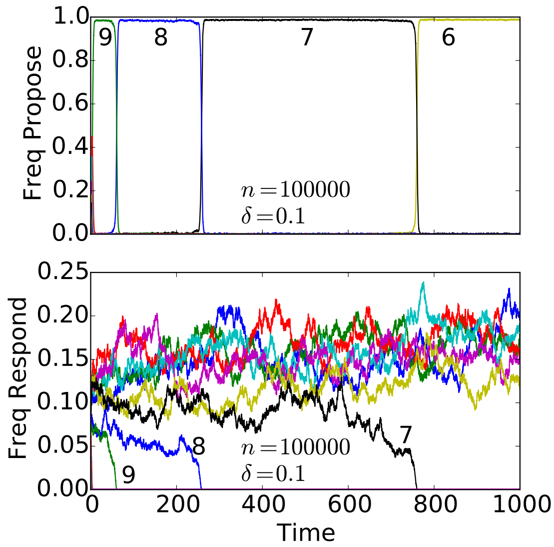
[36] This does not always occur with these parameters: in most of them the system reaches $t = 1\ 000$ still in an equilibrium with offers of 7. However, if the simulations are left to run longer, they eventually all transition to a P6 equilibrium at some point after $t = 1\ 000$.



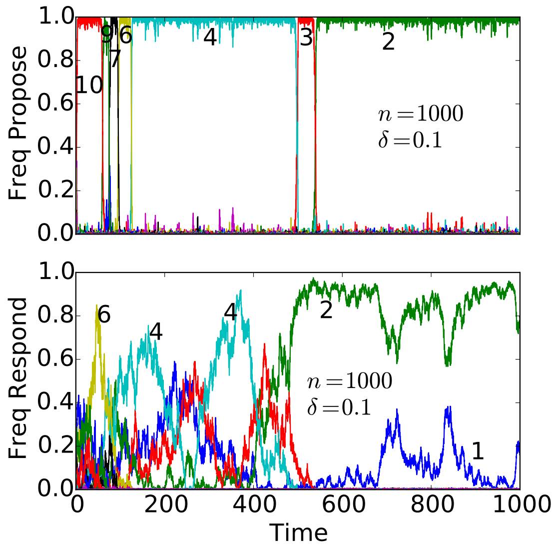
(a) $N = 100\,000\,000$



(b) $N = 1\,000\,000$



(c) $N = 100\,000$



(d) $N = 1\,000$

Figure 3.6.2: Stochastic UG Simulations

quency with which the equilibrium offer, is demanded.

For the case where differential selection and mutation rate parameters are used for proposers and responders results are generally similar. I also ran simulations with $\delta^P = 0.01$ and $\delta^R = 0.1$. Recall that an equilibrium offer of 9 can be stabilized in the deterministic model with these parameters. Simulations of the stochastic model with these parameters resulted in equilibrium offers of 9 for $N = 100\ 000$ at $t = 50\ 000$, i.e. the same result, but offers as low as 5 for $N = 10\ 000$, and offers as low as 2 for $N = 1\ 000$. The timing of equilibrium transitions can be highly variable between different runs with the same parameter values, but, similarly to the deterministic version, the system transitions from higher to lower-offer equilibria and eventually tends to settle at some equilibrium above one without additional transitions taking place during time-spans that can feasibly be simulated. The offers at these "final" equilibria are either the same as at the deterministic model (e.g. for $N = 1\ 000\ 000$ and higher) or lower (for smaller N).

For either set of mutation rate parameters, if N is small enough, e.g. in simulations with $N = 520$, the system does eventually reach the SPNE offer of 1, thus showing that the large population assumption is vital for the GBS result. While it is not perfectly clear what an appropriate value for N would be for interpreting experimental results, it would seem that for real-world settings in a broader social context, there would be enough noise in learning, with occasional mistakes, for the subgame-perfect result to be possible in principle. But the time required to reach the subgame-perfect result might be so large that we should expect to observe the subgame-perfect result only very rarely, unless initial conditions are at or near the SPNE, in which we may expect such conditions to be maintained.[37]

3.7 Conclusion

The results presented in GBS remain as interesting and important as ever from an evolutionary game theoretic perspective, but the relevance of the model and results to the real world has not been clear. Since the paper was published in 1995, a large amount of experimental research has been conducted around the world, using the ultimatum game (UG) as a basis to learn about human behaviour and social preferences in simple strategic interactions. There has been less enthusiasm for evolutionary explanations of experimental results than for explanations based on rational behaviour combined with social pref-

[37] I ran 215 simulations with $\delta^P = 0.01$ and $\delta^R = 0.1$ and $N = 520$; on average the first time 95% of proposers offer 1 was at $t = 8591.5$, though it is also extremely variable, with standard deviation 8038.2. The same set of simulations also confirmed that the probability of a transition from a lower to a higher offer equilibrium is too low to be of practical significance – not a single such transition was observed (though for even smaller N it may well occur).

erences, and it is clear that evolutionary models by themselves cannot explain some features of experimental data. Nevertheless, a number of experimental researchers have found some evidence of learning taking place as subjects play the same game repeatedly with different opponents, and certain features of the evolutionary model, e.g. that learning takes place more clearly among proposers than among responders, have been validated.

In retrospect, however, it seems evident that expecting evolutionary models, understood solely as models of interactive learning involving boundedly rational yet narrowly self-interested agents, to account for all or even the majority of the rich behaviour and social preferences observed in experimental data is an unrealistic pursuit. Fairness norms and social preferences are deeply held and not simply forgotten after a few rounds of play. This points to the need to better understand the origins of initial behaviour in experimental data.

Evolution is a powerful idea that can apply to different contexts and different timescales, and applying evolutionary thinking and modelling in an attempt to understand ultimate, rather than proximate, explanations for observed behaviour remains a promising prospect. I argued that a reinterpretation of the GBS model along such lines could prove insightful, but careful attention needs to be given to issues of interpretation of a model that works only at an aggregate level. An individual-level learning model based on social imitation was therefore developed, to serve as plausible microfoundations for the standard replicator aggregate dynamics model. I have argued that the imitation model combines simplicity of interpretation and is particularly well suited to modelling cultural evolution, where people copy other people's behaviour, with a bias towards strategies that perform well individually. This seems most compatible with the idea of humans as a social creature, driven by a combination of social norms and material incentives.

However, once the individual-level model has been adopted, it was found that there are some important implications for the interpretation of the aggregate model. Specifically, in the standard replicator dynamics employed by GBS, which shows the result of asymptotically stable non-subgame-perfect equilibria most clearly, individual agents learn faster the higher the average payoff for their population. The adjusted, or Maynard Smith, replicator dynamics do not have this effect. This implies that the form of the model chosen by GBS favours selection for proposers relative to responders, which tends to strengthen their result. This is not problematic in itself, as there is a good argument that proposers should pay more attention because they have more to lose if they play a poor strategy, but GBS then also argue in favour of applying a higher mutation rate parameter value to responders, which multiplies the effect and ends up producing a system in which the effective balance of selection and mutation may have been pushed further in a specific direction than might be considered reasonable.

Even with equal selection and mutation rates in the standard replicator dynamics, the basic GBS result still appears, just a little more weakly. But the result is sensitive to parameters,[38] and will also be affected by other factors not considered in detail, such as what monetary units are used to measure payoffs. Given the sensitivity of the main result to model setup and parameters, it is unsurprising that the result also does not hold up well in finite populations, where stochastic disturbances can upset the delicate balancing act required to maintain stable non-subgame-perfect equilibria. The agent-based model simulations presented in section 3.6 show that the aggregate dynamics are a good approximation for the underlying individual-level stochastic model only for very large populations in excess of a million agents per population, contrary to Friedman (1998, p. 20)'s arguments suggesting that much smaller numbers can often be approximated by infinite populations. If finer details like the timing of equilibrium transitions are to be matched, then even more agents, e.g. a hundred million, are required.

Overall, these results suggest that less relevance should be accorded to the GBS result of asymptotically stable subgame imperfect equilibria in cultural evolutionary processes. The model may be able to explain why non-subgame-perfect behaviour may persist for extended periods of evolutionary time, but it is unlikely to be able to account for its indefinite persistence. The long-run tendency of the model, accounting for stochastic shocks typically expected in finite populations, is to converge towards the SPNE.[39] Moreover, as indicated by the simulation results in this chapter, the highly unequal subgame-perfect equilibrium, once reached by the evolutionary model, is more robust to stochastic disturbances than other equilibria. This suggests that the GBS result depends strongly on initial strategies being distributed so that there is substantial weight on higher offers and higher demands.

The GBS model can therefore not, on its own, provide an explanation for the origins of fairness norms and social preferences seen in experimental findings. This should not come as a surprise, given the argument that these proclivities likely originated from long-term exposure to a fluid and rich mixture of real-life interactions, which humans have engaged in daily over extended periods of genetic and cultural evolution. This emphasises the need for further research on evolutionary models that go beyond the standard ultimatum game.

GBS's paper and those that have followed it remain valuable for allowing a greater understanding of the ultimatum game's structure and in particular an appreciation of the different learning environments faced by proposers and responders. Regardless of whether a particular non-subgame-perfect equilibrium is asymptotically stable or not, we now understand that in its vicinity, responders have only a small incentive to deviate from it, while proposers have

[38] Harms (1997) and Akdeniz and Van Veelen (2023) confirm that the GBS result is sensitive to the precise way mutation is modelled, and GBS (p. 69) themselves also acknowledge this.

[39] This should not necessarily be interpreted as support for a prediction of SPNE play in any particular context, given the slow speed of adjustment in the model.

strong incentives to maintain it.

Chapter 4

An Evolutionary Perspective on Good and Bad Reputations in the Ultimatum Game

4.1 Introduction

A person with a reputation for being a tough negotiator may be able to obtain favourable outcomes in bargaining interactions. But how can such a reputation be obtained? Cheap talk is not an option in conflictual situations where there would be obvious incentives to dishonestly inflate one's apparent toughness. A more credible method might be a public demonstration of willingness to suffer a cost rather than accept an inferior deal. This could be rational if there were a high enough probability that this act will become known to future bargaining counterparts, and that their behaviour will be changed in the desired way through this knowledge. But what is often overlooked is that such a strategy can only be effective if there are sufficient opportunities to build up a track record – it is necessary to receive some low offers to be able to show that you will reject them.

The effectiveness of a signalling strategy to convey useful information is endogenous if the opportunities to build a reputation depends on the distribution of counterpart behaviours among those with whom an agent interacts. There seems to be an inherent instability when the reputation's purpose is to deter unwanted behaviour: if the reputation-building strategy were effective, it might undermine itself by disincentivizing the behaviours that allow it to effectively convey useful information.

In a bargaining situation, if one player uses a tough strategy of rejecting bad offers, and the other accordingly refrains from making bad offers, both players could be playing sequentially rational strategies despite there being no effective demonstration that low offers would be rejected. But if you acted "as if" your opponent was a tough negotiator, rather than on the basis of credible information, you may not even notice if your opponent's strategy changed, so a point may be reached where you are no longer playing a best response to the actual strategy of the other player. Such situations could not be regarded as stable. It is also not clear how such an equilibrium can be reached in the first place. In repeated games, if there are equilibria supported by contingent punishments and/or rewards, there will generally also be other equilibria where contingent strategies are not used (Mailath and Samuelson, 2006, p. 4), and the parties may not face adequate incentives to learn the behaviours associated with a particular equilibrium, especially if, as indicated, the information on which the equilibrium depends may not be strong at critical times. Equilibrium selection issues can therefore not be ignored.

These questions call for an evolutionary analysis, which this chapter uses to study one of the simplest bargaining games, the ultimatum game. In the ultimatum game, the first player, the proposer, gets a single opportunity to choose a share of some fixed total amount of money to offer to the second player, the responder, who in turn gets a single opportunity to accept or reject the proposal. If accepted, the players are rewarded according to the proposal – the responder gets what was offered and the proposer the remainder – while rejection leaves both players with nothing. Researchers have been fascinated by the stark disparity between the predictions of theory, based on rational action and foresight, and experimental results. The game's unique subgame-perfect Nash equilibrium (SPNE) is an entirely lopsided division in favour of the proposer and no rejections, while experiments mostly show relatively equal divisions proposed and unequal divisions often rejected.[1]

It is plausible that norms, preferences and behaviour that experimental subjects bring to the laboratory may be the result of genetic and cultural evolutionary processes that have taken place outside the laboratory, in everyday life situations, over different timescales (Gale et al., 1995, p. 70, Mailath, 1998, p. 1350, Güth and Napel, 2006, p. 1038, Skyrms, 2014). For this reason, it is interesting to consider evolutionary dynamics for a system of strategically interacting agents engaging in repeated interactions that can plausibly be approximated by the ultimatum game as played in a research laboratory, but with some details of the interaction changed to be more reflective of real-world scenarios. Developing a reputation for being a tough negotiator is typically not possible

[1] Human proposers typically offer between 30 and 50 per cent of the money and responders often reject offers below 20 per cent (Camerer and Thaler, 1995, p. 210, Oosterbeek et al., 2004, Güth and Kocher, 2014, p. 398). The dominant interpretation is that responders experience positive utility from rejecting what they regard as inferior or unfair offers (Fehr and Schmidt, 2006, p. 630), and proposers are motivated by the need to avoid rejection and (to varying degrees) other-regarding preferences e.g. altruism, egalitarianism or social norms (Cooper and Kagel, 2016).

in a laboratory,[2] where great care is usually taken to ensure that interactions are anonymous, but is likely a central concern in many real-world bargaining situations.

Previous authors (Nowak et al., 2000; Poulsen, 2007; Debove et al., 2016; Zhang et al., 2023; Akdeniz and Van Veelen, 2023) have reported encouraging results from evolutionary models of the ultimatum game with some form of reputation. In their original paper, Nowak et al. (2000) present a deterministic analysis of an infinite-population minigame using replicator dynamics, alongside results from agent-based computer simulations that involve a finite population and a larger strategy space. The latter type of model is more flexible in terms of what can be modelled, and can easily be made more detailed and realistic, but it can sometimes be difficult to provide comprehensive and tractable explanations for the results, so it is sensible to begin with a thorough analysis of a simple minigame.

In Nowak et al. (2000)'s minigame model, there are only two possible offers proposers can make: High (H) and Low (L)[3] and two corresponding responder strategies, namely to demand H or to demand L. If a responder "demands" i it means that the proposer adopted a monotonic minimum acceptable offer (MAO) strategy, accepting offers of i or higher and rejecting any lower offer. H can be thought of as the "fair" offer, e.g. half of the total amount, while L can be thought of as a low amount that would leave the proposer with a much larger payoff if accepted.

To model the effects of reputation in their minigame analysis, Nowak et al. (2000) assume that whenever an L-responder (that would accept L) meets an H-proposer (who would ordinarily offer H), the proposer reduces her offer by a constant amount reflecting a possibility that the proposer finds out about and exploits the information that the responder would accept L. We can call this kind of information a negative reputation as it is something that a responder would prefer not to be known. H-proposers can only gain by offering less to responders that will definitely accept L. The conditionality in the H-proposers' strategy creates a direct incentive for responders to switch to demanding H (i.e. rejecting L offers), and if responders demand H, proposers have an incentive to offer H (to avoid the risk of rejections) so this creates an asymptotically stable equilibrium. But there is also a second stable equilibrium, namely the UG's SPNE where L offers are made and accepted, because if the proportion of H-proposers is low, the incentive to avoid a bad reputation is diminished, and if demands are low the strategy making unconditional low offers is better for proposers.

The finding of "fair" outcomes (and more generally, non-subgame-perfect

[2] A notable exception is Poulsen and Tan (2007).

[3] I use the following terminology: if a proposer "offers" i then the proposed division is $\$ - i$ to the proposer and i to the responder, where \$ is the total amount to be divided.

stable equilibria) in an evolutionary model of the UG is undoubtedly interesting, but there is a problem: if all proposers are offering H, then how can they ever learn that some responders would also accept L? There would be no instances of this to know about. Under endogenous information there would be zero information at this equilibrium. Responders who demand H and L would then be treated the same by the H-offering proposers, and they would get the same expected payoff, thus the stability of the equilibrium is no longer assured as the frequency of L-demanders may drift upwards to a point where proposers can gain by switching to L offers, thus unravelling the equilibrium. Fortunately, as I will show in this chapter (section 4.5.3), it turns out that in this case the problem can be resolved fairly simply by adding a tiny amount of mutation to the model, which stabilizes the H-offer equilibrium by increasing information slightly – enough to provide the necessary incentives that stabilizes the equilibrium.

This example should illustrate the need for a proper consideration of endogenous information in cases where the opportunity to signal useful information depends on strategy choices of others. Of course, information also affects behaviour (which is the point of it), so an equilibrium of a sort is needed where the information generated by patterns of behaviour leads to behaviour that generates the information. In this chapter, I develop and apply to the ultimatum game a general framework for determining such informational equilibria for a two-player sequential-move game. For every *strategy* profile, where strategies of the first player is contingent on information about the second player's past *actions* that the first player may have gained, an endogenous information equilibrium can be determined. An *endogenous information equilibrium*[4] is an information state that induces an *action* profile which itself generates the same information state. In the simple example above, if *all* proposers offer H then the information state generated cannot include any positive probability of a proposer gaining the knowledge that a responder accepted a low offer. Proposers will only be able to know that all responders they interact with accepted high offers, which is not useful information.

Realistically, information is partial, so a given responder action will only be known to future bargaining counterparts with some probability. I will assume that proposers have a fixed probability of finding out about any specific act of the responder that occurred within a randomly selected sample of a given size of the responder's interactions.[5] This allows proposers multiple chances of ob-

[4]I use this slightly cumbersome term to avoid confusion with the "informational equilibrium" concept due to Riley (1979).

[5]This scheme is superficially similar to the one used in the agent-based model in Nowak et al. (2000), where agents have a given probability of observing any of the interactions that have taken place in the current evolutionary generation, with the total number of interactions per generation fixed, so it acts like a variable-size sample. However, agents in their model record historical instances of past behaviour within a timeframe as generated by their computer sim-

serving a particular action as well as the possibility of observing multiple distinct actions (e.g. the responder rejected X and also accepted Y) at the same decision point.

The general framework allows all possible combinations of all possible outcomes of individual interactions to influence proposer behaviour. For example, even in the two-offer minigame, in the most general case, the following possible outcomes from single interactions can occur:

- Accepted L (negative reputation)
- Rejected L (positive reputation)
- Accepted H
- Rejected H

I call each of these bits of information a *signal*. Each of them, or any subset of them (including none) can make up an information bundle, called a *signal set*, available to a proposer when the proposer has to choose an offer. The proposer's strategy set becomes the set of possible mappings from the set of possible signal sets to possible offer amounts. Without restriction, there would therefore be $2^4 = 16$ distinct proposer strategies even in this simple minigame. Naturally, not all possible signals, signal sets or proposer strategies are reasonable or interesting, so restrictions can be applied to make the analysis tractable and computer simulations (of a deterministic evolutionary system based on the framework) feasible.

Among the signals above, we see not only one that indicates a negative reputation, as considered by Nowak et al. (2000), but also one indicating a positive reputation: proposers may want to *increase* their offers to responders they know to have rejected a low offer. As will be shown, an evolutionary model based on positive rather than negative reputations in the ultimatum game gives entirely different dynamics and results, so a holistic analysis needs to consider both kinds of reputation.

After presenting key conceptual ideas (section 4.2) and the general information framework (section 4.3), the framework is used to generate a series of minigame models that are analysed in detail. The first model includes only the negative reputation signal, which yields an endogenous information variant of the Nowak et al. (2000) minigame. In section 4.4, I demonstrate how the endogenous information equilibrium for this game can be explicitly solved. I also show that under endogenous information, both players' expected payoffs are influenced not just by their own strategy choices but also by the distribution

ulation, while in my framework, information at every point in time is the calculated probability distribution over the possible knowledge states a proposer can hold for any responder.

of proposer strategies, which determines the probability of the signal being observed. I also consider the implications of endogenous information on stability of equilibria, an issue already alluded to above. This section also investigates the conditions required to ensure that the information state can be uniquely determined. This generally requires that at least a small frequency of all possible proposer strategies are present, which can be achieved simply by adding a small mutation rate to an evolutionary model.

In section 4.5, deterministic computer simulation results based on noisy replicator dynamics are presented for a series of minigames – negative reputation, positive reputation and a model that combines both – analysing evolutionary trajectories, equilibria and stability in each case. The results show that the two types of reputation both have important and complementary roles in explaining relatively egalitarian outcomes in the ultimatum game. The section ends with a brief exploration of larger models including a model with more than two possible offer amounts, where more sophisticated ways of exploiting information is possible, and the final section concludes.

4.2 The value of reputation

In economics literature, the concept of reputation is often used together with concepts like trust and commitment. Trust is a willingness to take a risk in which a favourable outcome hinges on another agent's behaviour complying with an agreement or rule, even if it will not be in the other agent's best interest to comply. Commitment is a decision to act in a certain way, even if at that point it would not be the decision-maker's optimal decision. Both trust and commitment are tied to future actions, both can lead to current benefits for one (or both) players; and in both cases a similar dilemma arises – how can a person convince others that she will comply with the required behaviour, given that it will be against her interest at the later time?

Schelling (1956) describes a wide-ranging array of devices by which the desired credibility could be established in bargaining situations, including methods to reduce future freedom to act or make non-compliant actions costly, e.g. delegation, contracts and penalties of various kinds. In some of these methods, the concept of reputation plays an important role, especially in the sense of one's reputation acting as a bond that is posted to ensure that promises are kept: "if the buyer can assert that he will pay no more than $16 000 so firmly that he would suffer intolerable loss of personal prestige or bargaining reputation by paying more, and if the fact of his paying more would necessarily be known, and if the seller appreciates all this, then a loud declaration by itself may provide the commitment" (p. 284).

The strategies Schelling describes are deliberate, calculated, clearly communicated and understood by the other party, which is quite different to the instinctive behaviour seen in ultimatum game experiments, where responders simply have an emotional preference for rejecting unfair (or rude[6]) offers. Nevertheless, the value people attach to maintaining their reputations could still be a part of the ultimate explanation for why the behaviour we see evolved, provided it can be shown that the strategy can actually work in typical bargaining situations faced by people in everyday life (or faced by them in the past).

From an economist's perspective, the key aspect of reputation is that it is a type of information about a person's conduct that could affect how other people interact with that person, and the person may therefore have some interest in what information gets transmitted.[7] Reputation could affect not only the conduct of another player in a bargaining interaction, but also the choice of whom to bargain with,[8] what roles players adopt (e.g. who gets to be the first mover) and what rules to follow. Consider, for example, the situation where players could choose whether or not to engage in an ultimatum game interaction with another person. If there were no opportunity costs, there would be no reason to decline any opportunity to interact, as the worst that could happen would be a payoff of zero. But if there were some opportunity cost, e.g. interacting with A means not immediately being able to interact with B, then players could use reputational information to decline interactions where there is a reasonable prospect of a poor outcome. Effectively, players would be "shopping around" for a good bargain. Proposers may have to adjust their offers to ensure getting chosen by responders for a transaction, or responders may have to adjust their acceptable offers to be selected by proposers. If there were many sellers (proposers) as well as buyers (responders), the price might settle at a competitive equilibrium market-clearing level.

Despite considerable added complexity, this "ultimatum marketplace" concept is a promising avenue for further research to explain observed behaviour in ultimatum games, because in the real world people do have some control over who they interact with and there are likely signifiant opportunity costs.[9]

---

[6]See Camerer and Thaler (1995, p. 217).

[7]The term reputation sometimes refers directly to information about the quality of a person's abilities or a product's quality, rather than conduct per se, e.g. Rogerson (1983); Jackson (2005); Dellarocas (2006). Nevertheless, strategic concerns, e.g. whether to improve quality, or whether to hide or reveal information, and under what conditions the information can be regarded as credible, always seem to be salient.

[8]Reputation as a device to regulate partner choice and thus enable trading systems to function is an important theme in historical institutions literature (Milgrom et al., 1990). In law and economics, reputation could be related more to a general assessment of a person's moral character, but the end result is similar: "A good reputation implies that people are eager to transact with the individual, and a bad one that they are averse to transacting with him. Reputation affects the individual's wealth by determining the terms that people will offer him in transactions." (Posner, 1978, p. 11)

[9]The theory would be that evolution has sculpted human behaviour to work well in environments where walking away from a bad deal can be sensible because in the real world there are generally better outside options available. This could explain why responders feel justi-

The problem, however, is that if agents could choose who they interact with, it takes the sting out of the ultimatum game, perhaps to the point where the essence of the game, namely the extreme inequality in roles, is lost. If we allow a responder to “reject” a proposer before the proposer has even made an offer, we may be changing the interaction too drastically for it to remain a recognisable ultimatum game.

Perhaps the ultimatum game is artificial by design. However, it captures an important aspect of many real-world interactions: the inequality of power and institutions that determine bargaining rules that favour one agent above another – something humans must have evolved strategies to recognize and deal with. If we are to retain this aspect, we need to focus on interactions that are random and compulsory, where roles are fixed,[10] and there is no marketplace. Under these conditions, there seems to be no obvious role for proposers' reputations – their offers are on the table when responders choose to accept or reject, so what more information about them could be relevant to responders' decisions?[11] On the other hand, since responders move second, proposers should be interested in what offers the responders are likely to accept. Knowledge of the responders' past interactions could be informative, and responders should therefore value the kind of reputation that results in favourable offers.

4.2.1 The classical perspective: repeated games

The standard approach to modelling reputation in classical game theory uses the framework of repeated games. Here, the original game becomes a stage game that is played repeatedly, and players' strategies are mappings from complete histories to stage game actions (Mailath and Samuelson, 2006, p. 19). Various folk theorems show that in such settings there are generally many equilibria, including equilibria where players can incentivize other players to play certain actions using threats and/or promises about their own future behaviour.

fied in rejecting low offers – they instinctively feel that they could get a better deal. Of course, in a typical experimental design, such better offers will not be forthcoming, so there may be disappointment.

[10]Another avenue through which reputations could matter if roles are not fixed is positive reciprocity. A person known to make generous offers in the proposer role may attract generous offers by other proposers when in the responder role. See Fehr and Gächter (2000) for a general discussion and arguments suggesting that reciprocal motives are widespread in many contexts in which humans interact. Zhang et al. (2023) include reputations for proposers in the ultimatum game in their model, motivated by the notion of indirect reciprocity.

[11]Could a proposer develop a reputation for toughness in the sense of making low proposals *even to* a responder that is known to reject low offers? Could such a strategy be effective in persuading responders to capitulate, their bluff thus being called? If so, a responder would not really need to have knowledge of *past* interactions of the proposer, because the low offer on the table should already be a clear indication of the proposer's intransigence. More important is that the responder knows what the proposer knew when making the offer. While this is an intriguing possibility, it will not be pursued in this chapter.

The most well known example is the repeated prisoners' dilemma, where it is easy to show that it can be optimal for players to cooperate rather than defect, provided backwards induction cannot be used, so the theory requires infinite repetitions or that the end is stochastic and therefore unpredictable.

It is straightforward to apply a similar treatment to the ultimatum game. Consider a simple minigame example, where the surplus to be divided is 4, the proposer has two possible offers, $S^P = \{\mathrm{H}, \mathrm{L}\} = \{2, 1\}$ and the responder has the same two amounts as minimum acceptable offers (demands). The strategic form of this game is represented in table 4.2.1. In the once-off game, there are two Nash equilibria (NE), (H, H) and (L, L), while only the latter is subgame perfect if it is taken into account that the proposer moves first.[12]

		Responder	
		H	L
Proposer	H	2, 2	2, 2
	L	0, 0	3, 1

Table 4.2.1: Strategic-form representation of an ultimatum minigame

Suppose the game is repeated infinitely. Consider the strategy profile (C_H,H), where C_H means the proposer plays a default action of H every round, unless the responder was ever observed accepting an L offer, in which case the proposer switches to L every round. The responder simply plays H every round. If ever a subgame in which L was offered was reached, the responder would face a tempting deviation from her strategy of rejecting low offers. Accepting instead would gain $1 - 0 = 1$, but lose $2 - 0 = 2$ in every subsequent round, assuming the deviation was perfectly observable by the responder and that the proposer's action was a one-shot deviation, so in future rounds both players return to their stated strategies. Alternatively, the responder can switch to accepting low offers permanently, in which case the initial gain is still 1 but the subsequent losses would be $2 - 1 = 1$ per round. In both cases, the sum of losses exceeds the gain, so a rational responder would not deviate. The proposer would also not deviate since, given the responder strategy, the proposer can only lose by making L offers that would be rejected, so C_H is a best response. Thus, (C_H,H) is a subgame-perfect Nash equilibrium (SPNE). The result could of course be different if we incorporated a high discount rate for the responder: the immediate gain from accepting a low offer could then outweigh the eternal losses that would follow.

[12]The strategy of demanding H for the responder is irrational in the subgame that follows once the proposer chose L.

In this simple example, we clearly see the rationale for reputation in the ultimatum game, but there are several concerns. The most immediate problem is that we are not necessarily interested in a repeated game scenario as it would be a poor proxy for modelling once-off interactions.[13] But the repeated game can also be interpreted as a fixed responder interacting repeatedly with different proposers (see Mailath and Samuelson, 2006, p. 19), in which case the game might seem somewhat more like a one-shot interaction to the participants, though with the requirement that proposers have knowledge of the responder's past interactions with other proposers. Even if past interactions could be observed only imperfectly, there would still be a SPNE in which low offers would be rejected if the probability of discovery were high enough and the responder's discount rate low enough. Perhaps evolution has sculpted behaviour that implicitly recognizes the risk of any action becoming part of one's public record, even if the framing of the situation suggests that it is a one-shot interaction.

Another major problem in this simple model is that there is never any real information revelation – observed instances of low offers getting rejected is zero – thus the equilibrium is supported by hypothetical behaviour that is never actually observed – effectively an assumption on the part of proposers maintained throughout infinite repetitions of the same game. There are more sophisticated repeated-interactions models in which the one player has a degree of uncertainty about the other player's utility function, and updates beliefs using Bayes' rule (Kreps and Wilson, 1982, Milgrom and Roberts, 1982, Kreps et al., 1982, Mailath and Samuelson, 2006, p. 460). In the ultimatum game, this would entail a (possibly small) probability of the responder being a committed type, who would always reject low offers. A "normal" responder could then find it worthwhile to mimic the committed type, being willing to suffer the costs of a number of initial rejections because this would cause proposers to believe with sufficient probability that the responder would reject low offers, thus making high offers a best response. While this gives a more concrete role to information, in the steady-state long-run equilibria there is likewise no further flow of information.[14]

Adding some kind of noise (e.g. trembles or observation errors), so that responders occasionally see low offers, could help to test the robustness of the equilibrium and illustrate revelation of information. What would still be miss-

[13] As Akdeniz and Van Veelen (2021, p. 4) put it, "it is somewhat hard to reconcile the idea that people have a hard time differentiating between repeated and one-shot games with the finding that people can and do differentiate rather accurately between repeated games with high and with low probabilities of repetition [in experiments]". See also Fehr and Schmidt (2006, p. 629).

[14] Another issue is that equilibria in these models are foreseeable by both players, so deterrence is typically effective from the first round, hence there is no actual information revelation at any stage; for example Kreps and Wilson (1982, p. 262). Some models may have mixed equilibria though.

ing is an account of how the players arrive at their equilibrium strategies, and whether it is possible for boundedly rational agents to learn to behave in such a way. And if low offers are infrequent and arise only because of noise, is the information they generate sufficient for proposers to stick with their strategies? There is also the problem of equilibrium selection. In addition to the equilibrium in which reputation matters, there is an alternative equilibrium in repeated games, the familiar SPNE, (L,L), played every round. In fact, the folk theorem tells us there is generally a plethora of equilibria in repeated games, including in mixed strategies. Is it possible for responders, in an environment where the norm is that low offers are made and accepted, to improve their situation by building a reputation? Would there be enough information of the right kind, at the right time for high-offer equilibria to be reachable, under various initial conditions? Answering such questions calls for an evolutionary analysis.

4.2.2 Evolutionary models

Evolutionary models can account for the development of heuristic behaviours that are well-adapted to the behaviour of other agents without requiring sophisticated rational reasoning. Such behaviour seems more suited to explain observed behaviour in ultimatum games, where responders appear to be motivated by ingrained emotional reactions when rejecting low offers, rather than rational forward-looking calculations of the effects on their behaviour on their reputations and future payoffs. Another reason to use evolutionary models is that they can give insight in terms of which equilibria can be reached and which are stable, which is especially valuable in situations where multiple plausible equilibria exist without clear reasons to focus on any particular one.

An evolutionary theory about the effects of responder reputations in the UG requires a plausible account for how proposers come to possess credible knowledge about responders. In the existing literature featuring evolutionary models of the UG, rather arbitrary assumptions on the availability of information have been made. In Nowak et al. (2000)'s first model (the minigame), offer amounts are simply reduced by a constant valued parameter when facing a responder who would accept a low offer. Güth and Napel (2006) assume that proposers know responders' utility functions without uncertainty. Akdeniz and Van Veelen (2023, p. 581) assume that, with a fixed probability, proposers know their counterpart's strategy. Interestingly, they claim that, since reputation is not strictly necessary for commitment, and there are other means by which proposers can predict rejecting behaviour of a committed responder,[15] their model is justified as being more general. In Poulsen (2007), the focus is on the value

[15]This includes picking up responders' behavioural cues that accurately reveal emotional commitment, for which there is some empirical evidence. See also Akdeniz and Van Veelen (2021, p. 14) and Frank (1988, p. 169).

of information to proposers. Two models are offered; in the first, there is fixed probability of a responder learning about a responder's strategy, similar to the last model, while in the second model, proposers can pay an exogenously fixed fee to learn the responder's acceptance threshold. This model has endogenous information because information depends on proposers' decisions to buy information, but the information they buy is the responders' MAO strategies directly, rather than their histories of past interactions, so there is still no explicit link between interactions taking place and proposers' available information.[16]

These models demonstrate that credible information, if available, can result in higher offers and demands, but full explanations about where the knowledge comes from, why it is credible, etc., are suggestive rather than integrated into the models. To my knowledge, there are only two exceptions, where the source of information is made explicit. The first is Nowak et al. (2000)'s second model, which uses a computer simulation to track 100 individual agents and the information they gather on accepted proposer offers. Here, the links between past behaviour and information is credible and concrete. Information is endogenous because the probability with which responder strategies can be discovered is linked to the probability that they will receive low offers. However, the description and analysis of results in their paper is very compact and only one type of reputation is modelled. Furthermore, since the model is implemented by agent-based simulation, deterministic results cannot be obtained from it, which hinders full analysis of causal mechanisms, stability and general implications. The second is a model by Zhang et al. (2023), who investigate whether specific (exogenously specified) social norms can lead to the emergence of fairness (equal divisions) in the ultimatum game. Adherence to the social norm gives an agent a good reputation, which can then result in others behaving favourably towards him. One of the norms they consider is that responders should reject low offers, and a reputation based on this norm can only be obtained by responders who receive low offers, which therefore qualifies as endogenous information.[17] However, the focus of their research is not on the value of the information an agent can obtain via their opponent's reputation per se, but rather on the notion of indirect reciprocity, i.e. the use of reputations to effect collective punishment of social-norm violators.[18]

[16]Their results are interesting: if costs are not too high, proposers buy information, which gives responders an incentive to raise their demands, to the ultimate collective detriment of proposers.

[17]They do not refer to the concept of endogenous information in their paper, though it is implicit in the "reputation dynamics" section (p. 3).

[18]In addition to the aforementioned norm for responder behaviour, they consider numerous relatively complicated ("third-order") norms related to the behaviour of proposers. For example, a proposer might gain a bad reputation by failing to punish another proposer who made a high offer instead of a low offer to a responder with a bad reputation. Interestingly, none of the more complicated norms they consider are as effective as the simple responder's norm of rejecting low offers at raising aggregate "fairness levels". There are many other note-

4.3 A general framework for endogenous information

In this section, a theoretical specification for an ultimatum population game[19] with endogenous information is developed, which will form the basis for evolutionary analyses in sections to follow. It will be seen that the framework is general, and could be adapted to other sequential games with two players, the first player's behaviour being contingent on imperfectly observed information about the second player's past behaviour. To avoid unnecessary abstraction, I will use terminology appropriate to the ultimatum game. The framework defines how reputational information is represented, how information flows from a pattern of interactions in the population, and it derives expected payoffs for players with a given pair of strategies in a once-off game, given population strategy frequencies. In subsequent sections, these expected payoffs will then be used as the basis for evolutionary models.

The interactions driving information is to be understood as having occurred in a shared socio-economic environment, so that the first player (the proposer) may have some probability of learning about some past interactions involving the second player (the responder), by direct observation, word-of-mouth or other means. There are two populations, a population of proposers and a population of responders. The populations need not be interpreted as mutually exclusive real-world groups of people – an "individual" in a population merely represents an instance of a specific way of behaving in a specific role; a real person may have multiple roles. The payoff to a player is the expected payoff from an interaction with a randomly selected member of the opposite population, implying that populations are infinitely sized, and lack any structure determining who interacts with whom.

worthy differences in their methods as compared to mine. They do not consider reputational information using generalized formulations (relying instead on exhaustive calculations for all possible combinations of reputations and strategies), they only include two reputational states – an agent is either "good" or "bad" – and past actions are perfectly observable. This means that there is never "no reputation" and their model can therefore not be used to investigate a concern highlighted in this research, namely that the effectiveness of the reputation mechanism may be inhibited by lack of information. Moreover, they model their dynamic system as a Markov process that transitions between states characterized by monomorphic populations (using role-contingent strategies), hence the dynamics they consider are limited to a small number of states and transitions always occur along a single dimension at a time. This would therefore exclude the possibility of many interesting dynamic phenomena such as the mixed-strategy evolutionary equilibrium considered in section 4.5.4.

[19] Sandholm (2015, p. 705) defines population games as models "in which the number of agents is large, each agent is small, and agents are anonymous, with each agent's payoffs depending on his own strategy and the distribution of others' strategies ... One typically imposes further restrictions on the agents' diversity: there are a finite number of populations, and agents in each population are identical, in that they choose from the same set of strategies and have identical payoff functions."

The crux of the endogenous information approach is to recognize that, for given strategy distributions of both populations, there is interdependence between information and actions. The first players' strategies specify actions to take in response to available information on the second player's behaviour pattern (i.e. reputation). The second players' strategies specify actions to take in response to the first players' actions. Conversely, the pattern of actions thus induced determines what information is available. This suggests a dynamic process, where actions and information continually adjust to each other until they converge to a steady state, where the information available to the first players induces exactly the action pattern that generates the same information. This configuration can be called an *endogenous information equilibrium*.

The equations below specify how *prior* information, which are in the form of probability distributions over signal sets for each responder type, induces a pattern of actions, which, if sampled according to the specified procedure, results in the *posterior* information available to proposers. An endogenous information equilibrium occurs when prior and posterior information coincides, so that the system has reached a fixed point. The system of equations implicitly defines information as a function of strategies.

Reaching an endogenous information equilibrium may potentially require many iterations – rounds of play and information updating – before players with a given strategy have built up a stable reputational profile. In this framework, however, this iterative process is considered a short-run process. From an evolutionary time perspective, the process concludes instantaneously. This means that each agent's strategies as well as aggregate strategy frequencies can be regarded as fixed during the process of adjustment to an endogenous information equilibrium.[20] The long-run evolutionary process depends only on average payoffs for given strategies at each point in (evolutionary) time *after* information and actions have converged to an endogenous information equilibrium. In short, populations are large, the number of interactions is high, information adjusts quickly and evolution is slow. Henceforth, any reference to time should be understood as referring to the evolutionary timescale, unless noted otherwise.

4.3.1 Representing information

Proposers have a fixed probability of finding out about any specific act of the responder that occurred within a randomly selected sample of a given size of the responder's previous interactions. To make the nature of information that

[20]This assumption was formulated independently in this research, but was subsequently found to be similar to Zhang et al. (2023, p. 3). Key differences between the two approaches are mentioned in footnote 18. In particular, their information equilibrium is defined as the steady-state distribution of a Markov chain that accommodates only two types of reputation and two strategies (at a time).

the framework generates precise, we first need to specify what can occur in an interaction that can be observed, bearing in mind that full strategies cannot be observed, at least not directly. Proposers choose an action a^P from the proposer action set $A^P \subset \mathbb{R}_{>0}$, which will be the set of possible offers that proposers can make in the ultimatum game.[21] Responders then choose an action a^R from the responder action set $A^R = \{1, 0\}$ where the action 1 indicates acceptance of the offer and 0 indicates rejection. An observer of this interaction would therefore see the action profile (a^P, a^R), which is the basic unit of information that a single interaction can generate. For example, $(4, 1)$ would indicate that an amount of 4 was offered to the responder and the responder accepted the proposal.

Suppose before every interaction, a random sample of $z \geq 0$ past interactions of the responder is drawn. Note that this sample is taken from interactions occurring at the present stage of evolutionary time, so they reflect the current, fixed strategy distributions. For each interaction in the sample, the outcome (a^P, a^R) for that interaction can be observed with a probability of $\alpha : 0 \leq \alpha \leq 1$. This scheme clearly reflects partial observability of past results. It allows a proposer to observe results from multiple past interactions of the proposer, or none at all. If an outcome (a^P, a^R) is successfully observed, it becomes a *signal*. This term "signal" is used here in a generic sense to mean a quantum of information, though the connotation from the signalling literature of a costly action that transmits information about a player's type is not entirely inappropriate.[22]

Given the above, the set of possible signals M is a subset of the Cartesian product of the two players' action sets,

$$M \subseteq A^P \times A^R.$$

This full set of possible signals therefore contains two elements for every possible offer in the game, one for the signal where the amount was accepted and one for where it was rejected. The elements of $A^P \times A^R$ to include in M is a modelling choice, similar to the choice of what offer amounts to include in A^P. It is often desirable not to work with the entire set of possible signals for reasons of analytical convenience and to reduce computational intensity of simulations.

[21] Zero is not allowed as an offer. If zero were included, then the proposer offering zero and any strategy of the responder would be a NE.

[22] To avoid confusion, note the following differences: Firstly, In the signalling literature, "type" usually refers to preferences over action profiles; here it refers to phenotype in the evolutionary game theoretic sense, equivalent to "strategy". Secondly, here, the action does not always succeed in transmitting information to a particular opposite party, it only does so probabilistically. On the other hand, the action's effects may be transmitted to a potentially large number of future bargaining partners. Thirdly, here, all actions available to the responder can result in signals being sent, not only costly ones with beneficial future effects. Finally, in signalling games, the signalling player's choice to send a signal is typically the first move of the game, while here it comes from past interactions (though of course one can conceive of all the interactions together as a multi-player supergame).

Later, I investigate models with negative and positive reputations separately before I consider a model that contains both signal types. Some signals may be of no interest to proposers, or impossible to observe for any responder strategy, e.g. the signal that the highest possible offer was rejected, so excluding such signals would be sensible simplifications.

When a proposer has to choose an offer amount, she will have received a set of signals $q \in Q$, where the set of all possible signal sets Q is a subset of the set of all possible subsets – the *power set* – of M,

$$Q \subseteq \mathcal{P}(M),$$

where $q \in Q$ if there is at least one responder type that can exhibit q. For example, q could be $\{(2,0),(4,1)\}$, indicating knowledge that the responder rejected an offer of 2 and accepted an offer of 4.[23] It is assumed that duplicate signals carry no additional information. For example, if a proposer observes the sequence (No signal), $(4,1)$, $(4,1)$ for a given responder, the signal set from which a decision is to be made is simply $q = \{(4,1)\}$.[24]

Q should include all feasible elements of $\mathcal{P}(M)$, where feasible means at least one responder type is able to generate the combination of signals from random interactions with proposers under any information condition. Not all signals can co-exist in the same signal set, for example $\{(4,1),(4,0)\}$ does not need to be included in Q as it indicates that the responder accepted 4 and also rejected 4, which is impossible unless mixed strategies were allowed (which I do not in this chapter) or strategies could change while being sampled (also not allowed). Furthermore, if responder strategies are monotonic minimum acceptable offer (MAO) strategies, signal sets like $\{(2,1),(4,0)\}$ may also be excluded from Q as there would be no possibility of a proposer having such knowledge. Including infeasible signal sets in Q is not invalid, but can generate a large number of unnecessary proposer strategies (see below) that can complicate analysis and computation. If $\alpha < 1$ there is always the possibility that no signal at all is received, in which case the proposer's knowledge is given by the empty set $\{\}$ which must always be included in Q. The first two minigame models analysed in this chapter only include $\{\}$ and a single signal in Q.

It should be clear by now that each proposer strategy is a mapping from signal sets to offer amounts, $s_i^P : Q \mapsto A^P$. The strategy sets can thus be specified:

$$s_i^P \in S^P \subseteq (A^P)^Q$$

[23] A signal set can be seen to contain exactly one binary bit of information for each possible signal in M – a signal is either included or excluded from a signal set. If there were eight possible signals, e.g. four possible offer amounts, a signal set would contain one byte of information.

[24] This would not be an appropriate assumption if mixed strategies were allowed, in which case multiple observations of the same outcome could be interpreted statistically to estimate the responder's mixed strategy. Note also how important it is for the responder's strategy to remain constant while being sampled – without this, proposers would have a considerably harder task interpreting the data.

$$s_j^R \in S^R \subseteq A^P$$

where i and j are index variables enumerating the strategy sets and $(A^P)^Q$ contains all possible (q, a^P) pairs. S^P and S^R are shown as subsets because the strategies to include in a specific model are also modelling choices and it may be desirable to work with restricted strategy sets. $s_i^P(q)$ then returns the offer that proposer strategy s_i^P makes to a responder given signal set q. Responder strategies are simply an amount, interpreted as a minimum acceptable offer (MAO) and drawn from the same set as proposers' possible offers. It would obviously also have been possible to specify responder strategies in a more general way by dropping the monotonicity assumption. The responder strategy set could then be specified as the power set of offer amounts, with each strategy being the set of offers to accept. This is not necessary for the objectives of this chapter, but could be explored in future work.

4.3.2 What information is generated?

In order to eventually determine expected payoffs for both players, given strategies, we will first need to calculate, for each responder type j, the probability distribution over signal sets that results from sampling z past interactions of that responder type, with each sampled interaction's outcome being observed by the current proposer with a probability of α.[25] Let $y_j(q)$ be the probability with which a proposer's knowledge about a responder of type j is constituted by the signal set q. Recalling that the empty signal set {} is included in Q, we must have,

$$\sum_{q \in Q} y_j(q) = 1.$$

We know that the responder can get information on a sample of z interactions of the responder. If $y_j^n(q)$ is defined as the probability of observing signal set q from n samples of a responder with strategy j's interactions, then $y_j(q) = y_j^z(q)$, and $y_j^1(q)$ is the probability of observing signal set q from a single sample. From a single interaction, the only signal sets that can be observed are those containing a single, or zero signals. This leads to

$$y_j^1(\{m\}) = y_j^1(m) \tag{4.1}$$

$$y_j^1(\{\}) = 1 - \sum_{m \in M} y_j^1(m) \tag{4.2}$$

[25]A proposer would in principle be sampling from the history of the specific responder he is interacting with, but the average probability of any signal from a random responder of a given type would be the same as that generated for the type overall. One could think of each responder strategy being represented by a single individual, with sampling taking place from that individual's interaction history once there has been enough interactions for the statistical distribution over outcomes in it to converge.

This indicates that the probability of observing a signal set containing a single signal m in a single interaction is equal to the probability of observing that signal in a single interaction,[26] and that the probability of observing an empty signal set in a single interaction is the complement of the combined probability of observing one of the signals in a single interaction's outcome. The probability of observing, in a single interaction, any signal sets containing more than one signal is zero, but this case will be dealt with using more general formulae below.

In a single interaction, a given responder with strategy s_j^R will be paired with a random proposer and the proposer will see a random signal set, so expectations need to be taken along two dimensions to calculate $y_j^1(m)$:

$$y_j^1(m) := \alpha \sum_i \sum_{q \in Q} x_i^P y_j(q) I(\psi(s_i^P, s_j^R, q) = m) \tag{4.3}$$

where x_i^P is the relative frequency of proposer strategy s_i^P (with $\sum_i x_i^P = 1$), $I()$ is the indicator function that returns 1 if the condition is true and 0 if it is false, and

$$\psi(s_i^P, s_j^R, q) := (s_i^P(q), s_j^R(s_i^P(q))) \tag{4.4}$$

returns a tuple that reflects the outcome from an interaction between a responder with strategy s_j^R and a proposer with strategy s_i^P observing signal set q. The first component of the tuple is the amount offered by the proposer and the second component is 1 or 0 indicating acceptance or rejection. Since mixed strategies are not allowed, the outcome is fully deterministic for a given (s_i^P, s_j^R, q). Equation (4.3) adds up the weights of all possible combinations of s_i^P and q that results in outcome m for a s_j^R-responder, and multiplies this with the probability that the outcome will turn into an observed signal, α. The $y_j(q)$ on the RHS of the equation can be regarded as reflecting prior information, which the function transforms into posterior information in the form of the $y_j^1(m)$ on the LHS. $y_j(q)$ itself will be defined as a function of $y_j^1(m)$ below; together, the equations require the prior and posterior to be equal, which defines an endogenous information equilibrium.

It is worth emphasizing again that the endogenous information framework being developed here does not depend on the ultimatum game's payoff structure or strategy sets, so other two-player sequential-move games with similar information structure can be substituted.

[26]To understand the distinction, note that it is *not* generally true, for example, that $y_j^2(\{m\}) = y_j^2(m)$. If a sample of two interactions were taken, the probability that m will be observed at least once in one of the two interactions, $y_j^2(m)$, is not the same as the probability that m will be the *only* signal observed in the sample, $y_j^2(\{m\})$.

We now know from equations (4.1) to (4.4) how to calculate the probability of any signal set being observed from a sample of responder's interactions when the sample size is one. In order to calculate $y_j(q)$, we must be able to calculate $y_j^z(q)$ for any sample size $z \geq 1$ and any $q \in Q$. This is an exercise in probability calculus. Consider a generic example with four possible signals. The probability of observing signal set $q = \{m_1, m_2\}$ is the joint probability of observing at least one m_1, observing at least one m_2 and not observing any other signals (e.g. m_3 and m_4) among the z samples taken:

$$\Pr[q = \{m_1, m_2\}] = \Pr[m_1 \cap m_2 \cap \neg m_3 \cap \neg m_4]$$

where $\Pr[m]$ should be read as the probability of *at least one* instance of m occurring in the sample of size z and $\Pr[\neg m]$ should be read as the probability of *zero* instances of m occurring in the sample of size z. The probabilities of observing no instances of a signal, or of observing at least one (the latter's complement) are straightforward,[27] e.g.

$$\Pr[\neg m_3] = (1 - y_j^1(m_3))^z \tag{4.5}$$

$$\Pr[m_1] = 1 - (1 - y_j^1(m_1))^z \tag{4.6}$$

The probabilities of neither of, or at least one of, two signals are also simple,[28] respectively,

$$\Pr\left[\neg m_3 \cap \neg m_4\right] = \Pr[\neg(m_3 \cup m_4)] = (1 - (y_j^1(m_3) + y_j^1(m_4)))^z \tag{4.7}$$

$$\Pr[m_1 \cup m_2] = 1 - (1 - (y_j^1(m_1) + y_j^1(m_2)))^z \tag{4.8}$$

These formulae extend easily for any number of signals. Define

$$K := 1 - y_j^1(m_3) - y_j^1(m_4)$$

as the probability of signals excluded from the signal set *not* occurring in a single sample,[29] where $y_j^1(\,)$ is as defined in equation (4.3). We can now make use of the multiplication rule and the inclusion-exclusion principle:

$$\Pr[q = \{m_1, m_2\}] = \Pr[m_1 \cap m_2 \cap \neg m_3 \cap \neg m_4]$$
$$= \Pr[m_1 \cap m_2 | \neg m_3 \cap \neg m_4] \Pr[\neg m_3 \cap \neg m_4]$$

[27] The probability of a signal not occurring in one interaction is independent of the probability of the signal not occurring in another interaction, so the multiplication rule for independent events can be used.

[28] In a single sample any two signals are mutually exclusive so their probabilities can be added to get the probability of "either of them", and the probability of "neither of them" is again the complement. As with one signal, the probability of a group of signals not occurring in an interaction is independent of the probability of it not occurring in another interaction.

[29] A more general definition for K will be provided below.

$$
\begin{aligned}
&= [Pr[m_1|\neg m_3 \cap \neg m_4] + Pr[m_2|\neg m_3 \cap \neg m_4] - Pr[m_1 \cup m_2|\neg m_3 \cap \neg m_4]] \\
&\qquad \times \Pr[\neg m_3 \cap \neg m_4] \\
&= \Bigg[\left(1 - \left(1 - \frac{y_j^1(m_1)}{K}\right)^z\right) + \left(1 - \left(1 - \frac{y_j^1(m_2)}{K}\right)^z\right) \\
&\qquad\qquad - \left(1 - \left(1 - \frac{y_j^1(m_1) + y_j^1(m_2)}{K}\right)^z\right)\Bigg] K^z \\
&= \left(K^z - (K - y_j^1(m_1))^z\right) + \left(K^z - (K - y_j^1(m_2))^z\right) \\
&\qquad\qquad - \left(K^z - (K - y_j^1(m_1) - y_j^1(m_2))^z\right) \qquad (4.9) \\
&= \Pr[m_1 \cap \neg m_3 \cap \neg m_4] + \Pr[m_2 \cap \neg m_3 \cap \neg m_4] \\
&\qquad\qquad - \Pr[(m_1 \cup m_2) \cap \neg m_3 \cap \neg m_4] \qquad (4.10)
\end{aligned}
$$

The solution is in the second-last line (4.9); the last line has been added to illustrate the logic of the inclusion-exclusion principle: to find the probability of the intersection of two overlapping sets, one can subtract the probability of their union from the sum of their individual probabilities. Figure 4.3.1 illustrates the principle for the calculation of $\Pr[m_1 \cap m_2 \cap \neg m_3 \cap \neg m_4]$.

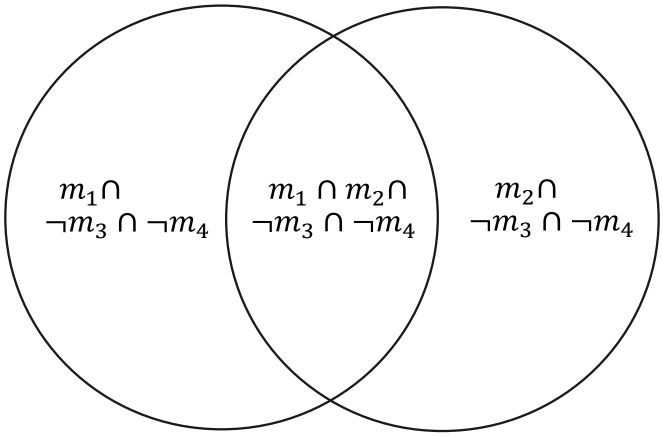


Figure 4.3.1: The inclusion-exclusion principle for a signal set with two signals

These methods can be extended to calculate the probability of signal sets with any number of signals, using applicable versions of the inclusion-exclusion principle. Notice in equation (4.10) that the last term relates to the probability of the union of all of the signals in q, jointly with the non-occurrence of all signals not in q. This is the only term that relates to all of the included signals, while all of the the other terms relate to the occurrence of subsets of the included signals (in the example, $\{m_1\}$ and $\{m_2\}$). This pattern holds generally and allows a recursive solution to the general problem.

Before stating the general solution, the convenience symbol K in the example needs to be generalized. Define,

$$K_j(E) := 1 - \sum_{m \in E} y_j^1(m) \tag{4.11}$$

which gives the probability of no signal in E occurring in a single interaction. The general solution for $y_j(q)$ then relies on the following recursive function:

$$\begin{aligned}\Upsilon_j(I,E) := (-1)^{(|I|+1)} &\left(\left(K_j(E)\right)^z - \left(K_j(E) - \sum_{m \in I} y_j^1(m)\right)^z \right) \\ &+ \sum_{g=1}^{|I|-1} \sum_{A \in \binom{I}{g}} (-1)^{(|I|+g+1)} \Upsilon_j(A,E)\end{aligned} \tag{4.12}$$

The function takes two sets as arguments, namely signals to include (I) and signals to exclude (E), and returns the joint probability of the I-signals occurring and the E-signals not occurring in a sample of z interactions of responder strategy j. The first term captures the probability of the union of all of the I-signals occurring, jointly with the non-occurrence of all the E-signals. According to the inclusion-exclusion principle, the sign for this term should be either positive or negative, depending on whether there is an odd or equal number of elements in I, hence a power of -1 factor is added. The rest of the terms needed for the inclusion-exclusion procedure are generated by iteration and recursion. For each positive number g of signals smaller than the number in I, the joint probability of all g-combinations of I occurring and E not occurring in a z-sample is added or subtracted, the sign alternating as g increases. The formula works correctly for any positive number of signals in I, including one, in which case the iteration and recursion fall away and the first term reduces to the joint probability of that signal occurring and E-signals not occurring. In cases where $|I| > z$, the formula correctly returns zero as it is impossible to see more signals than the sample size.

To see how the formula works, consider an example with three signals included and one excluded:

$$\begin{aligned}&\Upsilon_j(\{m_1, m_2, m_3\}, \{m_4\}) \\ &= (-1)^4 \left(\left(1 - y_j^1(m_4)\right)^z - \left(1 - y_j^1(m_4) - y_j^1(m_1) - y_j^1(m_2) - y_j^1(m_3)\right)^z \right) \\ &\quad + \sum_{A \in \binom{I}{1}} (-1)^5 \Upsilon_j(A, \{m_4\})\end{aligned}$$

$$
\begin{aligned}
&+ \sum_{A \in \binom{I}{2}} (-1)^6 \Upsilon_j(A, \{m_4\}) \\
&= \left(1 - \left(1 - \frac{y_j^1(m_1) + y_j^1(m_2) + y_j^1(m_3)}{1 - y_j^1(m_4)}\right)^z\right)\left(1 - y_j^1(m_4)\right)^z \\
&\quad - \Upsilon_j(\{m_1\}, \{m_4\}) - \Upsilon_j(\{m_2\}, \{m_4\}) - \Upsilon_j(\{m_3\}, \{m_4\}) \\
&\quad + \Upsilon_j(\{m_1, m_2\}, \{m_4\}) + \Upsilon_j(\{m_1, m_3\}, \{m_4\}) + \Upsilon_j(\{m_2, m_3\}, \{m_4\}) \\
&= \Pr[(m_1 \cup m_2 \cup m_3) \cap \neg m_4] \\
&\quad - \Pr[m_1 \cap \neg m_4] - \Pr[m_2 \cap \neg m_4] - \Pr[m_3 \cap \neg m_4] \\
&\quad + \Pr[m_1 \cap m_2 \cap \neg m_4] + \Pr[m_1 \cap m_3 \cap \neg m_4] + \Pr[m_2 \cap m_3 \cap \neg m_4]
\end{aligned}
$$

It can be seen in figure 4.3.2 (in which it can be assumed that only the parts of the sets that are not in m_4 are shown) that this is the correct application of the inclusion-exclusion principle for three sets: start with the union of all three sets, deduct the three circles, add back the three almond shaped two-way intersections and what remains is the Reuleaux-triangle-shaped, three-way intersection in the middle that we wanted to calculate.

The Υ function can be used directly to calculate the probability of every signal set $q \in Q$, with one exception: the empty signal set $\{\}$. The exception is due to the fact that the inclusion-exclusion principle is not clearly defined for zero sets and the formula's first term (which returns zero if $I = \{\}$) has no sensible meaning in such a case. Therefore,

$$
y_j(q) := \begin{cases} (K_j(M))^z, & \text{if } q = \{\} \\ \Upsilon_j(q, M \setminus q), & \text{otherwise.} \end{cases} \tag{4.13}
$$

This defines an implicit function for $y_j(q)$ as both $K_j()$ and $\Upsilon_j()$ depend on $y_j^1(m)$ which in turn depends on $y_j(q)$ – see equation (4.3). Analytical solutions for the system of equations including (4.3) and (4.12) may be feasible for simple instances, while numerical solutions may be obtained more generally. The computer simulations for the evolutionary models that follow in this chapter embed these functions and use an iterative procedure to yield solutions at each point in time as they are needed to calculate expected payoffs for proposers and responders when information is endogenous (see section 4.5.2).

4.3.3 Expected payoffs

Having characterized information produced by random interactions, all that remains to be done is to specify expected payoffs given strategy frequencies. Start

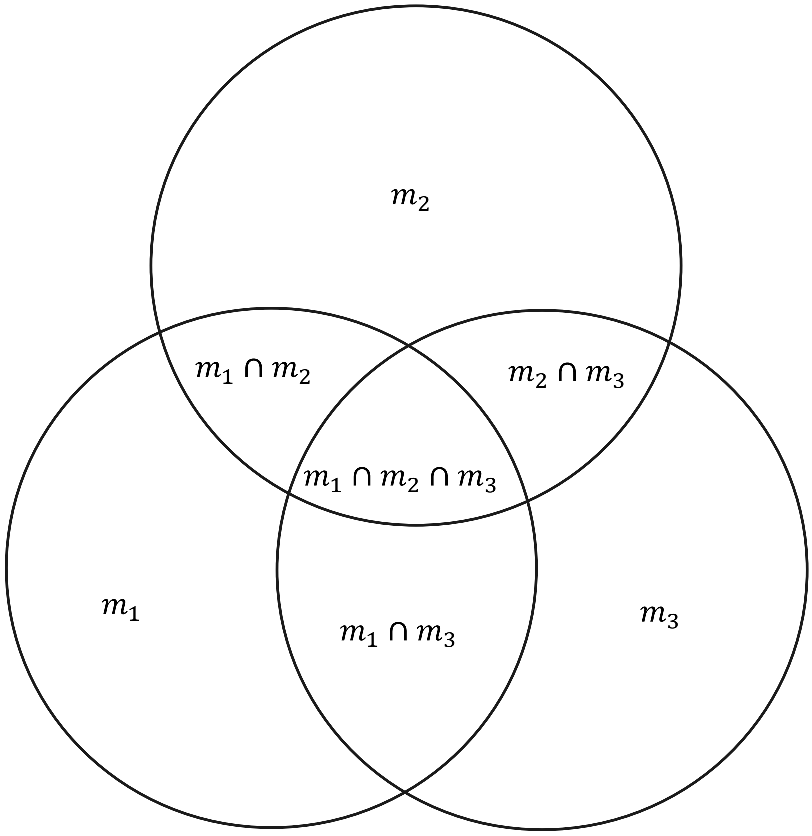


Figure 4.3.2: The inclusion-exclusion principle for a signal set with three signals

with the assumption that we already know how to relate action profiles to payoffs, i.e. we know $\pi^P(a^P, a^R)$ and $\pi^R(a^P, a^R)$. For the ultimatum game, if the full amount being divided is \$, payoffs are $(\$ - a^P, a^P)$ if $a^R = 1$ and $(0, 0)$ if $a^R = 0$. To get expected payoffs as a function of strategies, it is necessary to take expectations over information:

$$\begin{aligned}\pi^P(s_i^P, s_j^R) &:= \sum_{q \in Q} y_j(q) \pi^P(s_i^P(q), s_j^R(s_i^P(q))) \\ \pi^R(s_i^P, s_j^R) &:= \sum_{q \in Q} y_j(q) \pi^R(s_i^P(q), s_j^R(s_i^P(q)))\end{aligned} \tag{4.14}$$

Note that information available to a proposer does not depend on the proposer's strategy, but does depend the frequency distribution over all proposer strategies through (4.3) – this is the sense in which information is endogenous in this framework. Having completed the framework, specific models can be generated by it by selecting different money amounts, \$, sets of possible offers, A^P, possibles signals M and strategy sets S^P and S^R.

4.4 A simple minigame with endogenous information

In this section, we consider a minigame similar to Nowak et al. (2000)'s first model, using the aforegoing section's endogenous information framework. For concreteness, I will use the values from table 4.2.1, i.e. a total money amount to be divided of 4, and possible offers 1 (L) and 2 (H). This leads to four possible signals and $2^4 = 16$ possible signal sets, though for this simple example I will include in M only the one signal considered in Nowak et al. (2000)'s paper, namely the one indicating that the responder accepted a low offer, i.e. a negative reputation. Q contains only the empty signal set and one signal set containing the above-mentioned signal.

This leads to four possible proposer strategies (i.e. mappings from Q to A^P). To summarize,

$$\begin{aligned}
\$ &= 4\\
A^P &= \{2, 1\}\\
M &= \{(1,1)\}\\
Q &= \{\{\}, \{(1,1)\}\}\\
S^P &= \{s_1^P, s_2^P, s_3^P, s_4^P\}
\end{aligned}$$

where

$$s_1^P = \begin{cases} 2, & \text{if } q = \{\} \\ 2, & \text{if } q = \{(1,1)\} \end{cases} \quad (\text{'HH'}) \qquad s_2^P = \begin{cases} 1, & \text{if } q = \{\} \\ 1, & \text{if } q = \{(1,1)\} \end{cases} \quad (\text{'LL'})$$

$$s_3^P = \begin{cases} 2, & \text{if } q = \{\} \\ 1, & \text{if } q = \{(1,1)\} \end{cases} \quad (\text{'HL'}) \qquad s_4^P = \begin{cases} 1, & \text{if } q = \{\} \\ 2, & \text{if } q = \{(1,1)\} \end{cases} \quad (\text{'LH'})$$

The first two strategies ignore the signal and offer 2 (H) and 1 (L) unconditionally, respectively. The third strategy is the sophisticated strategy of offering H, except if it is known that the responder accepted L, in which case, offer L. This is the strategy that corresponds to Nowak et al. (2000)'s proposer strategy of offering $h-a$, i.e. the normal high offer "shaded" by deducting a for responders that accept low offers. The difference is that Nowak et al. (2000) condition on the responder strategy while s_3^P conditions on available information. The last proposer strategy is a "Robin Hood" strategy of offering H to those who would accept L and L to those that demand H (which would then be rejected).[30] The game can be represented by table 4.4.1.

[30]The four proposer strategies identified here do not correspond to the four proposer strategies in Nowak et al. (2000)'s minigame. The latter uses a symmetrized game in which a player can take on both proposer and responder roles randomly, so strategies need to be role contingent. They do not include the strategies that I label HH and LH, which I also discard below.

			Responder	
			H	L
Proposer	HH		2, 2	2, 2
	LL		0, 0	3, 1
	HL	if q={}	2, 2	2, 2
		if q={(1,1)}		3, 1
	LH	if q={}	0, 0	3, 1
		if q={(1,1)}		2, 2

Table 4.4.1: Strategic-form representation of an ultimatum minigame with negative reputations

Endogenous information is evident: the payoffs for the conditional strategies depend on whether the (1, 1) signal (indicating that the responder accepted a low offer) was received or not. Since an H-responder can never emit this signal, the relevant cells of the table have been left blank. As I will show, expected payoffs for the various strategies can be calculated using equation (4.14).

For illustrative purposes, the game can be simplified by eliminating two weakly dominated proposer strategies.[31] HH is weakly dominated by HL because a proposer who makes low offers only to responders who have a reputation of accepting them gains 1 if the signal is received and never risks rejection. LH ("Robin Hood") is weakly dominated by LL because it responds to information that a low offer will be acceptable by changing the offer to H. The simplified game is illustrated in table 4.4.2. Note for now that LL can be better than HL for the proposer if the probability that the responder demands L is high *and* the probability of the signal being received is low.

			Responder	
			H	L
Proposer	HL	if q={}	2, 2	2, 2
		if q={(1,1)}		3, 1
	LL		0, 0	3, 1

Table 4.4.2: Strategic-form representation of an ultimatum minigame with negative reputations (simplified)

We can see that (HL,H) and (LL,L) are both NE regardless of whether the signal is received or not, but stability of both equilibria depends on the state of

[31] I include these strategies again in simulations (section 4.5.3), to see what effects their presence has.

available information. I will apply the definitions for static evolutionary stable strategies (ESS) for two groups in Cressman (1995, p. 241), according to which an equilibrium is an ESS, if, from any point in its neighbourhood, at least one player would get a higher payoff if it switches back to its equilibrium strategy. (HL,H) would be an ESS if the signal were present, but not if the signal were absent, because then the equilibrium could be invaded by L-responder mutants, who would effectively be undetected if the signal were absent and they would therefore still get H offers. In contrast, (LL,L) – the UG's SPNE – would *not* be an ESS if the signal were present, because HL-proposers could then invade, but it would be an ESS if the signal were absent, as in the standard ultimatum game without reputations.

Clearly, it is important to take information into account when analysing stability. We will therefore need to calculate $y_L((1,1))$, the probability that a proposer matched with an L-responder will receive the signal indicating that the responder has accepted a low offer.

4.4.1 The case where $z = 1$

Since there are no signal sets with more than one signal in this example, it seems we should be able to simplify by assuming that the proposer only samples a single past interaction of the responder, so the first case to be considered is where $z = 1$. This collapses equations (4.14) and (4.12) to simply $y_L(\{(1,1)\}) = y_L^1((1,1))$. Call this value y^* for convenience. It is the probability with which a proposer will know about a past interaction of the responder in which the responder accepted a low offer (i.e. a negative reputation), for a responder whose strategy is to accept low offers. Since there are just two possible signal sets, the empty signal set's probability is just the probability of not getting the signal, $y_L(\{\}) = 1 - y^*$. Using equation (4.3), we then see that,

$$\begin{aligned} y^* &= \alpha\left[x^P_{HL}((1-y^*)0 + y^*1) + x^P_{LL}((1-y^*)1 + y^*1)\right] \qquad (4.15)\\ &= \alpha\left[x^P_{HL}y^* + x^P_{LL}\right],\\ &= \alpha\left[x^P_{HL}y^* + (1-x^P_{HL})\right],\end{aligned}$$

therefore,

$$y^* = \frac{\alpha(1-x^P_{HL})}{1-\alpha x^P_{HL}}. \qquad (4.16)$$

The first line (4.15) shows that outcome of a low offer being accepted by the L-responder can result from three distinct possibilities: A HL-proposer seeing the signal (thus making a low offer), a LL-proposer not seeing the signal and a LL-proposer seeing the signal. If an interaction with this outcome is sampled by a proposer, the proposer will see the signal with probability α.

Equation (4.16) indicates that if there are only LL-proposers, i.e. if $x^P_{HL} = 0$, every interaction will result in a low offer getting accepted and y^*'s value becomes α. Furthermore, y^* approaches zero as α approaches zero and approaches one as α approaches one, except when $x^P_{HL} = 1$, in which case $y^* = 0$ for any $\alpha < 1$. This indicates it is impossible to receive the signal if $x^P_{HL} = 1$ and $\alpha < 1$.

But there is a special case: when $\alpha = 1$ and $x^P_{HL} = 1$, the expression in (4.16) is undefined. The problem is division by zero in the last step. If we backtrack and take $x^P_{HL} = 1$, this gives $y^* = \alpha y^*$, indicating that y^* can take any value if $\alpha = 1$. This means that, given any value y^*, the information generated by interactions in which proposers see the signal with that probability is exactly y^* once again. We get out exactly what information we put in. There is some intuitive sense to this: if the fraction of L-responders arriving at interactions with the signal is y^*, then HL-proposers will make low offers a fraction y^* of the time exactly, thus leading to the signal being observed $\alpha y^* = y^*$ times per interaction on average if $\alpha = 1$. Nevertheless, this demonstrates that the framework can in some cases have indeterminate outcomes, which is generally problematic. I discuss this issue and suggested solutions below. If $\alpha < 1$, the information is attenuated and the only possible solution is $y^* = 0$, thus indeterminacy is avoided, but this is specific to the case where $z = 1$ (see below for $z > 1$).

For now, assume $0 < \alpha < 1$ and use (4.16) with (4.14) to calculate expected payoffs for both players when a HL-proposer interacts with an L-responder,

$$\begin{aligned}\pi^P(\mathrm{HL},\mathrm{L}) &= \frac{\alpha(1-x^P_{HL})}{1-\alpha x^P_{HL}}3 + \left(1 - \frac{\alpha(1-x^P_{HL})}{1-\alpha x^P_{HL}}\right)2 \\ &= \frac{2+\alpha-3\alpha x^P_{HL}}{1-\alpha x^P_{HL}},\end{aligned} \tag{4.17}$$

and,

$$\begin{aligned}\pi^R(\mathrm{HL},\mathrm{L}) &= \frac{\alpha(1-x^P_{HL})}{1-\alpha x^P_{HL}}1 + \left(1 - \frac{\alpha(1-x^P_{HL})}{1-\alpha x^P_{HL}}\right)2 \\ &= \frac{2-\alpha-\alpha x^P_{HL}}{1-\alpha x^P_{HL}}.\end{aligned} \tag{4.18}$$

Table 4.4.3 incorporates these expressions and can be regarded as reflecting a game with the endogenous information aspect fully solved. Remember that x^P_{HL} reflects the frequency of the HL strategy in the population and should not be confused with the proposer's strategy in a particular interaction. The expected payoffs to both HL-proposers and L-responders (but not LL-proposers or H-responders) depend not only on the strategy of their opponents in a particular game but also on the population frequency of HL-proposers, which means this game features own-population effects for proposers and nonlinear state dependence (Friedman, 1998, p. 22–23).

		Responder	
		H	L
Proposer	HL	2, 2	$\frac{2+\alpha-3\alpha x_{HL}^{P}}{1-\alpha x_{HL}^{P}}, \frac{2-\alpha-\alpha x_{HL}^{P}}{1-\alpha x_{HL}^{P}}$
	LL	0, 0	3, 1

Table 4.4.3: Strategic-form representation of an ultimatum minigame (solved)

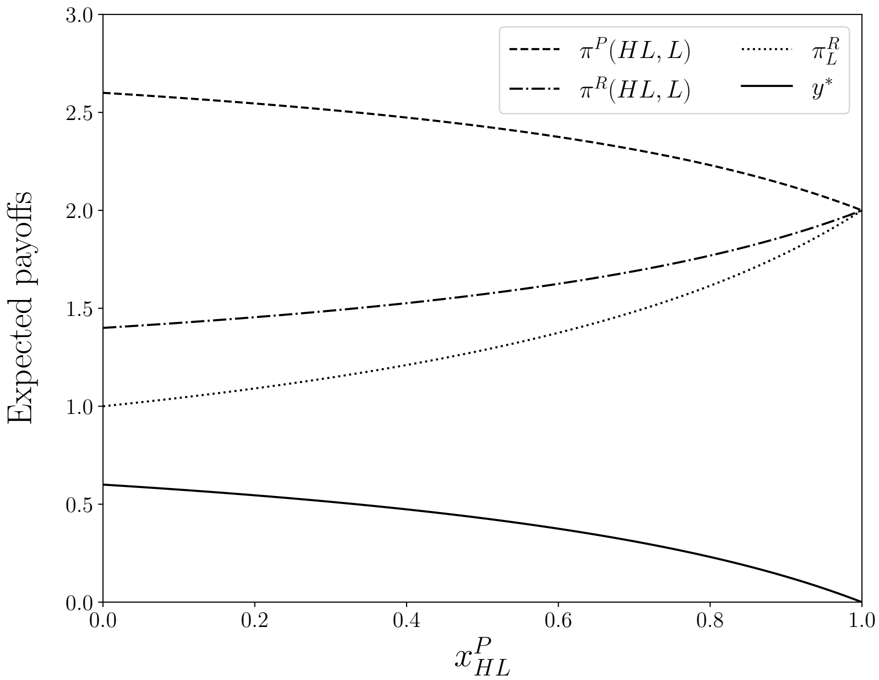


Figure 4.4.1: Expected payoffs for HL-Proposers and L-Responders with endogenous information ($\alpha = 0.6$)

Figure 4.4.1 shows expected payoffs for HL-proposers and L-responders as x_{HL}^P is varied from zero to one with $\alpha = 0.6$. A higher value for α shows the nonlinearity of the curves more clearly. The expected payoff for a HL-proposer proposing to an L-responder decreases from 2.6 to 2. An L-responder (who never rejects any offers) gets an expected payoff of 4 minus the HL-proposer's expected payoff, which increases from 1.4 to 2. These effects are caused by the decrease in the probability with which the proposer receives the signal (and thus knows that the responder accepts L and thus reduces the offer accordingly), which decreases from α to zero. On the left-hand-side of the graph, HL-proposers are benefiting from the abundance of LL-proposers in the population who make L offers to the L-responders, hence there is a relatively high probability that the signal will be generated. But on the right-hand-side information is relatively scarce so HL-proposers end up making more H offers, which is to their detriment and to the benefit of the L-responders. Also shown on the graph is π_L^R, the average payoff to L-responders against a random member of the proposer population, which is relatively low on the left-hand-side because of the high frequency of LL-proposers they interact with, who always offers L, not just a fraction α of the time as the HL-proposers do.

The analysis helps to settle the question of stability for $\alpha < 1$ and $z = 1$. On the left of the graph, where $x_{HL}^P = 0$, the expected payoff of a HL-proposer, $\alpha 3 + (1-\alpha)2 = 2+\alpha = 2.6$ is strictly less than the payoff of a LL-proposer, 3, against a population of all L-responders, which makes (LL,L) an ESS. On the right, all responders get an expected payoff of 2, as there is no information to distinguish them, so HL-proposers offer H to both H- and L-responders. Therefore (HL,H) is not an ESS, contrary to Nowak et al. (2000)'s finding that both high and low offer equilibria are stable in their minigame. The difference is precisely due to their minigame analysis not taking into account endogenous information. However, is could be argued that ESS stability gives a limited (static) view of stability, and I will show below (section 4.5.3) that there can indeed be high-offer stable evolutionary equilibria if minor modifications are made.

4.4.2 The case where $z > 1$ and the general problem of information indeterminacy

It would seem from the previous analysis that the problem of information indeterminacy can be eliminated by setting $\alpha < 1$, but unfortunately this is not generally the case. If proposers can gain information from more than one past interaction of a responder, i.e. if the sample size $z > 1$, a small probability α of observing a signal in a single interaction can easily turn into a large probability of the responder's behaviour being detected by a proposer.

Instead of analysing the full minigame again, the focus here is specifically on the case where $x_{HL}^P = 1$ and an interaction takes place between a HL-proposer and an L-responder. Using (4.13) and (4.3), we can calculate the probability of the $(1,1)$ signal being received as,

$$\begin{aligned} y^* &= 1 - \left(1 - \alpha\left(x_{HL}^P y^* + (1 - x_{HL}^P)\right)\right)^z . \\ &= 1 - (1 - \alpha y^*)^z \end{aligned} \tag{4.19}$$

which is still an implicit function. Figure 4.4.2 shows plots of "output" (LHS) values of y^* against "input" (RHS) values for y^* for different α and z parameters. A solution of (4.19) is indicated by the curve crossing the dotted 45° line. Recall that these are values for y^* that induce distributions of action profiles, which, if sampled according to the specified sampling procedure, result in the same information again.

From the bottom, the ($\alpha = 0.3, z = 1$) curve is linear and produces only one solution at $y^* = 0$ – the solution I worked with above to claim that (HL,H) was not an ESS. If $\alpha = 1$ and $z = 1$, the linear curve would lie exactly on top of the 45° line, explaining the complete indeterminacy found in that case. The next curve, ($\alpha = 0.4, z = 2$), is nonlinear, but the sampling parameters are still too weak for negative reputations to survive – the only solution is again at $y^* = 0$ and the implications for stability of the equilibria would be similar to the above. The three other curves all cross the 45° line twice, at $y^* = 0$ and again at some positive value. This can either be achieved by increasing α or by increasing z. Increasing the value of z increases nonlinearity but does not alter the number of crossing points within the boundaries considered.

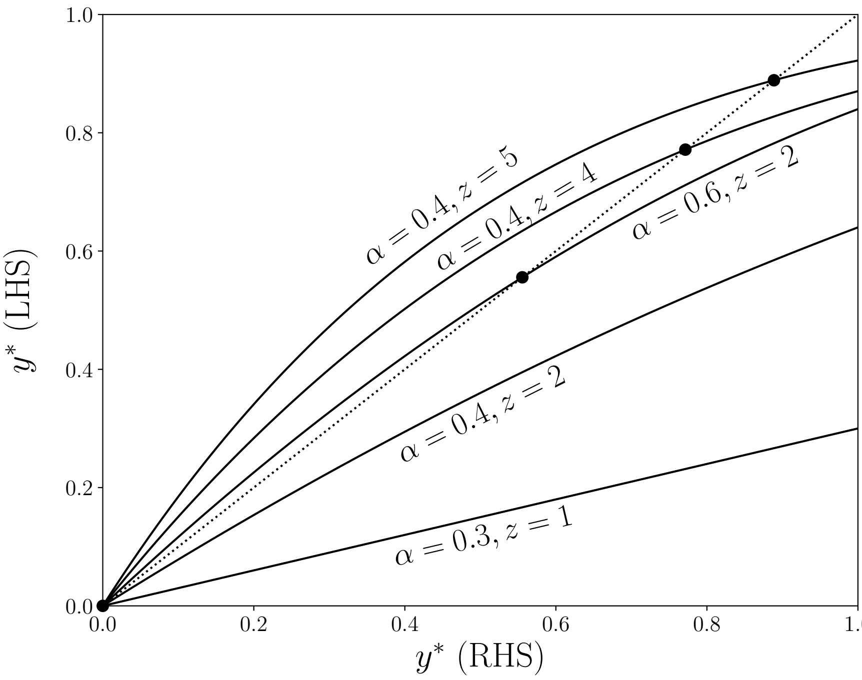


Figure 4.4.2: Information function solutions ($x^P_{HL} = 1$)

The curves that cross the 45° line twice indicates that there are multiple solutions for those parameter values, and we now see that setting $\alpha < 1$ does not always prevent this. This is because setting $z > 1$ boosts the probability with which a signal is observed in a sample beyond the frequency with which it can be observed in a single interaction, so the information does not attenuate to zero as above with $\alpha < 1$ and $z = 1$. Thus we see multiple solutions are common rather than just a special case due to extreme parameter values.

This indeterminacy should not be considered a nonsensical result. Consider two L-responders: one with an existing negative reputation and one without, entering an environment filled with HL-proposers. Under the right conditions, the responder with the negative reputation gets offered low amounts, which he accepts, thus his negative reputation is reproduced, while the responder about whom nothing is known gets offered only high amounts, thus she is never observed to accept a low offer and her reputation remains intact.

Yet indeterminacy introduces three problems in analysing the UG. Firstly, the information framework only accommodates one signal set probability distribution for each responder type, effectively making it a representative agent model. There cannot be two agents of the same type, one with a clean reputation and one without. It is not hard to imagine that in some situations a permanently clean reputation may be a matter of individual luck, but the use of expected values in the framework removes the possibility of such stochastic effects. This links to the second problem, namely that a particular result, even a relatively stable one, may not be robust against large unmodelled stochastic effects if there are multiple endogenous information equilibria. Perhaps a responder with a negative reputation can be lucky enough to avoid low offers for long enough that the negative reputation is forgotten.

Finally, a problem of a more practical nature is that, if there are multiple endogenous information equilibria, the prior values one uses in (4.3) will determine the result, and it is hard to think of a strong *general* principle by which to choose prior values. Conceptually the information is supposed to come from the interactions between proposers and responders at a point in evolutionary time. Should everyone start with a clean slate? This is debatable. It can be even less clear when we consider that reputations in more complicated scenarios can consist of bundles of signals of various kinds.

Fortunately, there is a reasonable way to avoid the problem altogether. If one is concerned about realistic unmodelled stochastic effects causing a responder to permanently lose a negative reputation, then it could be argued that it also makes sense to be concerned about the realistic possibility of the opposite, i.e. an L-responder with a clean reputation having the bad luck to receive a rare low offer, thus gaining a permanent negative reputation. We must therefore ensure that such possibilities exist. If a signal type is included in the analysis, there should be a way for this signal to become part of an agent's reputation even if it is not there initially.

In this minigame, this means it must be ensured that there is always some possibility of low offers being received. This is easily achieved by adding a small mutation rate to the evolutionary model, which I do in the next section. This effectively ensures that $x_{HL}^{P} < 1$ at all times, i.e. there is always at least a positive (though possibly very small) frequency of LL-proposers in the environment, who can act as the source of occasional low offers to responders with clean reputations. It is interesting to note that the only other model with endogenous information, Nowak et al. (2000)'s second (agent-based simulation) model, includes a small exogenous probability of H-proposers making low offers – possibly this was also added partly to "seed" negative reputations among L-responders to avoid zero-information outcomes.[32]

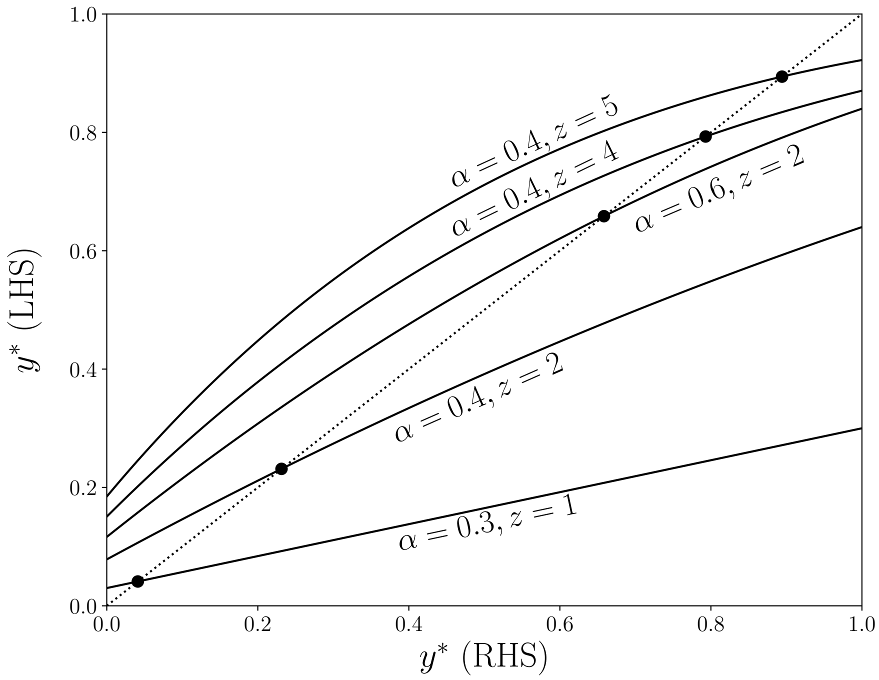


Figure 4.4.3: Information function solutions ($x^P_{HL} = 0.9$)

Figure 4.4.3 shows the effect of the addition of LL-proposers, whose frequency is set to 0.1 for illustration. A clean reputational slate offers no defence against LL-proposers, who ignore reputations and make low offers indiscriminately. This can be seen clearly on the left-hand side of the graph, where positive probabilities of observing the signal being emitted result from interactions in which no signals have been received by proposers. Once L-responders have a small positive probability of being detected, the probability increases even further because now interactions with both types of proposers can result in the signal being emitted.

For all of sampling parameter configurations, there is now exactly one value of y^* that solves the information equation, so the problems associated with indeterminacy are gone. There is always a positive probability of observing a negative reputation for all L-responders. As I will show in section 4.5.3, the

[32]Akdeniz and Van Veelen (2023) also included mutation in their model though the lack of endogenous information in their model suggests the purpose of this could not have been to avoid zero-information outcomes.

high-offer equilibrium in a model with a small rate of mutation can be asymptotically stable.

Do we not lose something important when we discard the $y^* = 0$ solutions? I think that it is important to know about their theoretical existence, but they need not be dwelt upon. The addition of mutation is arguably a step towards greater realism and is required in any case to flush out equilibria that are not subgame perfect (explained further in the next section).

Furthermore, the $y^* = 0$ solutions are clearly unstable in the information framework in cases where another solution also exists – notice in figure 4.4.2 that the slopes of the relevant curves at 0 are always greater than 45°, indicating that in the positive neighbourhood around 0, the equation will emit larger values than we put in. The positive-information solutions, on the other hand, are stable: the equation crosses the 45° line from above, demonstrating that to the left of the solution y^* is boosted and to the right it is attenuated. This brief stability analysis also suggests that a straightforward iterative numerical procedure can be to used solve the equation; this is implemented in section 4.5.2.

In summary, this section has illustrated how the information framework can be explicitly solved for very simple cases, and how endogenous information affects a simple minigame, including important effects on stability of equilibria. It also pointed out how information indeterminacy can arise in certain situations, and suggested a reasonable solution to the problem, namely to ensure, using mutation or some other device, that there are always some positive frequencies of proposer types that can cause all signals included in an analysis to occur with strictly positive probability regardless of prior reputations.

4.5 Deterministic simulation results

In this section the aforegoing framework is used to generate a series of simple models that are analysed using computer simulations of deterministic aggregate dynamics. The first models have only negative reputations, while those that follow have only positive reputations and finally both negative and positive are combined in the same model. The last section briefly describes examples of more sophisticated strategies that can evolve to take advantage of more complex information structures in larger models.

Most of the models are minigames with only two possible offer amounts, and restricted strategy sets, which conveniently allows two-dimensional graphical analysis, though models with fewer restrictions are also described in section 4.5.3 and again in section 4.5.6, which gives an indication of the relevance of excluded (generally weakly dominated) strategies. The section begins with a mathematical description of the deterministic dynamics and implementation.

4.5.1 Deterministic evolutionary dynamics

Section 4.3 explains how expected payoffs can be calculated, given strategy frequency distributions for the populations of proposers and responders. These expected payoffs reflect an endogenous information equilibrium at each point in evolutionary time, in which current strategy frequencies induce a distribution of action profiles and a distribution of signal sets for each responder type that are mutually consistent. In this section, the framework is turned into an evolutionary model by adding standard continuous replicator dynamics.[33] These equations determine how strategy frequencies evolve over time in response to expected payoffs. By working with expected payoffs, rather than actual payoffs in random encounters, we are effectively assuming that the population size is infinite so all stochastic payoff effects cancel out (see Sandholm, 2010, p. 119, Friedman, 1998, p. 20). The populations of proposers and responders evolve separately:

$$\begin{aligned} dx_i^P/dt &= x_i^P(\pi_i^P - \overline{\pi}^P) \\ dx_j^R/dt &= x_j^R(\pi_j^R - \overline{\pi}^R) \end{aligned} \tag{4.20}$$

where x_i^P is the current fraction of the proposer population adopting strategy s_i^P, $\pi_i^P = \sum_j x_j^R \pi^P(s_i^P, s_j^R)$ is the expected payoff to s_i^P against a randomly selected responder, using (4.14) to determine the expected payoff for a specific strategy profile, and $\overline{\pi}^P = \sum_i x_i^P \pi_i^P$ is the current average expected payoff in the proposer population (and similar for responders). It can be seen that strategies that earn an expected payoff higher than the average payoff will tend to increase their share in the population and strategies earning expected payoffs below the average will be adopted by a decreasing fraction of the population.

In addition to selection, I also add mutation to the dynamics. At a low rate, δ, strategies change into a strategy selected at random from all possible strategies. The interpretation is that agents innovate or make mistakes in learning. Mutation ensures that there is always a small positive frequency in each population of each possible strategy.[34]

---

[33] While the replicator dynamics has biological origins (Taylor and Jonker, 1978), reflecting the dynamical effects of differential reproduction rates of different phenotypes, the model can also be interpreted as an imitative learning model (Weibull, 1995, p. 155, Björnerstedt and Weibull, 1995, p. 163, Gale et al., 1995, p. 85, Sandholm, 2010, p. 154, Hofbauer and Schlag, 2000, p. 529).

[34] The replicator dynamics is said to be *invariant* on any face or in the interior of the simplex representing the state space, where a face corresponds to situations where some strategies have zero probability and the interior situations where all strategies have positive probability (see Hofbauer and Sigmund, 2003, p. 482). Therefore, a strategy that is present in the population can never be entirely eliminated through selection, though its frequency can certainly become small enough to be insignificant for all other calculations. On the other hand, a strategy that is absent from a population can never be introduced through selection alone either.

This addresses the indeterminacy problem discussed in the previous section, provided that the set of possible strategies are chosen to include strategies that can lead to all possible signals being emitted regardless of prior reputations. In addition, mutation can flush out strategies that make irrational choices on off-equilibrium paths, e.g. responder strategies that reject low offers should not survive merely because no proposers are making low offers. A low frequency of various strategies due to mutation provides a minimal incentive to make rational choices against all possible opponent strategies. Mutation is implemented at a low rate, so it will not generally interfere with the relatively strong force of selection, but it can sometimes affect the system's dynamics in significant ways (Gale et al., 1995). In one of the models, discussed in section 4.5.4, mutation is shown to add subtle bias to the dynamics that can stabilize an interior equilibrium.

The replicator dynamics with mutation added result in the following dynamics:

$$\begin{aligned} dx_i^P/dt &= x_i^P(\pi_i^P - \overline{\pi}^P) + \delta\left(\frac{1}{|S^P|} - x_i^P\right) \\ dx_j^R/dt &= x_j^R(\pi_j^R - \overline{\pi}^R) + \delta\left(\frac{1}{|S^R|} - x_i^R\right) \end{aligned} \tag{4.21}$$

4.5.2 Implementation

To obtain simulation results from these continuous dynamics, they are discretized as follows:

$$\begin{aligned} x_i^P(t+\tau) - x_i^P(t) &= \tau\left[x_i^P(\pi_i^P - \overline{\pi}^P) + \delta\left(\frac{1}{|S^P|} - x_i^P\right)\right] \\ x_j^R(t+\tau) - x_j^R(t) &= \tau\left[x_j^R(\pi_j^R - \overline{\pi}^R) + \delta\left(\frac{1}{|S^R|} - x_i^R\right)\right], \end{aligned} \tag{4.22}$$

where my implementation uses $\tau = \frac{1}{20}$, which gives accurate results.[35] At every τ-step, signal set probabilities are needed to be able to calculate expected utilities (see equation 4.14). According to equations (4.3) and (4.12) these probabilities result from the same distribution of action profiles that are induced by those probabilities. In practice, the pattern of interactions are determined together with the signal set probabilities, using the previous time step's probabilities as prior values. The posterior probabilities that result from the pattern of

[35]As a check, I regenerated figure 4.5.1c with $\tau = \frac{1}{100}$, and the results were practically indistinguishable. This does not mean that small errors from the ideal ($\tau \to 0$) do not occur, but they do not appear to be significant, particularly in the vicinity of asymptotically stable equilibria where movement is very slow, so the approximation is more accurate. Away from such equilibria, on faster-moving trajectories with directional changes, errors could in theory be more serious, but I still found only very minor changes in such trajectories (e.g. the simulation in figure 4.5.3d). On the other hand, the computational time savings from a slightly higher value for τ were substantial.

interactions are then compared with the prior values to determine the degree of error. If the error is significant, the procedure is repeated, this time using the previous attempt's calculated probabilities as priors. This procedure loops until an acceptably small error is obtained, indicating a solution for the endogenous information equilibrium has been found. I consider a solution to be where the absolute value of the error for all probabilities are less than 0.000000001.

Pseudocode is provided in Algorithm 1.[36] Since the strategy frequencies themselves change in relatively small steps to simulate continuous movement, the prior signal set probabilities from the previous τ-step are typically quite close to the required solution so the algorithm can find them easily enough.[37]

4.5.3 Model 1: Negative reputations

The first model considered is the negative reputation minigame model based on Nowak et al. (2000)'s first model and described in detail in section 4.4. I summarize the model again here:

$$\begin{aligned}
\$ &= 4 \\
A^P &= \{2, 1\} \\
M &= \{(1,1)\} \\
Q &= \{\{\}, \{(1,1)\}\} \\
S^P &= \{s_1^P, s_2^P, s_3^P, s_4^P\}
\end{aligned}$$

where

$$s_1^P = \begin{cases} 2, & \text{if } q = \{\} \\ 2, & \text{if } q = \{(1,1)\} \end{cases} \quad (\text{'HH'}) \qquad s_2^P = \begin{cases} 1, & \text{if } q = \{\} \\ 1, & \text{if } q = \{(1,1)\} \end{cases} \quad (\text{'LL'})$$

$$s_3^P = \begin{cases} 2, & \text{if } q = \{\} \\ 1, & \text{if } q = \{(1,1)\} \end{cases} \quad (\text{'HL'}) \qquad s_4^P = \begin{cases} 1, & \text{if } q = \{\} \\ 2, & \text{if } q = \{(1,1)\} \end{cases} \quad (\text{'LH'})$$

Recall that S_1^P and S_4^P are weakly dominated (see table 4.4.1). It is known that the presence of strategies that are themselves suboptimal can sometimes have significant effects on the relative success of other strategies (Axelrod, 1984; Skyrms, 2014), so I first report simulation results with all four proposer strategies present. Our point of departure is to investigate whether the main conclusions of Nowak et al. (2000) hold in the more general model. Their minigame

[36] The full program was implemented using the Python programming language. Graphs were produced using the matplotlib library.

[37] Occasionally, however, the algorithm can overshoot the target while searching for an information solution. When this is detected (i.e. errors increasing instead of decreasing), the program interpolates between prior and calculated posterior probabilities with progressively decreasing weights given to the posterior, effectively reducing the search step size (not shown in pseudocode).

Algorithm 1 Simulating one τ-step

function INTERACTION_RESULTS_R(j, x^P, y_j)
    payoff $\leftarrow 0$
    posterior_y^1 $\leftarrow$ vector of zeroes
    for all i **do**
        for all signal sets q **do**
            weight $\leftarrow x_i^P y_j(q)$
            outcome $\leftarrow (s_i^P(q), s_j^R(s_i^P(q)))$
            increase *payoff* by *weight* $\times\ \pi^R(outcome)$
            increase *posterior_y*^{1}[*outcome*] by *weight* $\times\ \alpha$
    calculate *posterior_y* for all signal sets from y^1 given sample size z (equation 4.13)
    return *payoff*, *posterior_y*

function INTERACTION_RESULTS_P(i, x^R, y)
    payoff $\leftarrow 0$
    for all j **do**
        for all signal sets q **do**
            weight $\leftarrow x_j^R y_j(q)$
            outcome $\leftarrow (s_i^P(q), s_j^R(s_i^P(q)))$
            increase *payoff* by *weight* $\times\ \pi^P(outcome)$
    return *payoff*

procedure RUN_FOR_1_STEP
    repeat
        for all j **do**
            $\pi_j^R, y_j' \leftarrow$ INTERACTION_RESULTS_R(j, x^P, y_j)
        info_err $\leftarrow \max_{j,q}(|y_j'(q) - y_j(q)|)$
        $y \leftarrow y'$
    until *info_err* < 0.000000001
    for all i **do**
        $\pi_i^P \leftarrow$ INTERACTION_RESULTS_P(i, x^R, y)
    for all i **do**
        $x_i^P \leftarrow x_i^P + \tau\left[x_i^P(\pi_i^P - \overline{\pi}^P) + \delta\left(\frac{1}{|S^P|} - x_i^P\right)\right]$
    for all j **do**
        $x_j^R \leftarrow x_j^R + \tau\left[x_j^R(\pi_j^R - \overline{\pi}^R) + \delta\left(\frac{1}{|S^R|} - x_i^R\right)\right]$

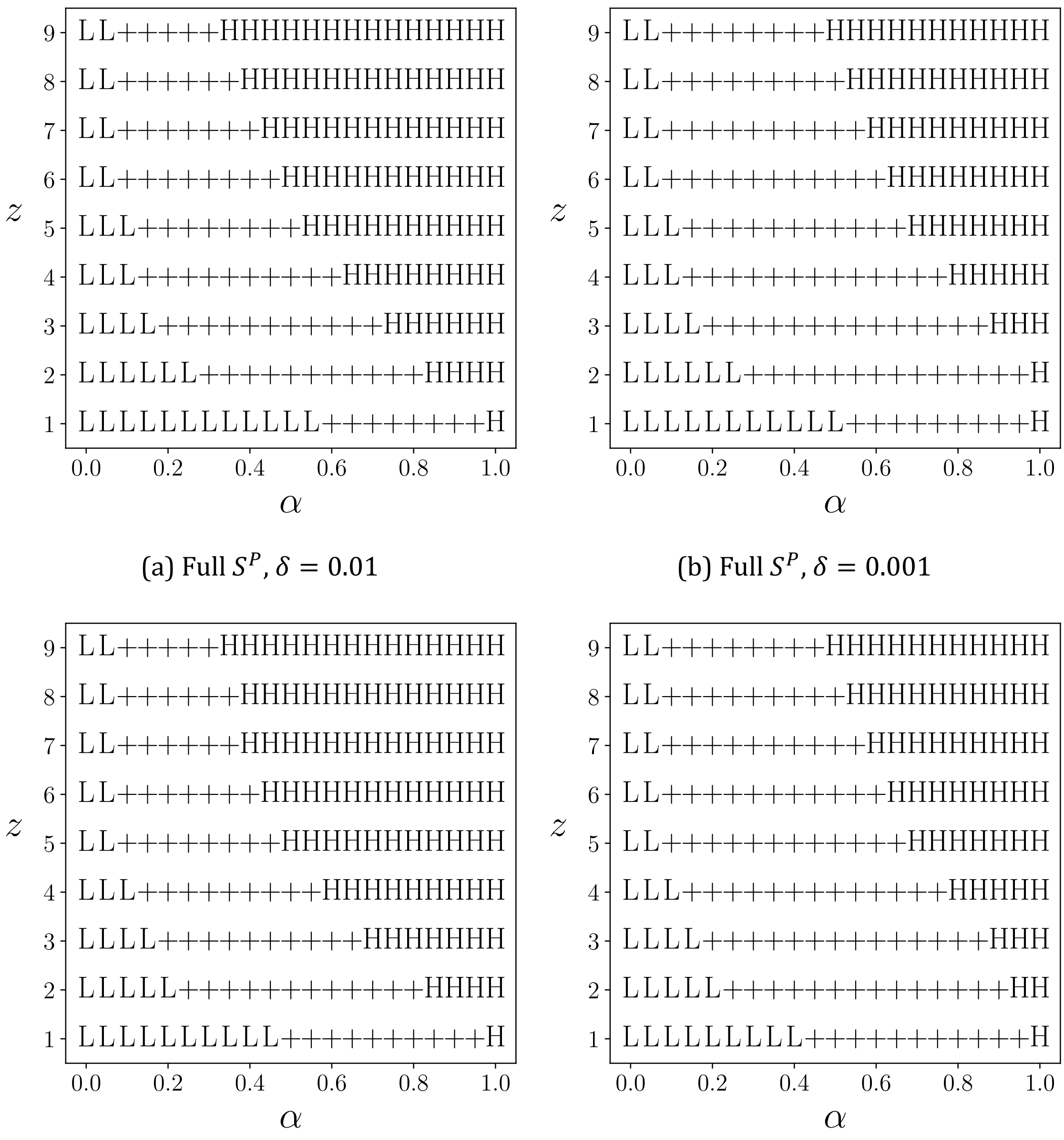
model includes only two proposer strategies, but they assume that when an H-proposer is paired with an L-responder, the offer is reduced by a fixed amount – interpreted as reflecting a proposer's average response given imperfect knowledge of the responder's strategy. This means their L-proposer is similar to my LL-proposer and their H-proposer is similar to my HL-proposer. They report that their model produces a bistable dynamic system: both H- and L-offer equilibria are stable. But their model does not feature endogenous information or mutation, so it would be interesting to establish whether a high-offer equilibrium, where information should be scarce, could be maintained under endogenous information, with the presence of a low rate of mutation. The second question is whether the endogenous information model also produces a bistable dynamic system, or alternatively whether there are reputation mechanisms that can emerge *ex nihilo*, in other words whether a high-offer equilibrium based on reputations can evolve from initial conditions in which offers are low and there are no reputations.

For the first results, I set the mutation rate to a fairly low 0.01. I ran simulations stepping through parameter values for α from 0 to 1 and z from 1 to 9. For each information parameter value pair, I used three initial conditions, the first with only LL-proposers and L-responders (labelled 'L' in table 4.5.1 below), the second with a uniform distribution over all strategies for each population ('U') and the third with only HL-proposers and H-responders ('H'). This allows both high-offer-equilibria and low-offer-equilibria to be found for all combinations of parameter values. Each simulation was allowed to run until the state stabilized.[38] The results are shown in figure 4.5.1a. Equilibria were of two types, high-offer and low-offer. H offers were made in at least 98.9% of interactions in each H-equilibrium, and L offers were made in at least 97.9% of interactions in each L-equilibrium.[39] If all three sets of initial conditions lead to the same equilibrium for given parameter values, "L" or "H" is indicated on the graph, while if different initial conditions resulted in different equilibria, "+" is indicated for a bistable dynamic system. As can be seen, if information is scarce, only the low-offer equilibrium (close to the UG's SPNE) can be stabilized. Even moderate information parameter values such as ($\alpha = 0.55, z = 1$) do not yield high-offer equilibria at all; the reasons are explored below. For higher parameter values, both low-offer and high-offer stable equilibria exist (the system is bistable), and for even higher parameters, only the high-offer equilibria are stable. The information parameters for which this occurs need to be very strong, as is shown below.

The same series of simulations were run again, this time with a substantially lower mutation rate, $\delta = 0.001$; results are shown in figure 4.5.1b. Results are

[38]This is interpreted to mean that in two successive time periods (separated by τ), the maximum change in any state variable (i.e. strategy frequency) is less than 0.00000001.

[39]These calculations Take into account both strategy and information frequencies at the equilibria.



(a) Full S^P, $\delta = 0.01$

(b) Full S^P, $\delta = 0.001$

(c) Restricted S^P, $\delta = 0.01$

(d) Restricted S^P, $\delta = 0.001$

Figure 4.5.1: Minigame equilibria with negative reputation
Key: L: low-offer equilibrium only H: high-offer equilibrium only +: bistable.

quite similar except that the system is now bistable for a larger range of information parameter values. Specifically, the low-offer equilibrium is stable in more settings with relatively strong information. This suggests that higher mutation rates tends to destabilize some low-offer equilibria so that the systems end up at high-offer equilibria instead. On the other hand, mutation seems to have a surprisingly small effect on the stability of high-offer equilibria – in other words, they seem to depend more on information parameters than on the mutation rate.

	Conditions			Proposers				Responders			
No.	z	α	Initial	LL	LH	HL	HH	L	H	${y_L}^*$	Reject
1	1	0	Any	0.497	0.497	0.003	0.003	0.995	0.005	0.000	0.005
2	1	0.4	Any	0.987	0.006	0.004	0.003	0.995	0.005	0.397	0.005
3	4	0.1	Any	0.986	0.007	0.004	0.003	0.995	0.005	0.342	0.005
4	4	0.15	L or U	0.987	0.005	0.005	0.003	0.995	0.005	0.475	0.005
5	4	0.15	H	0.007	0.007	0.775	0.211	0.554	0.446	0.016	0.006
6	4	0.2	L or U	0.987	0.004	0.006	0.003	0.995	0.005	0.587	0.005
7	4	0.2	H	0.004	0.004	0.725	0.268	0.438	0.562	0.013	0.004
8	4	0.4	L	0.974	0.003	0.020	0.003	0.995	0.005	0.868	0.005
9	4	0.4	U or H	0.001	0.001	0.747	0.250	0.016	0.984	0.407	0.003
10	4	0.6	L	0.831	0.003	0.164	0.003	0.993	0.007	0.973	0.006
11	4	0.6	U or H	0.001	0.001	0.748	0.250	0.008	0.992	0.858	0.002
12	4	0.8	Any	0.001	0.001	0.748	0.250	0.007	0.993	0.969	0.002

Table 4.5.1: Minigame 1 (Negative reputations) characteristics of selected equilibria ($\delta = 0.01$)

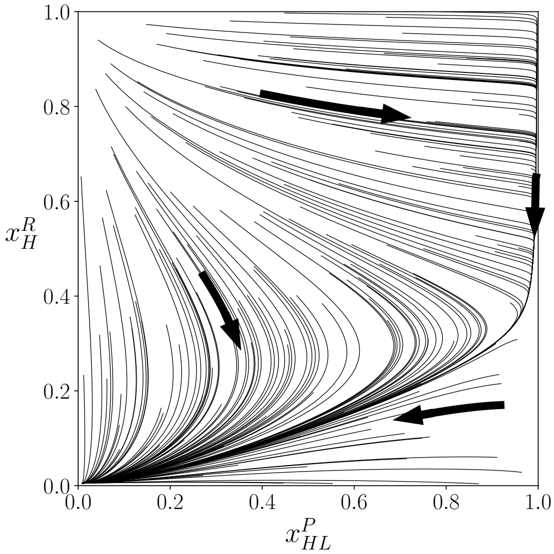
Some characteristics of selected equilibria are given in table 4.5.1. Each line represents an equilibrium, which can in some cases be reached from any initial distribution ("Any"), but in cases of bistable systems, only from certain initial distributions ('H', 'U', or 'L', as described above). The equilibrium frequencies of each strategy are given, as well as the probability with which a responder can observe the $(1, 1)$ signal when interacting with an L-responder, i.e. the probability that the proposer knows that the responder will accept an L offer (indicated in the table as y_L^*). The last column shows the rejection rate, which is quite low in all cases – strategies that result in substantial rejection are rapidly selected against, though the presence of mutation ensures that all strategies are present at minimal frequencies at least.

Only in the first case (no. 1) is there any significant presence of the irrational LH proposer strategy. In this case, the probability of the signal being observed is zero so LH is equivalent to LL, while in all cases where the signal can be observed, the LH strategy frequency is driven down close to zero rapidly. When information is present, but weak (nos. 2 and 3), as indicated previously, there is only one equilibrium with low offers, so only LL-proposers and L-responders have high frequencies. If the information parameters are strengthened, the system becomes bistable, so stable low-offer equilibria can be reached from some but not all initial distributions (nos. 4, 6, 8 and 10). Even when the signal probability is at 0.973 (no. 10), which is almost a certainty, this is still not enough to get out of the low-offer equilibrium. The reason for this is that if most offers are low, switching to demanding H will immediately trigger rejections. Only when there is a substantial enough frequency of HL-proposers is it possible to move out of an L-offer equilibrium, but HL-proposers are at a disadvantage to LL-proposers when the vast majority of responders accept low offers, the only exception being if they can observe the L-responders' negative reputations with a probability approaching certainty.

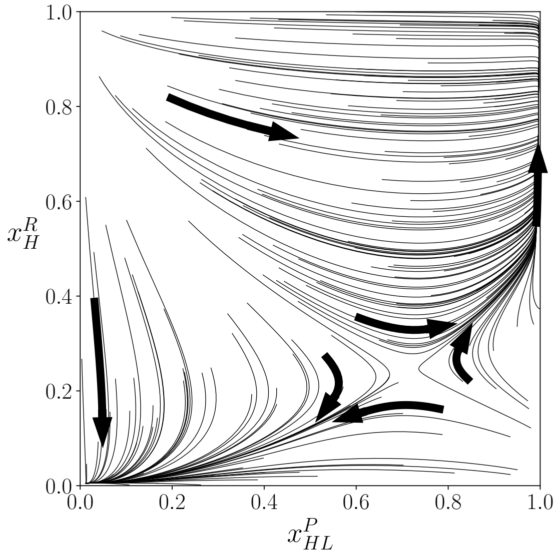
On the other hand, it takes extremely little information to stabilize a high-offer equilibrium. In the seventh case, the probability of the signal being observed is a mere 0.013, yet the high-offer equilibrium is stable. This shows endogenous information at work: despite fairly robust information parameters ($z = 4, \alpha = 0.2$), the signal is observed only very infrequently because few interactions take place where L-responders receive low offers when the system is at a high-offer equilibrium.[40] The small probability of the signal being received is barely enough to distinguish HL- from HH-proposers, but there is a slight payoff advantage to HL-proposers (2.0059 vs 2.0000) so their frequency remains just high enough for H-responders to get very slightly higher expected payoffs than L-responders (1.9856 vs 1.9832). The small payoff difference allows L-responders' frequency to become substantial, which is a result of mutation, that tends to pull the responder distribution towards uniformity when the payoff gradient is flat. A high-offer equilibrium can be maintained with even weaker information parameters (no. 5) where L-responders even have a slight payoff advantage over H-responders, but the difference is small enough that mutation can maintain the frequency of H-responders at a level high enough to keep the proposers' payoff advantage with high offers.[41]

[40] There are almost no LL-proposers (who would generate the signal often when paired with an L-responder) and a substantial share of HH-proposers whose interactions never generate the signal. The signal can be generated by HL-proposers interacting with L-responders but because information is so scarce these proposers rarely make low offers to them.

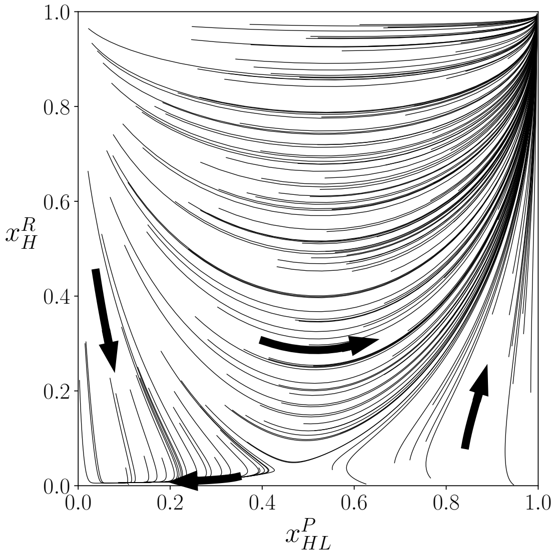
[41] This is a rare case of a high-offer equilibrium being maintained by the force of mutation. Figure 4.5.1b indicates that the rate of mutation is not critical: the same thing happens at a much lower mutation rate as well. This is similar to the mechanism investigated by Gale et al. (1995). In their model, the high-offer equilibrium is maintained by boosting the rate of responder mu-



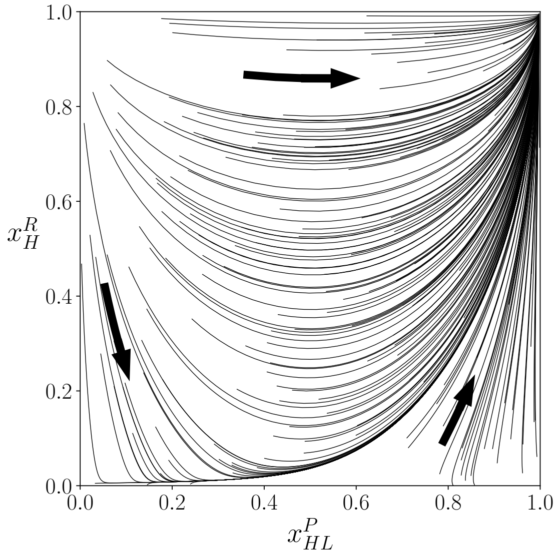
(a) $z = 1, \alpha = 0.4$



(b) $z = 4, \alpha = 0.2$



(c) $z = 4, \alpha = 0.5$



(d) $z = 4, \alpha = 0.8$

Figure 4.5.2: Phase portraits for a minigame with negative reputations

The presence of a significant frequency of the weakly dominated HH-responders in H-equilibria, and the fact that interactions with these proposers produce no information (i.e. accepted low offers), suggest their presence may have a negative impact on stability of high-offer equilibria. I ran all of the aforegoing simulations again with only two possible proposer strategies included in S^P, LL and HL. The results, shown in figures 4.5.1c and 4.5.1d indicate that the impact is relatively unimportant: for a small number of parameter values the dynamics now give only low-offer equilibria (e.g. $z = 5$, $\alpha = 0.1$, $\delta = 0.01$), where previously there were also high-offer equilibria, but for most parameter values the stable outcomes are the same.[42]

Since it appears that the weakly dominated strategies HH and LH are not important, I remove them for the remainder of the analysis in this section. This very usefully allows two-dimensional phase portraits to be used to illustrate trajectories of the dynamic systems for all areas of the state space. I also fix $\delta = 0.01$. Trajectories are plotted for a number of selected parameter value combinations to illustrate dynamics when there is only a low-offer equilibrium (figure 4.5.2a), when the system is bistable (figures 4.5.2b and 4.5.2c) and when there is only a high-offer equilibrium (figure 4.5.2d). In all cases, there is strong selection towards high offers when x_H^R is high, and strong selection towards low responder demands when x_{HL}^P is low. Both effects aim to avoid high probabilities of rejections.

There is also strong selection pressure on proposers to make low offers when x_H^R is close to zero, *except* when information parameters are extremely strong. In the lower-left-hand-side corner of figure 4.5.2d, the probability of learning about an L-responder's acceptance of low offers is almost one, and escape from a low-offer equilibrium is possible. Note that mutation will tend to (slightly) push both X_{HL}^P and x_H^R higher, thus aiding the eventual transition to a situation where high offers are the norm.

On the other hand, the downwards selective pressure on responders when x_{HL}^P is close to one (seen in figure 4.5.2a) is weak and can easily be reversed if there is even a modest probability of proposers being able to find out about L-responder's acceptance of low offers, so upwards movements are observed on the right-hand side in all cases except the first. As indicated above, even though the probability of the signal is quite small, it still requires fairly strong selection parameters because information is inherently weak when there are very few LL-proposers.

To summarize, with the possibility of negative reputations, it will under most circumstances be difficult to escape low-offer equilibria, similar to the

tation relative to that of proposers, but here the mutation rates are equal for both populations; the nudge required to reach the same outcome is instead provided by reputations.

[42]There are also isolated cases where low-offer equilibria disappear when S^P is restricted (e.g. $z = 3$, $\alpha = 0.7$, $\delta = 0.01$). This could be due to mutation towards HH-proposers making H-responders' expected payoffs slightly higher.

UG's SPNE. On the other hand, it is also quite easy to stabilize a high-offer equilibrium in many cases, as the force of selection on responders towards accepting lower offers is rather weak when most offers are high, and not much help is required from the reputation mechanism, which is actually rather weak in terms of delivering reliable information. Not much information is required, and not much information is available, but occasional low offers being accepted and knowledge thereof serve as enough of a disincentive effect to responders to maintain their high demands.

These results therefore broadly concur with Nowak et al. (2000) in that both high-offer and low-offer equilibria can be stable, though now accounting properly for endogenous information. This suggests that evolution will sometimes result in either outcome. But experimental results, even while there is diversity in outcomes, almost never exhibit the low-offer SPNE result where the proposer takes practically all of the surplus and the responder accepts (Henrich et al., 2005), suggesting that other models also need to be considered.

4.5.4 Model 2: Positive reputations

While the negative reputation mechanism considered above is a natural and intuitive way to account for information that proposers could have about past behaviour of responders, it is not the only possible way to think about reputations. If only negative reputations are modelled, a proposer who observes a responder rejecting a low offer would not treat that responder any differently, though there is clearly valuable information in such an observation. A reasonable proposer should conclude that it is in his best interests to make higher offers to such a responder, to avoid similar rejection, thus such reputations are *positive*. Allowing proposers to be sensitive to positive reputations, as will be shown, leads to totally different results to those of the negative reputation model above.

The second model I consider therefore allows for positive reputations by including the signal sets with the signal $(1,0)$ in the set of possible signal sets Q:

$$\begin{aligned}
\$ &= 4\\
A^P &= \{2,1\}\\
M &= \{(1,0)\}\\
Q &= \{\{\},\{(1,0)\}\}\\
S^P &= \{s_1^P, s_2^P\}
\end{aligned}$$

where

$$s_1^P = \begin{cases} 1, & \text{if } q = \{\} \\ 2, & \text{if } q = \{(1,0)\} \end{cases} \quad \text{('LH')} \qquad s_2^P = \begin{cases} 2, & \text{if } q = \{\} \\ 2, & \text{if } q = \{(1,0)\} \end{cases} \quad \text{('HH')}$$

I only allow two proposer strategies, LH, which makes low offers by default, but high offers to a responder with a positive reputation (i.e. known to reject low offers), and HH, which makes high offers unconditionally. The other possible proposer strategies, not included, are LL, which is weakly dominated by LH, and HL, which is weakly dominated by HH as it makes low offers specifically to responders known to reject them, with predictably poor results.

<table>
<tr><td></td><td></td><td></td><td colspan="2">Responder</td></tr>
<tr><td></td><td></td><td></td><td>H</td><td>L</td></tr>
<tr><td rowspan="3">Proposer</td><td colspan="2">HH</td><td>2, 2</td><td>2, 2</td></tr>
<tr><td rowspan="2">LH</td><td>if q={}</td><td>0, 0</td><td>3, 1</td></tr>
<tr><td>if q={(1,0)}</td><td>2, 2</td><td></td></tr>
</table>

Table 4.5.2: Strategic-form representation of an ultimatum minigame with positive reputations

The strategic-form representation of this game is shown in table 4.5.2. Note, only H-responders can exhibit the signal. The negative reputations considered in the previous section benefited proposers by allowing them to lower their offers safely, to the detriment of responders. Positive reputations benefit proposers in a different way, by helping them avoid rejections when they make low offers by default. Positive reputations help H-responders in two ways: by raising the offers they receive and by helping them avoid rejections.

It can be seen from the table that, without any information, both (HH, H) and (LH, L) are NE, but only the latter is subgame-perfect. On the other hand, if the signal were always received when interacting with H-responders, then (HH, H) and (LH, H) would both be SPNE. But such perfect observability is hardly realistic. If proposers could only observe the signal with a positive probability smaller than one, then LH would no longer be a best response to H as its payoff would be some weighted average of 0 and 2, while HH gives 2 with certainty.

Endogenous information plays an integral role in the analysis. Responders, facing a mix of HH- and LH-proposers, get the same payoff against HH-proposers regardless of what their demands are, so their best response between H and L will be determined solely by what expected payoff they obtain against LH-proposers, even if the latter's frequency is very low in the proposer population. Responders' best response against LH-proposers depend critically on information. If the probability with which an H-responder's $(1, 0)$ signal is observed by responders – call this probability y_H^* – were less than 0.5, L would be best, while if the probability were more than 0.5, H would be best.

Suppose responders play H with high frequency. Then proposers would do better to play HH (supposing $y_H^* < 1$ at all times). But if proposers mostly switched to HH, then very few low offers would be made, and y_H^* would consequently be low. This means responders would do better if they switched to L. If enough of them do so, proposers would do better to switch to LH. But this raises y_H^*, which creates and incentive for responders to switch to H again, and so forth. We conclude that, if endogenous information is taken into account, there cannot be an equilibrium in pure strategies, indicating that there must instead be a mixed-strategy equilibrium.

The mixed strategy equilibrium can be calculated using the indifference principle of mixed strategies: y_H^* would need to be 0.5 to make the expected payoff of H- and L-responders equal. Responders would need to mix H and L with equal probabilities to make proposers indifferent between HH and LH. The proposer mix in the equilibrium would need to ensure that $y_H^* = 0.5$, so would depend on the information parameters. From (4.13), taking note that the signal can only result from an interaction between an H-responder and an LH-proposer when the signal is *not* observed by that proposer,[43]

$$\begin{aligned} y_H^* &= y_H(\{(1,0)\}) \\ &= 1 - \left(1 - \alpha x_{LH}^P(1 - y_H^*)\right)^z . \end{aligned}$$

Taking $y_H^* = 0.5$ on both sides and solving,

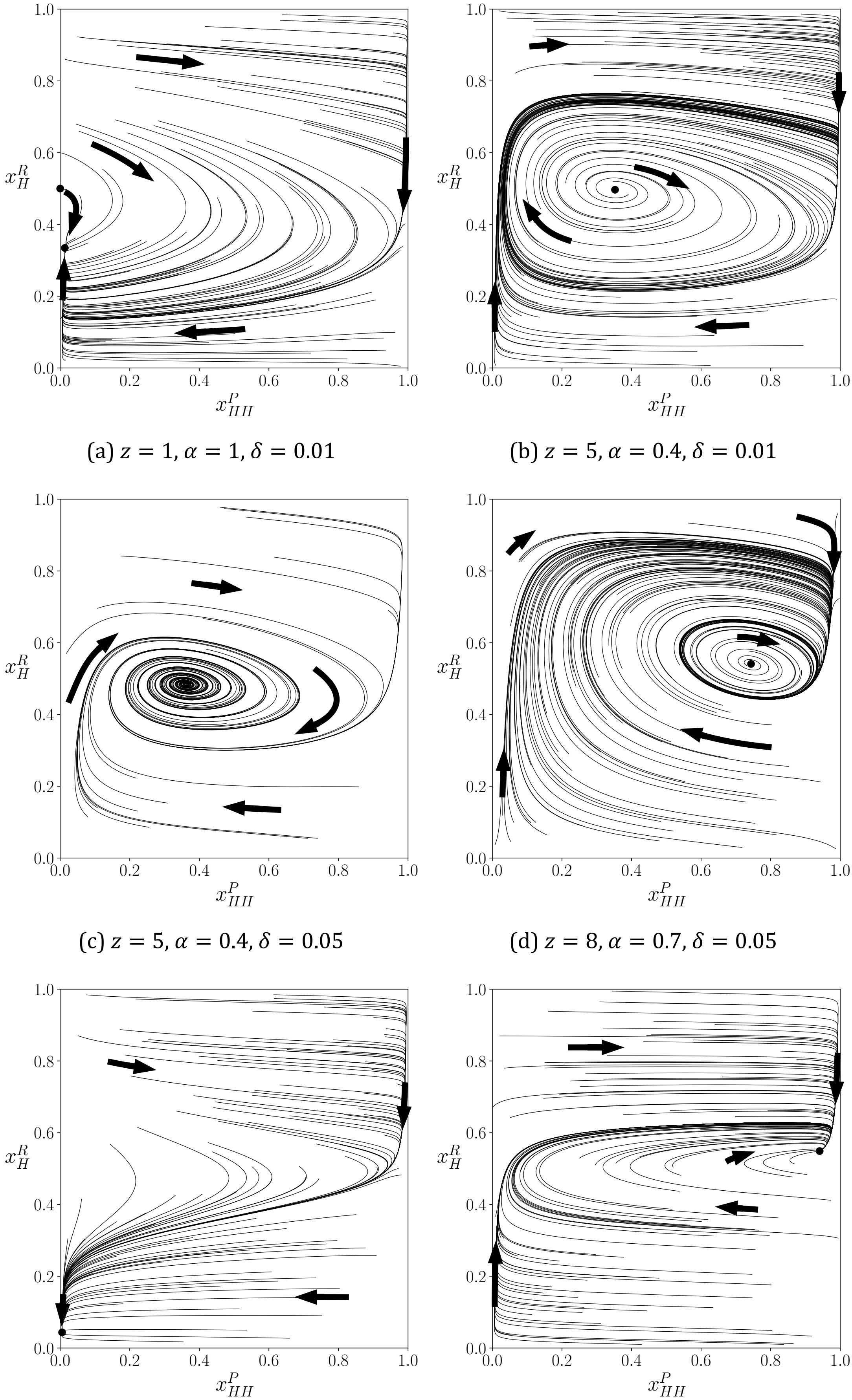
$$\sqrt[z]{\frac{1}{2}} = \left(1 - \alpha x_{LH}^P\left(1 - \frac{1}{2}\right)\right)$$

$$x_{LH}^P = \frac{2 - 2\left(\sqrt[z]{\frac{1}{2}}\right)}{\alpha} \tag{4.23}$$

If $z = 1$, this cannot yield values for x_{LH}^P that are in the interval $[0, 1]$, except if $\alpha = 1$, in which case there is a mixed equilibrium where $x_{LH}^P = 1$. For values of $\alpha < 1$, there are no mixed strategy equilibria. There are more possibilities when $z > 1$, with mixed equilibria in the interior, i.e. $0 < x_{LH}^P < 1$.

Before interpreting the meaning of these mixed strategy equilibria, we need to investigate whether they can be reached and maintained by evolutionary dynamics. One would normally expect evolutionary fixed points near NE, but they may not necessarily be stable and the dynamics are complicated by mutation.

[43]This arrangement results in somewhat surprising calculations, e.g. if $z = 1$, $\alpha = 1$ and every interaction is between an LH-proposer and an H-responder, what is the probability with which the $(1, 0)$ signal can be observed by the LH-proposers? Suppose the answer is 1; but then LH-proposers would always make H offers to H-responders and the signal could never be observed, so 1 cannot be correct. There is a strong negative feedback effect for this type of signal. The correct answer is 0.5 – exactly half of these interactions must result in rejections for a random sample of one interaction, perfectly observed, to yield the signal with a probability of 0.5.



(a) $z = 1, \alpha = 1, \delta = 0.01$

(b) $z = 5, \alpha = 0.4, \delta = 0.01$

(c) $z = 5, \alpha = 0.4, \delta = 0.05$

(d) $z = 8, \alpha = 0.7, \delta = 0.05$

(e) $y_H^* = 0.44$ (fixed)

(f) $y_H^* = 0.52$ (fixed)

Figure 4.5.3: Phase portraits for a minigame with positive reputations

A trajectory plot for $z = 1$, $\alpha = 1$ is shown in figure 4.5.3a. Note that these trajectory plots show the frequency of HH on the horizontal axis to allow easy comparison with previous figures, where the higher offers are on the right-hand-side.[44] All trajectories converge to a stable equilibrium point where $x^P_{LH} = 0.986$ and $x^R_H = 0.335$ (indicated by a dot on the graph), which is not particularly close to the mixed strategy NE indicated above (indicated by another dot). The difference can be explained by the effect of mutation.[45]

Further detail of the state at the evolutionary equilibrium is given in the first row of table 4.5.3. It confirms that an H-responder will be recognized as such from his reputation with a probability of 0.496 which is near the theoretical maximum of 0.5 for the information parameters. The substantial frequency of H-responders combined with availability of information makes LH-proposers' expected payoffs slightly higher than HH-proposers' (2.33 vs. 2.00). Both types of responders get quite low expected payoffs, 1.015 for L-responders and 1.008 for H-responders, only slightly above the maximin payoff 1, indicating that practically the entire benefit of the information accrues to LH-proposers. High offers get made approximately 18% of the time, so the average offer is approximately 1.18, or 29.4% of the total amount. The equilibrium is inefficient, with a rejection rate of 16.6%. Since positive reputations can only come from rejections, an equilibrium based on positive reputations will necessarily have some level of rejections, hence inefficiency.

Can more favourable information parameters be more beneficial to responders? Can they lead to more efficient equilibria? The problem is that if LH-proposers only look at a sample of one past interaction, they need to observe, on average, one rejection of a low offer for every high offer they make, since high offers do not generate information. If responders considered a larger sample, then a small rejection rate could generate a proportionally higher probability of the positive reputation being observed by proposers.

Consider the case where $z = 5$ and $\alpha = 0.4$. Using the above reasoning and equation (4.23), the mixed NE here can be calculated to be at the coordinates $(x^P_{HH}, x^R_H) = (0.3528, 0.5)$. Trajectories for this case are plotted in figure 4.5.3b which shows somewhat exotic dynamics (see also no. 2 in table 4.5.3). There is

[44] Here, however, offers can be high on the left-hand-side as well, depending on the probability with which the LH-observers see the $(1, 0)$ signal, in which case they make high offers. Since this probability is ordinarily below one, we have higher average offers towards the right-hand side.

[45] Trace a trajectory (indicated by an arrow) that starts at the NE coordinates $(0, 0.5)$. We see rightwards movement from this point, which is due to mutation. At the NE, proposers get the same expected payoff from LH and HH, so there is no selection pressure to counteract the rightwards force of mutation. But this causes y^*_H to decline slightly below 0.5, which reduces the expected payoff to H-responders a little (refer to table 4.5.2), so there is also downwards movement. The greater frequency of L-responders again favours LH-proposers, so there is leftwards movement again. At the evolutionary equilibrium, the forces of selection and mutation are perfectly balanced, which is not the case at the NE.

	Conditions				Proposers		Responders			
No.	z	α	δ	Initial	LH	HH	L	H	$y_H{}^*$	Reject
1	1	1	0.01	Any	0.985	0.015	0.666	0.335	0.496	0.166
2	5	0.4	0.01	None	0.353	0.647	0.503	0.497	0.500	0.161
3	5	0.4	0.05	Any	0.361	0.640	0.516	0.484	0.497	0.155
4	5	0.7	0.05	None	0.256	0.744	0.459	0.541	0.516	0.067
5	8	0.7	0.10	Any	0.304	0.697	0.424	0.576	0.551	0.078
6	1	0.3	0.01	Any	0.995	0.005	0.991	0.009	0.230	0.007
7	9	0.99	0.01	None	0.166	0.834	0.487	0.513	0.502	0.042

Table 4.5.3: Minigame 2 (Positive reputations) characteristics of selected equilibria

a central point, indicated by a dot, which is in fact the system's only evolutionary equilibrium, almost exactly at the mixed NE's coordinates, and it happens to be unstable.[46]

Trajectories spiral *outwards* from this point, until it reaches a regular pattern with large oscillations, indicated by the dark orbit that all of the trajectories eventually converge to. This is a limit cycle. The orbit's clockwise direction can be explained by the same reasoning I used above to argue that there can be no NE in pure strategies: every time y_H^* is high, H-responders surge, followed by a surge in HH-proposers (who do not risk rejection as the LH proposers do), which destroys the ability of H-responders to maintain positive reputations (y_H^* drops to practically zero), causing HH-proposers' frequency to collapse, and the cycle is repeated again. This dynamical behaviour reveals a big problem with positive reputations: their very success leads to their undoing, because a high frequency of H-responders is inevitably followed by proposers switching to unconditional high offers. Lack of information is particularly damaging to H-responders' relative fitness as they must reject the resulting low offers from LH-proposers.[47]

Can the evolutionary equilibrium in this model ever be stable? The general pattern of stable oscillations seems to hold for many parameter values, but I will illustrate a modification that can give a different result. In figure 4.5.3c,

[46]The precise coordinates $(0.3526, 0.497)$ were determined by running the computer simulation *backwards* from a point in the interior of the orbit. This was achieved by setting τ to a small negative value. Initially, this did not work as anticipated due to small errors that accumulated when the simulation was reversed, which tended to break the unity sums of the distributions. This could be countered by re-normalizing frequency distributions at every step.

[47]In the negative reputation model, lack of information is not so damaging, as the default offer of the dominant HL-proposers in high-offer equilibria is high.

the dynamics converge towards the mixed evolutionary equilibrium from all initial points. The only difference is that the mutation rate has been increased ($\delta = 0.05$); the mechanism is, straightforwardly, mutation pushing the state towards the interior from all directions, thus the trajectories are altered slightly so they spiral inwards instead of outwards. It is not clear which of the two systems is more realistic, but it is useful to know that both possibilities exist. In figure 4.5.3d, the information parameters have been substantially increased ($z = 8$, $\alpha = 0.7$) while maintaining the higher mutation rate $\delta = 0.05$. The equilibrium is further to the right, but the higher mutation rate now fails to stabilize it – trajectories again converge to a limit cycle, similar to figure 4.5.3b, but less dramatic. Further boosting mutation does stabilize it (no. 5 in table 4.5.3).

With such strong information parameters, these equilibria are much more efficient due to the high frequency of HH-proposers, and average responder payoffs are also higher, approximately 1.72. Even when the system oscillates, offers are high 78.4% of the time (averaging over time), so on average proposers offer 44.6% of the total amount to responders, which may be comparable to various experimental results. The average rejection rate is still positive but reasonably low at 7.5%.[48]

It must be emphasized that the oscillating behaviour and instability of some equilibria are purely due to endogenous information. Figures 4.5.3e and 4.5.3f show systems in which endogenous information has been removed by fixing y_H^* at two different levels.[49] This shows that an analysis of positive reputations that neglects to incorporate endogenous information could result in misleading conclusions.

To summarize the results for the positive reputation models: these models show mixed strategy equilibria due to the inherent tendency of strong information on positive reputations to undermine itself. Dynamic systems based on positive reputations often do not show a stable equilibrium point but limit cycles and oscillatory behaviour. When stable, the equilibria are inefficient, because maintaining positive reputations require regular rejections. One attrac-

[48]Oosterbeek et al. (2004)'s meta-study finds an average rejection rate of 16% in UG experiments, with high variation between studies. There are numerous possible explanations for this apart from an instinctive need to maintain a positive reputation, so I would not necessarily expect any of the simple evolutionary models to match empirical results closely, though it remains a plausible theory that there is a link between positive rejection rates and concern with positive reputations.

[49]The critical aspect of endogenous information is the own-population effects for proposers. It can clearly be seen by comparing the locus of points of the trajectories where $dx^P/dt = 0$, i.e. the points where the curves have a vertical slope. With endogenous information, the locus traces out a downwards-sloping line, indicating that the best response for proposers are affected not only by responder strategies but also by x^P, while in figures 4.5.3e and 4.5.3f these loci are perfectly horizontal and indicate the level of x_H^R where the proposer's best response changes from LH to HH.

tive aspect of the positive reputation model is that it avoids the bistabilism seen in the negative reputation model. A consequence of this is that, provided information parameters are strong enough, reputations can arise *ex nihilo*, so the model can give an account of how a society can develop norms for more equal distributions and rejections of low offers where such norms did not exist before.[50]

4.5.5 Model 3: Negative and positive reputations

Responders have no natural choice to utilize negative or positive reputations or not – they only accept or reject offers. We have seen that if proposers are responsive only to negative reputations, we can obtain positive results, but it can be difficult to get out of a low-offer equilibrium. A high-offer equilibrium can be sustained by negative reputations, despite the endogenous information model indicating that in such situations the actual amount of information available to proposers is minimal. On the other hand, if proposers are sensitive only to positive reputations, it is possible to get out of low-offer equilibriums if information parameters are reasonable. However, the equilibria they produce appear to be inherently unstable as a result of endogenous information, and oscillations and inefficiency typically result.

There is no natural reason why proposers should be sensitive to the one or the other, but not to both types of information simultaneously. Therefore, I investigate a model with both types of information, which is still a minigame with only two possible offers. There are now four conceivable information sets: the empty set, the positive and the negative signals and their combination. But a responder with a fixed pure strategy can never be observed to both accept and reject low offers, so only three signal sets can be observed. Initially, I only include two proposer strategies, as summarized below:

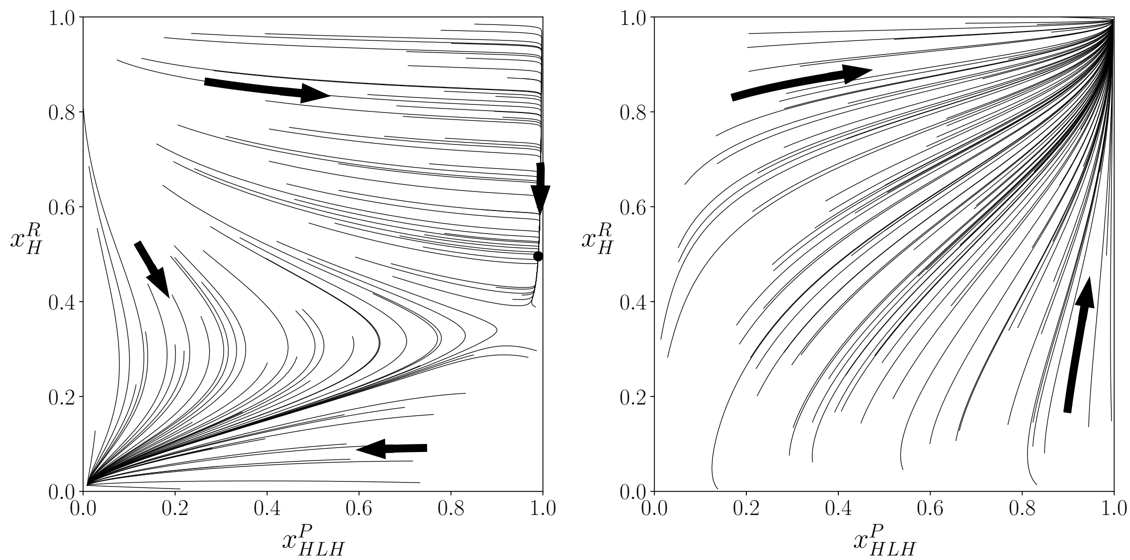
$$\begin{aligned}
\$ &= 4 \\
A^P &= \{2, 1\} \\
M &= \{(1,1), (1,0)\} \\
Q &= \{\{\}, \{(1,1)\}, \{(1,0)\}\} \\
S^P &= \{s_1^P, s_2^P\}
\end{aligned}$$

[50] The key to understanding why the UG's SPNE can be escaped with positive reputations is that the proposer strategy that is conditionally responsive to positive reputations weakly dominates one that makes low offers unconditionally. Even if all responders accepted low offers, it would cost nothing for proposers to switch to a strategy like the LH strategy in the positive-reputation model that makes high offers to a responder only if the responder has been observed to reject a low offer. Additionally, the probability of such a signal being observed, y_H^*, is high when most offers are low. In the negative reputations model, information is also strong near the UG's SPNE, but HL strategies only become better responses than LL once a substantial share of responders demand H.

where

$$s_1^P = \begin{cases} 1, & \text{if } q = \{\} \\ 1, & \text{if } q = \{(1,1)\} \\ 2, & \text{if } q = \{(1,0)\} \end{cases} \quad (\text{‘LLH’}) \qquad s_2^P = \begin{cases} 2, & \text{if } q = \{\} \\ 1, & \text{if } q = \{(1,1)\} \\ 2, & \text{if } q = \{(1,0)\} \end{cases} \quad (\text{‘HLH’})$$

The two proposer strategies are effectively the two sophisticated strategies from the aforegoing models, respectively LLH responds to positive reputations as the LH strategy in the positive reputation model and HLH responds to the negative strategy as HL does in the negative reputation model. They only differ in their default offer if they see the empty signal set {}. While having only two proposer strategies is convenient for allowing two-dimensional phase portraits, there is some justification for the exclusions as all other possible responder strategies are weakly dominated by these two, including LLL and HHH which are weakly dominated by LLH and HLH respectively.[51]



(a) $z = 5, \alpha = 0.1$

(b) $z = 5, \alpha = 0.4$

Figure 4.5.4: Phase portraits for a minigame with negative and positive reputations ($\delta = 0.01$)

Two trajectory plots are shown for low and moderate information parameters in figure 4.5.4, with the frequency of HLH-proposers measured on the horizontal axis. The first graph looks much like the bistable negative reputation

[51]This leaves HHL, HLL, LHL and LHH, which are weakly dominated by HHH, HLH, LLL and LLH respectively.

model, e.g. figure 4.5.2b,[52]. It is a true high-offer equilibrium with high offers being made 99% of the time, and both L- and H-responders enjoying very high average payoffs of 1.98. As in the negative reputation model, the equilibrium is supported by very little information - the probability of observing a negative signal for a L-responder is 0.01 and the probability of observing a positive signal for an H-responder is 0.005. The equilibrium is efficient as the overall rejection rate is very low (0.005). Since proposers are almost all of the HLH type, it must be mostly the negative reputation mechanism that maintains the equilibrium.

Figure 4.5.4b shows the situation if α is raised. The system is no longer bistable, and contains only the high-offer equilibrium. This is somewhat similar to the negative reputation model when there is only a high-offer equilibrium, e.g. figure 4.5.2d, but note from figure 4.5.1c that for these parameter values, the negative reputation model would have been bistable. The direction of movement is also different from figure 4.5.2d, especially near the lower-left corner (the UG SPNE), where there is upwards movement, suggesting H-responders have an expected payoff advantage in this area. This suggests that positive reputations can be effective in bootstrapping a system out of a low-offer equilibrium with modest information parameters, which accords with what the analysis for positive reputations above suggests.

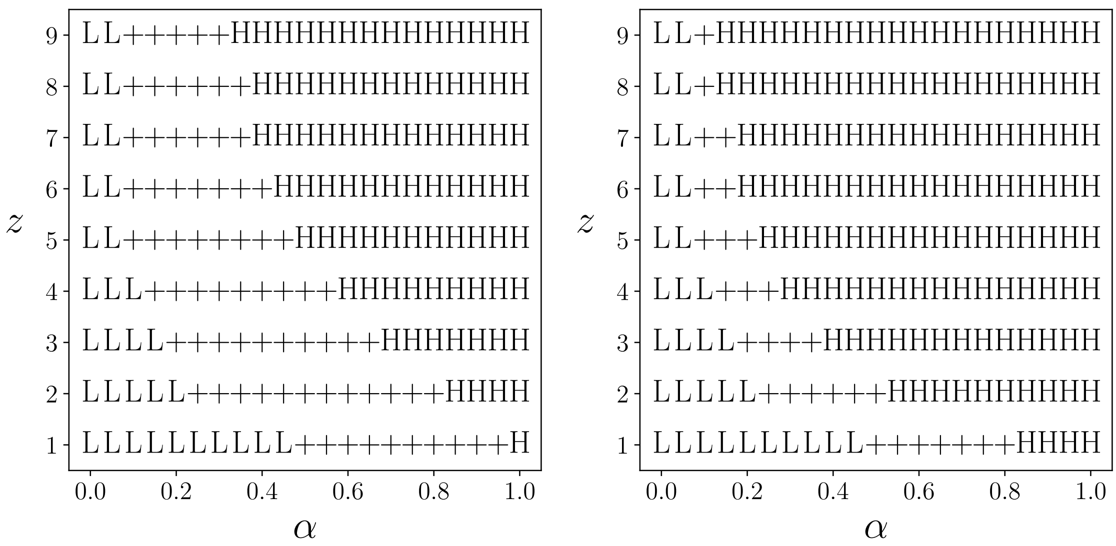
I ran a full sweep of simulations for the combined negative and positive reputation model under different information parameter values, each with three different initial states as before; the results are presented in figure 4.5.5b. For side-by-side comparison, I reproduced figure 4.5.1c next to it as figure 4.5.5a. Interestingly, there are no additional parameter value combinations for which high-offer equilibria are stable, but there are many more parameter combinations that were bistable in the negative-reputation model but now have only a high-offer equilibrium. In no case did the system exhibit oscillatory behaviour, as in the positive-reputation models.

It can be concluded that positive reputations are more relevant in states where mostly low offers are made, as they provide a potential mechanism that rewards responders for rejecting low offers even when all offers are low, while negative reputations are more effective at maintaining high-offer equilibria efficiently and with low informational requirements.

4.5.6 Beyond minigames

I have considered three minigames, as 2×2 models are analytically convenient to work with, but there are many more possibilities that can be generated by the information framework. I will briefly explore a few possibilities here without

[52]The high-offer equilibrium (indicated by a dot) is at a lower frequency of H-responders than in figure 4.5.4, but this is mainly just because parameter values α and z are different



(a) Negative reputation only

(b) Negative and positive reputations

Figure 4.5.5: Minigame equilibria with negative and positive reputations ($\delta = 0.01$)
Key: L: low-offer equilibrium only H: high-offer equilibrium only +: bistable.

detailed analysis.

Firstly, consider a model with combined negative and positive reputations with an unrestricted proposer strategy set, thus including all eight proposer strategies, using parameter values $z = 5$, $\alpha = 0.4$ and $\delta = 0.01$. From different initial states (including unconditional low-offers and low demands), this model reaches a high-offer equilibrium quite similar to the one above in figure 4.5.4b. Most of the weakly dominated strategies dropped to approximately zero, but the proposer strategy HHH[53] (making high offers unconditionally) shows a frequency of 0.15 in the equilibrium, which it could reach due to mutational drift and the fact that its expected payoff is only slightly less than that of HLH, which shows a frequency of 0.71. There are also smaller positive frequencies of HLL and HHL, for similar reasons. The presence of some unconditional high offers then results in the equilibrium frequency of H-responders being much higher at 0.991. Running the model again with $\alpha = 0.01$ resulted in a low-offer equilibrium similar to the one shown in figure 4.5.4a, this time with a mix of LLH (frequency 0.834) and unconditional low offers LLL (frequency 0.153).

Next, consider a model with another signal type included, namely $(2, 1)$. The motivation for this is that an observation of a responder receiving a high

[53]The three H's are respectively the offers made in response to observing $\{\}$, $\{(1, 1)\}$ and $\{(1, 0)\}$.

offer and accepting it may be informative because the proposer who earlier made that high offer may have had better information about the responder than is available presently. Including three possible signals results in six possible signal sets after eliminating two that contain contradictory signals that could never occur together. This results in a total of 64 possible proposer strategies.

If signal set ___ observed,	then offer ___
{}	L
{(1,1)}	L
{(1,0)}	H
{(2,1)}	H
{(1,1),(2,1)}	L
{(1,0),(2,1)}	H

Table 4.5.4: Sophisticated proposer strategy in game with three possible signals (low-offer equilibrium)

Despite repeating the simulations under numerous parameter value combinations, I found that the $(2,1)$ signal was ignored by prevailing proposers in almost all cases. Since the signal is noisy (high offers could have been made to L-responders by some suboptimal mutant proposers), conditions would have to be just right for it to be optimal to condition offers on the information. Eventually, for the parameter values $z = 20$, $\alpha = 0.06$, $\delta = 0.3$, the expected sophisticated responder strategy described in table 4.5.4 was found to have the highest expected payoff of all (though by a tiny margin).[54] Generally, however, the usefulness of this kind of information in this setting is clearly limited. It may be more useful in settings in which a proposer's own observations can be erroneous, there are mixed strategies or if strategies evolve while being sampled.

Finally, consider a model in which the total money amount is increased to 6 and the set of possible offers increased to $\{1,2,3\}$. Even allowing only a limited set of signals, $\{(1,0), (1,1), (2,0), (2,1)\}$, results in eight possible signal sets and 6561 possible proposer strategies (and three responder minimum acceptable offer strategies). Tracking this many strategies is computationally intensive, but it was possible to identify a stable equilibrium.[55]

[54]The higher mutation rate is presumed to be necessary to induce more diversity in the populations so as to generate more instances of rare types of interactions. The high z (more so than a high α) appears to be necessary to increase the probability of observing multiple signals at the same time.

[55]The time to compute each τ-step is much longer, but also the number of time periods of simulation needed to reach a stable equilibrium was dramatically larger. I terminated the computer simulation after running for approximately 57 hours, with the maximum τ-step-change

If signal set ___ observed,	which predicts that responder's MAO is ___,	then offer ___
{}	2 (93.4%) or 3 (6.2%)	2
{(1,0)}	2 (65.1%) or 3 (34.9%)	3
{(1,1)}	1	1
{(2,0)}	3	3
{(2,1)}	2 (99.97%)	2
{(1,0),(2,0)}	3	3
{(1,0),(2,1)}	2	2
{(1,1),(2,1)}	1	1

Table 4.5.5: Sophisticated proposer strategy in game with four possible signals (medium-offer equilibrium)

The equilibrium reached (with $z = 5$, $\alpha = 0.4$, $\delta = 0.01$) is a medium-offer one with a modal offer of 2, i.e. a third of the total amount. Responders demand 2 with a frequency of 0.983 and 3 with a frequency of 0.013. The most frequent proser strategy, with a frequency of 0.869, also the one with highest expected payoff, is described in table 4.5.5. The middle column shows what a signal set predicts, given actual responder strategy and signal set frequencies per responder type. Proposers only see the signal sets, so need to learn what each one means.

The proposer strategy shown is quite sophisticated, correctly making optimal offers for every possible signal set. The most interesting case is when the $\{(1,0)\}$ signal set is observed, which shows the proposer knows that the responder had rejected an offer of 1 in the past, but nothing else. One could think that the best offer to make in this case would be 2, but, surprisingly, the proposer offers 3 instead. The middle column reveals the reason: while it is indeed most likely that the responder's MAO is 2, the probability that it is 3 is high enough at 34.9% that the proposer would be facing too high a risk of rejection by offering 2.[56] The $\{(1,0)\}$ signal set predicts a MAO of 3 with such

in all frequencies having reduced to below 0.0000001 (but not below 0.00000001 as before), which is good enough to indicate a stable equilibrium has indeed been found. Among other possible optimisations, the model could be simplified by eliminating proposer strategies that make different offers for strategy sets $\{(2,0)\}$ and $\{(1,0),(2,0)\}$, as they contain the same effective information given the MAO assumption for responder strategies, or similarly $\{(1,1)\}$ and $\{(1,1),(2,1)\}$, although one might also be interested to see if proposers can learn this for themselves.

[56] Offering 2 instead of 3 gains 1 with probability 0.651 and loses 3 with probability 0.349, which results in an expected loss of 0.396. The actual expected payoff advantage of offering 3 rather than 2 when observing $\{(1,0)\}$ is slight (3.9752 versus 3.9750) given the low frequency of responders with a MAO of 3 (1.3%). If mutation rates were higher or signals were observed

a high probability because responders with a MAO of 2 are highly likely to be offered 2, thus exhibiting alternative signal sets $\{(2,1)\}$ or $\{(1,0),(2,1)\}$ with high probability. It is therefore the absence of $(2,1)$ rather than the presence of $(1,0)$ in this information set that interests proposers.[57] Proposers have more accurate information on responders with a MAO of 2 than on those with a MAO of 3, and evolution has endowed them with behaviour that seems to recognize this.

As these examples show, there are many interesting opportunities to explore, though further development of practical methods to simulate large models with many strategies will be needed. The main limitation is the large number of proposer strategies that can be generated from even a small number of possible signals and offer amounts. Further progress will likely require sensible restrictions on proposer strategy sets, and methods to model only a small set of possible strategies at a time.

4.6 Conclusion

Many decades ago, Thomas Schelling (1956) explained that the need to maintain one's reputation can act as a powerful commitment device. A bargaining party is supposed to recognize that their opponent, concerned with her reputation, will not accept an inferior offer, because the cost of damaging a good reputation outweighs the cost of rejecting the offer. The opponent, sophisticated enough to know that her commitment will be recognized as credible for this reason, can increase her prospects of success by emphasizing and reinforcing the logic of her strategy during the bargaining process.

The adaptive, heuristic behaviour that evolutionary models seek to explain is much less sophisticated: here, agents simply learn through experience or imitation, or know from received norms or instincts that have evolved during some past time, that it is better to make more reasonable offers, and to reject bad ones. Both kinds of theory, however, encounter the same paradox, which Schelling identified but did not fully resolve: successful threats, unlike promises of reward, do not need to be executed. The conundrum is that it is not clear how a reputation can then be established. The opportunity to prove one's sincerity is inseparable from the very outcome one is trying to avoid. In the context of international military conflict, Lieberman (1994, p. 416) spells the implication out explicitly: some instances of failure, even *repeated* failures, are a necessary condition for successful deterrence.[58]

noisily, however, such sophistication could render greater rewards.

[57] Similarly, the probability of the responder being a MAO 3 type, conditional on observing the empty signal set {} is 6.1%, even though their overall frequency is only 1.3%. In this case, however, the probability is not high enough to incentivize proposers to offer 3 instead of 2.

[58] One should hope that war is not the *only* effective way to achieve peace. I would not make

The analysis and simulations results in this chapter support similar conclusions. A theory in which the outcome hinges on reaction to a demonstrated commitment to a specific course of action, i.e. reputations based on facts rather than cheap talk, cannot be complete unless the conditions needed to generate and communicate a track record are addressed. If those conditions are systematically linked to the outcome of the model, the model must have endogenous information in the sense that I have used it. If the outcome that generates useful information is an adverse one that is supposed to be deterred by reputation, then some positive frequency of deterrence failure must occur alongside successes. The questions then arise whether reputations can generate enough information, given endogeneity of information, for them to be effective, whether such behaviour can be learned and whether such outcomes can be stable.

The general framework developed in section 4.3 defines an endogenous information equilibrium for a population game where the information generated by a pattern of observable behaviour (actions) is consistent with the behaviour induced by the information for some distribution of information-contingent strategies. In the framework, information consists of a probability distribution over signal sets for each possible strategy of the first player. A signal set is a collection of signals, which are observations of specific action profiles involving the second player. The framework allows the first player to receive multiple signals of different kinds simultaneously and react to their combination.

By considering all possible action profiles as potential signals, a more holistic view of reputation applicable to the ultimatum game is formed. This leads to the identification of two kinds of reputation even in a simple minigame with only two possible offer amounts, namely negative reputations (knowledge of having accepted a low offer) and positive reputations (knowledge of having rejected a low offer), which improves on existing analyses that considers only negative reputations.[59]

The negative-reputation minigame (sections 4.4 and 4.5.3) results in a low-offer or a high-offer equilibrium under particularly weak or strong informational parameters (sample size z and observation probability α) respectively, but there is a large range of intermediate infomration parameter values that result in bistable systems with both low- and high-offer equilibria. Paradox-

such a claim, nor do I claim that, in the ultimatum game, reputation is the only mechanism that can ensure fair offers. Theories can be regarded as complete not when they answer every question, but when they are self-consistent and answer at least more than they ask.

[59] Akdeniz and Van Veelen (2023) do allow responders to raise their offers above their defaults if they learn a responder's MAO is higher, so their model includes negative and positive reputations. They do not identify or investigate them as such, nor do they offer a deterministic version of their model that allows a comprehensive analysis of equilibria and stability. Zhang et al. (2023) assign a "good" reputation whenever a responder rejects a low offer and a "bad" reputation whenever a responder accepts a low offer, but their model still has only two possible reputations (because there is never "no reputation"), so their effects cannot be identified independently.

ically, in the low-offer equilibrium, proposers have strong information about responders, but in the high-offer equilibria, they have very weak information. But the payoff structure of the game prevents a low-offer equilibrium from being escaped unless information is almost perfect, while on the other hand a high-offer equilibrium can be sustained by very little information. In the latter, the addition of a small mutation rate generates occasional random low offers, which provides just enough information to provide the necessary incentives to proposers to make high offers and responders to reject low offers.

Positive reputations (section 4.5.4) give completely different results. In contrast to negative reputations, positive reputations can be used effectively to get out of low-offer equilibria, thus allowing relatively "fair" outcomes to emerge without requiring prior behavioural norms. But these models show an inherent instability, because positive reputations, once established, have a strong negative effect on their own effectiveness: if responders are successful in deterring proposers from making low offers to them, they do not get opportunities to prove that they would reject them. The effect of a lost reputation is immediate and severe, because the default offer of prosers in this case is low, unlike in the negative reputation model where the default offer is high and *avoiding* a negative reputation when high offers are made is comparatively easy.[60] The positive reputation models have only mixed-strategy equilibria, and the equilibria are often unstable, leading to dynamic systems with limit cycles and oscillations. They also feature a significant positive rate of rejections, up to 16.6% in the results presented, which are inefficient but not necessarily unrealistic. Finally, I combine both reputation types in a single model and find that they have complementary effects: positive reputations can bootstrap the system out of a low-offer situation, while negative reputations are then useful to stabilize a high-offer equilibrium and improve efficiency.

These results contribute to the understanding of experimental data, where relatively equal divisions in the ultimatum game are commonly found. Further work is needed to analyse games with larger signal and strategy spaces. I briefly explored a few models in the last section, indicating that sophisticated strategies exhibiting subtle strategic uses of multiple kinds of information can evolve. The endogenous information framework that has been developed in this chapter is not tied to the ultimatum game or any specific kind of evolutionary dynamics so could easily be applied to any other situation where the first player reacts to the distribution of past play of the second player, including other types of bargaining interactions and traditional signalling applications.

[60] Lack of stable high-offer equilibria in the negative reputation model under weak information conditions arise due to a more indirect effect: if L-responders cannot be detected, the relative payoff advantage to H-responders are diminished, leading to a high frequency of L-responders invading and consequently HL-proposers being replaced by LL-proposers.

Chapter 5

Conclusion

Over the decades, the ultimatum game (UG) has emerged as one of the most extensively studied games, rivalling the prisoners' dilemma. Both games are characterized by their inherent simplicity, yet they yield paradoxical and unexpected outcomes. In the prisoners' dilemma, the paradox lies in the fact that the strategy that is individually optimal is detrimental to the collective, prompting generations of students to think deeply about the meaning of rational choices that lead to lower payoffs. In the UG, the paradoxical result is that players are expected to behave in ways that are generally perceived as highly self-serving and unfair, which has prompted deeper inquiries into human motives and social behaviour, leading to much greater understanding of these issues. A substantial body of experimental research have shown that individuals do not often behave in the manner predicted by the theory; rather, they appear to be concerned about the impact of their choices on others (Cooper and Kagel, 2016).

Yet there is another kind of paradox in the UG, which is that the selfish result that seems to follow very straightforwardly from basic game theory may not actually be a reasonable prediction, even for purely selfish people, if those people are imperfectly rational. In parallel to the ongoing experimental research, there has also been a smaller, but significant research stream considering evolutionary explanations for behaviour in the UG. It is not always recognized explicitly, but evolutionary models can make two distinct contributions: they can either serve as a crude proxy for boundedly rational learning in interactive situations, or they can generate simplified accounts of historical genetic and cultural evolutionary processes that have over time shaped human behavioural norms and social preferences.

Keeping these dual uses of evolutionary game theory in mind, I have used the model of Gale, Binmore, and Samuelson (1995) (GBS) as a baseline model to explore different research questions aimed at providing and critically interpreting explanations from evolutionary models for empirical results. The main result of GBS[1] is that the evolutionary dynamics can come to rest at an asymptotically stable evolutionary equilibrium where proposers in the UG make offers substantially higher than the minimum to responders, and responders reject low offers lower than the modal one. GBS add mutational noise to their model, at a low rate, showing that the shape of the noise is important for obtaining stable imperfect equilibria. But it would be wrong to think of the result as being *caused* by the noise, given that the added noise can be vanishingly small, and the result still obtained. Hence, I refer to such models as baseline models.

As I conclude at the end of chapter 3, the value of the GBS model (and similar models) and its results is not so much about the stability or instability of any particular equilibrium point, but rather what we learn about the UG's strategic structure: responders learn more slowly than proposers, due to differing payoff gradients, and overall evolution of the system in the vicinity of subgame-imperfect Nash equilibria is very slow. There is empirical support for these features from experiments in which players can gain experience.[2]

5.1 Using minigames to explain imperfect outcomes in the ultimatum game

Mailath (1998, p. 1349) makes the general point that evolutionary models can be valuable in illuminating the strategic structure of games. One could go further and argue that a good understanding of a games' evolutionary dynamics is necessary for a good understanding of its strategic structure. This is the departure point in chapter 2, where I argue that it has not previously been established that the minigames that have been used (by GBS and others) to analyse the UG's dynamics are adequate for this purpose, hence it is not clear that the full UG's strategic structure is well understood.

There is reason to be sceptical that the dynamic analysis for the minigame can be transferred to a larger game with larger strategy sets. In the minigame, an essential part of the explanation for the GBS result is that mutation keeps responder rejection of low offers at a high enough frequency for the equilibrium to be maintained. I show that maintaining such an equilibrium in the full game requires a specific responder strategy, namely the one that would reject a slightly lower offer than the current modal offer, to be above a given frequency.

[1] As indicated, the same result appears in Binmore and Samuelson (1994) and Roth and Erev (1995), the former using a the alternate version of the replicator dynamics, the so-called "adjusted" or discrete replicator dynamics, and the latter using reinforcement learning. It was useful to be able to work with a specific aggregate dynamics specification, but there is no reason to think that any of my findings would not apply in the alternative, or any similar, dynamics. A part of the research in chapter 3, however, specifically concerns the difference between the two variants of the replicator dynamics.

[2] See references in chapter 3.

But the required frequency is higher than the mutation target for this frequency, suggesting that mutation is pulling in the wrong direction for the explanation to work.

To resolve this puzzle and relate the minigame's analysis of dynamics to the full game, a more rigorous approach is developed, relying on the analysis of the full game's dynamics in the form of conditional frequencies. The particular structure of the ultimatum game causes dynamics in this conditional strategy space to be approximately independent of higher-amount strategy frequencies when the full system is near an associated equilibrium. The conditional dynamics take on the same form as those in a full ultimatum game with reduced strategy sets. Moreover, it is established that the dynamics of any particular conditional strategy are almost entirely determined by the conditional frequency of the corresponding strategy in the opposite population, and thus that a two-dimensional analysis based on the ultimatum minigame is feasible and appropriate. The minigame explanation works as expected on conditional frequencies; in particular mutation pulls the critical responder strategy in the same direction as in the minigame. The reason it does not work directly with unconditional frequencies is that there are indirect flows in the full game, which ultimately support the frequency of the required responder strategy, that are not easily identified with direct (naive) analysis.

A better understanding is also gained of the factors affecting the difficulty (i.e. the minimum required ratio of responder to proposer mutation rates) of stabilizing any particular imperfect equilibrium. Technical advances in this chapter include development of the technique of conditional analysis (which is likely to be useful in other contexts) and the graphical analysis based on selection-mutation equilibrium loci, which proved useful in isolating the forces of selection in mutations for each population in the dynamics.

5.2 Stochastic learning and emulation in the ultimatum game

Various evolutionary game theorists have referred to the possibility that the behavioural norms and preferences exhibited by subjects in experiments may have originated in the distant past through genetic evolution, or in the somewhat-less-distant past through cultural evolutionary processes. In chapter 3, I investigate whether the GBS result may be sensibly interpreted as a model of cultural evolution. Since the GBS model and its replicator dynamics equations describing aggregate strategy frequency dynamics remain the same, this is a question of reinterpretation. This requires a defensible account of what occurs at the level of individual agents – in other words, a microfoundations model for the GBS model's noisy replicator dynamics.

I argue that the aspirational learning microfoundations model GBS use to justify interest in the replicator dynamics is not well suited to explain cultural evolutionary processes, and develop an alternative stochastic individual-level model, similar to a model in Weibull (1995, p. 158), which is based on the notion that people revise their strategies by imitating their relatively successful social peers. Interpreting the GBS model in the light of this microfoundations model leads to implications that have a somewhat negative impact on the relevance of the GBS result for empirical results.

Specifically, in the standard replicator dynamics employed by GBS, which shows the GBS result (asymptotically stable imperfect equilibria) most clearly, individual agents learn faster the higher the average payoff for their population. The adjusted, or Maynard Smith, replicator dynamics do not have this effect. This implies that the form of the model chosen by GBS favours selection for proposers relative to responders, which tends to strengthen their result. This is not problematic in itself, as there is a strong argument that proposers should pay more attention since they have more to lose from poor strategies. However, GBS also argue in favour of applying a higher mutation rate to responders than to responders, which multiplies the effect. This approach therefore ends up producing a system in which the effective balance of selection and mutation may have been pushed further in a specific direction than what might be considered reasonable. GBS's assumption of higher responder mutation rates, used in conjunction with the standard replicator dynamics, do not appear to have been challenged before.

A second negative implication is identified through agent-based simulations of finite populations based on the individual-level microfoundations. It is shown that the stability of imperfect equilibria is fragile and not robust to stochastic disturbances found in finite populations. The aggregate dynamics are a good approximation for the underlying individual-level stochastic model only for very large populations well in excess of a million agents per population.

The chapter contributes to the literature on revision protocols linked to the replicator dynamics, and illustrates that, even though different microfoundations models may lead to the same aggregate dynamics, they are not interchangeable when a specific research question is addressed, and can have substantive implications for the interpretation and relevance of a model's results.

5.3 An evolutionary perspective on good and bad reputations in the ultimatum game

If the GBS result is fragile, the long-run tendency of the baseline model is inevitably to move towards the SPNE, which is – like the imperfect equilibria – a stable asymptotic attractor, but unlike them, the SPNE tends to be robust against even relatively large stochastic disturbances. The baseline model can therefore, at most, explain why more equal divisions survive evolutionary processes for extended periods of time, but the model cannot explain how such behaviours originated. Ideally, an account is needed of how such behaviours can arise in societies where they are not initially present.

Chapter 4 is aimed at meeting this standard, using evolutionary models of the UG where proposers are able to learn about past interactions of the responders they face, in other words where responders can develop reputations regarding their behaviour. The research in this chapter represents an advance over existing literature on at least three counts: firstly, by avoiding arbitrary assumptions about availability of information, which generally fail to link such knowledge to patterns of interactions that can conceivably take place given the strategy distributions of the players; secondly, by generalizing the notion of what knowledge of past interactions can be useful to proposers; and finally by performing rigorous and comprehensive analysis of dynamics in the models.

A significant contribution of this research is the development of a flexible, general endogenous information framework for two-player sequential games. This framework can be used to analyse any two-player sequential game where the first player reacts to knowledge of the distribution of play of the second player. It is applicable to other types of bargaining interactions and traditional signalling applications. The framework defines an endogenous information equilibrium for a population game wherein the information generated by a pattern of observable behaviour (actions) is consistent with the behaviour induced by the information for some distribution of information-contingent strategies. Information consists of a probability distribution over signal sets for each possible strategy of the second player. A signal set is a collection of signals, which are observations of specific action profiles involving the second player. The framework allows the first player to receive multiple signals of different kinds at the same time, and react to their combination.

By considering all possible action profiles as potential signals, a more holistic view of reputation applicable to the ultimatum game is formed. This leads to the identification of two kinds of reputation even in a simple minigame with only two possible offer amounts, namely negative reputations (knowledge of having accepted a low offer) and positive reputations (knowledge of having rejected a low offer). This improves on existing analyses that consider only negative reputations.

A series of minigame models generated using the general framework is explored in detail using a combination of explicit solutions and deterministic simulations based on the standard replicator dynamics. Endogenous information is shown to be crucial. When responders successfully use their reputations to deter low offers from proposers, opportunities to build reputations become limited. This indicates that equilibria relying on reputations naturally raise questions of stability. However, reputations can result in stable equilibria with relatively equal outcomes, i.e. in the minigame, equilibria where proposers make the high offer and responders would reject the low offer.

In the model with only negative reputations, bistable dynamic systems are found for a wide range of parameter values – both high-offer (i.e. equal division) and low-offer (i.e. the SPNE outcome) equilibria are stable. Paradoxically, in the low-offer equilibrium, proposers have strong information about responders, but in the high-offer equilibrium, they have very weak information. Despite this, the payoff structure of the game prevents a low-offer equilibrium from being escaped, while a high-offer equilibrium can be sustained by very little information. In contrast to negative reputations, positive-reputation models can evolve out of low-offer, low-demand states, thus providing a viable theory for how behavioural norms resulting in equal divisions could have emerged where they did not exist before. In models with only positive reputations, however, there is only a mixed-strategy equilibrium, leading to dynamic systems with oscillations. When both types of reputation are combined in the same model, they have complementary roles: positive reputations can bootstrap the system out of a low-offer state, while negative reputations are then useful to stabilize a high-offer equilibrium and improve efficiency. The final part of the results section briefly explores larger models with more complex information structures, showing that relatively sophisticated proposer strategies can evolve, though further work is needed in this area.

5.4 Limitations and opportunities for future work

An obvious limitation of the research in this dissertation is that it has considered only a single game, the ultimatum game. While the UG has its own particular evolutionary dynamics, which has been analysed in some detail, the ultimatum game is just one kind of interaction, and we are ultimately interested in features of games and behaviour that reflect real-world interactions, past and present. We have learned that many of the features of human behaviour that motivates interest in the ultimatum games, such as other-regarding preferences, reciprocity and reputations, are not limited to a single game.

An interesting area that has been explored by only a few researchers such as Samuelson (2001) and Zollman (2008) is to look at evolution in complex environments in which players engage in more than one type of game. These articles take seriously an idea that has also influenced this dissertation, namely that behavioural norms exhibited in a given type of game may have developed in evolutionary context where different kinds of interactions have taken place. Much further work is needed to clarify these ideas and produce defensible theories. There is a danger in this work, as in evolutionary theorizing generally, of coming up with too many theories and devoting too little attention to empirical validation of these theories. This concern also applies to the work done here, particularly in relation to the chapter on reputations in the ultimatum game. Further attention given to this aspect would therefore be warranted.

There may be opportunities to explore the application of the conditional frequency dynamics developed in chapter 2 to other games. Particular candidates of interest would be games with money-amount strategy sets, e.g. the Nash demand game, games with recursive structure such as the centipede game or more extensive alternating-offer bargaining games, and possibly extensive-form games more generally. The benefit would be the ability to isolate frequency flows within and between subsets of strategies of particular interest, as in chapter 2.

It is evident that the work started in chapter 4 can be taken further, both in terms of considering reputation in larger ultimatum games with more complex strategies, and in terms of applying the endogenous framework to other games. While the framework defines a population game, suggesting an evolutionary context, it is not dependent on any particular kind of evolutionary dynamics, so a more systematic and general treatment should be worthwhile.

More generally, there may be areas of real-world applications other than traditional experimental work where some notions such as noisy evolution (as in GBS) or reputation-building strategies may be able to provide useful insights. One area of interest is the case where firms set their prices using algorithms or artificial intelligence (AI) systems, in which case it would not be surprising to see implicit collusion (Calvano et al., 2020) and even simple implicit bargaining processes to divide the rents that may arise. Understanding whether these digital behaviours resemble the theoretical learning and reputation-building behaviours described in this dissertation, could open avenues for exploring the evolution of conventions and reputation-building behaviours in such systems. Potentially, insights from evolutionary game theory could also guide the development of regulatory frameworks aimed at achieving efficient and equitable outcomes in such contexts. These connections suggest a broader relevance of evolutionary models of bargaining processes to novel real-world phenomena.

List of references

Abbink, K., G. E. Bolton, A. Sadrieh, and F.-F. Tang (2001). Adaptive learning versus punishment in ultimatum bargaining. *Games and Economic Behavior 37*(1), 1–25.

Aina, C., P. Battigalli, and A. Gamba (2020). Frustration and anger in the ultimatum game: An experiment. *Games and Economic Behavior 122*, 150–167.

Akdeniz, A. and M. Van Veelen (2021). The evolution of morality and the role of commitment. *Evolutionary Human Sciences 3*(e41).

Akdeniz, A. and M. Van Veelen (2023). Evolution and the ultimatum game. *Games and Economic Behavior 142*, 570–612.

Andreoni, J. and J. Miller (2002). Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica 70*(2), 737–753.

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

Baumard, N. and D. Sperber (2010). Weird people, yes, but also weird experiments. *Behavioral and Brain Sciences 33*(2–3), 84–85.

Binmore, K. and L. Samuelson (1994). An economist's perspective on the evolution of norms. *Journal of Institutional and Theoretical Economics 150*(1), 45–63.

Binmore, K. and L. Samuelson (1997). Muddling through: Noisy equilibrium selection. *Journal of Economic Theory 74*(2), 235–265.

Binmore, K. G. and L. Samuelson (1999). Evolutionary drift and equilibrium selection. *The Review of Economic Studies 66*(2), 363–393.

Binmore, K. G., L. Samuelson, and R. Vaughan (1995). Musical chairs: Modeling noisy evolution. *Games and Economic Behavior 11*(1), 1–35.

Björnerstedt, J. and J. Weibull (1995). Nash equilibrium and evolution by imitation. In K. J. Arrow, E. Colombatto, M. Perlman, and C. Schmidt (Eds.), *The Rational Foundations of Economic Analysis*, Chapter 7, pp. 155–171. MacMillan, London.

Blanco, M., D. Engelmann, and H. T. Normann (2011). A within-subject analysis of other-regarding preferences. *Games and Economic Behavior 72*(2), 321–338.

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

Boylan, R. T. (1995). Continuous approximation of dynamical systems with randomly matched individuals. *Journal of Economic Theory 66*(2), 615–625.

Brenner, T. and N. J. Vriend (2006). On the behavior of proposers in ultimatum games. *Journal of Economic Behavior & Organization 61*(4), 617–631.

Burke, M. A. and H. P. Young (2011). Social norms. In A. Bisin, J. Benhabib, and M. Jackson (Eds.), *Handbook of Social Economics*, Volume 1, Chapter 8, pp. 311–338. Elsevier.

Calvano, E., G. Calzolari, V. Denicolo, and S. Pastorello (2020). Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review 110*(10), 3267–3297.

Camerer, C. and R. H. Thaler (1995). Anomalies: Ultimatums, dictators and manners. *The Journal of Economic Perspectives 9*(2), 209–219.

Charness, G. and M. Rabin (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics 117*(3), 817–869.

Cooper, D. J. and E. G. Dutcher (2011). The dynamics of responder behavior in ultimatum games: a meta-study. *Experimental Economics 14*(4), 519–546.

Cooper, D. J. and J. H. Kagel (2016). Other-regarding preferences. In J. H. Kagel and A. E. Roth (Eds.), *The Handbook of Experimental Economics, Volume 2*, Chapter 4, pp. 217–289.

Cox, J. C., D. Friedman, and S. Gjerstad (2007). A tractable model of reciprocity and fairness. *Games and Economic Behavior 59*(1), 17–45.

Cressman, R. (1995). Evolutionary game theory with two groups of individuals. *Games and Economic Behavior 11*(2), 237–253.

Debove, S., N. Baumard, and J.-B. André (2016). Models of the evolution of fairness in the ultimatum game: a review and classification. *Evolution and Human Behavior 37*(3), 245–254.

Dellarocas, C. (2006). Reputation mechanisms. In T. Hendershott (Ed.), *Handbook on Economics and Information Systems, Volume 1*, Chapter 13, pp. 629–660. Emerald Publishing Limited.

Easley, D. and A. Rustichini (1999). Choice without beliefs. *Econometrica 67*(5), 1157–1184.

Fehr, E. and U. Fischbacher (2003). The nature of human altruism. *Nature 425*(6960), 785–791.

Fehr, E. and S. Gächter (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives 14*(3), 159–181.

Fehr, E. and K. M. Schmidt (2006). The economics of fairness, reciprocity and altruism – experimental evidence and new theories. In S.-C. Kolm and J. M. Ythier (Eds.), *Handbook of the Economics of Giving, Altruism and Reciprocity*, Volume 1, Chapter 8, pp. 615–691.

Forber, P. and R. Smead (2014). The evolution of fairness through spite. *Proceedings of the Royal Society B 281*.

Frank, R. H. (1988). *Passion Within Reason*. W Norton.

Friedman, D. (1998). On economic applications of evolutionary game theory. *Journal of Evolutionary Economics 8*(1), 15–43.

Fudenberg, D. and D. K. Levine (1998). *The Theory of Learning in Games*. MIT press.

Gale, J., K. G. Binmore, and L. Samuelson (1995). Learning to be imperfect: The ultimatum game. *Games and Economic Behavior 8*(1), 56–90.

Güth, W. (1995). On ultimatum bargaining experiments – a personal review. *Journal of Economic Behavior & Organization 27*(3), 329–344.

Güth, W. and M. G. Kocher (2014). More than thirty years of ultimatum bargaining experiments: Motives, variations, and a survey of the recent literature. *Journal of Economic Behavior & Organization 108*(Supplement C), 396–409.

Güth, W. and S. Napel (2006). Inequality aversion in a variety of games–an indirect evolutionary analysis. *The Economic Journal 116*(514), 1037–1056.

Güth, W., R. Schmittberger, and B. Schwarze (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization 3*(4), 367–388.

Harms, W. (1997). Evolution and ultimatum bargaining. *Theory and Decision 42*, 147–175.

Harsanyi, J. C. (1961). On the rationality postulates underlying the theory of cooperative games. *Journal of Conflict Resolution 5*(2), 179–196.

Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, R. McElreath, M. Alvard, A. Barr, J. Ensminger, N. S. Henrich, K. Hill, F. Gil-White, M. Gurven, F. W. Marlowe, J. Q. Patton, and D. Tracer (2005). "Economic Man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences 28*(06), 795–815.

Hofbauer, J. and K. H. Schlag (2000). Sophisticated imitation in cyclic games. *Journal of Evolutionary Economics 10*(5), 523–543.

Hofbauer, J. and K. Sigmund (2003). Evolutionary game dynamics. *Bulletin of the American mathematical society 40*(4), 479–519.

Hoffman, E., K. McCabe, and V. L. Smith (1996). Social distance and other-regarding behavior in dictator games. *The American Economic Review 86*(3), 653–660.

Huck, S. and J. Oechssler (1999). The indirect evolutionary approach to explaining fair allocations. *Games and Economic Behavior 28*(1), 13–24.

Jackson, A. R. (2005). Trade generation, reputation, and sell-side analysts. *The Journal of Finance 60*(2), 673–717.

Kandori, M., G. J. Mailath, and R. Rob (1993). Learning, mutation, and long run equilibria in games. *Econometrica 61*(1), 29–56.

Konigstein, M. and W. Muller (2000). Combining rational choice and evolutionary dynamics: The indirect evolutionary approach. *Metroeconomica 51*(3), 235–256.

Kreps, D. M., P. Milgrom, J. Roberts, and R. Wilson (1982). Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory 27*(2), 245–252.

Kreps, D. M. and R. Wilson (1982). Reputation and imperfect information. *Journal of Economic Theory 27*(2), 253–279.

Lieberman, E. (1994). The rational deterrence theory debate: Is the dependent variable elusive? *Security Studies 3*(3), 384–427.

List, J. A. and T. L. Cherry (2000). Learning to accept in ultimatum games: Evidence from an experimental design that generates low offers. *Experimental Economics 3*, 11–29.

Mago, S. D., J. Pate, and L. Razzolini (2024). Experimental evidence on the role of outside obligations in wage negotiations. *Journal of Economic Behavior & Organization 219*, 528–548.

Mailath, G. J. (1998). Do people play Nash equilibrium? Lessons from evolutionary game theory. *Journal of Economic Literature 36*(3), 1347–1374.

Mailath, G. J. and L. Samuelson (2006). *Repeated Games and Reputations: Long-run Relationships*. Oxford university press.

Mäs, M. and H. H. Nax (2016). A behavioral study of "noise" in coordination games. *Journal of Economic Theory 162*, 195–208.

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

Milgrom, P. and J. Roberts (1982). Predation, reputation, and entry deterrence. *Journal of Economic Theory 27*(2), 280–312.

Milgrom, P. R., D. C. North, and B. R. Weingast (1990). The role of institutions in the revival of trade: The law merchant, private judges, and the champagne fairs. *Economics & Politics 2*(1), 1–23.

Napel, S. (2003). Aspiration adaptation in the ultimatum minigame. *Games and Economic Behavior 43*(1), 86–106.

Nowak, M. A., K. M. Page, and K. Sigmund (2000). Fairness versus reason in the ultimatum game. *Science 289*(5485), 1773–1775.

Oosterbeek, H., R. Sloof, and G. Van De Kuilen (2004). Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Experimental economics 7*, 171–188.

Page, K. M. and M. A. Nowak (2002). Unifying evolutionary dynamics. *Journal of Theoretical Biology 219*(1), 93–98.

Parry, H. R. and M. Bithell (2011). Large scale agent-based modelling: A review and guidelines for model scaling. In A. J. Heppenstall, A. T. Crooks, L. M. See, and M. Batty (Eds.), *Agent-Based Models of Geographical Systems*, Chapter 14, pp. 271–308. Springer.

Peters, R. (2000). Evolutionary stability in the ultimatum game. *Group Decision and Negotiation 9*(4), 315–324.

Posch, M. (1997). Cycling in a stochastic learning algorithm for normal form games. *Journal of Evolutionary Economics 7*(2), 193–207.

Posner, R. A. (1978). Privacy, secrecy, and reputation. *Buffalow Law Review 28*(1), 1–56.

Poulsen, A. U. (2007). Information and endogenous first mover advantages in the ultimatum game: An evolutionary approach. *Journal of Economic Behavior & Organization 64*(1), 129–143.

Poulsen, A. U. and J. H. Tan (2007). Information acquisition in the ultimatum game: An experimental study. *Experimental Economics 10*, 391–409.

Rand, D. G., C. E. Tarnita, H. Ohtsuki, and M. A. Nowak (2013). Evolution of fairness in the one-shot anonymous ultimatum game. *Proceedings of the National Academy of Sciences 110*(7), 2581–2586.

Riley, J. G. (1979). Informational equilibrium. *Econometrica*, 331–359.

Rogerson, W. P. (1983). Reputation and product quality. *The Bell Journal of Economics 14*(2), 508–516.

Roth, A. E. and I. Erev (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior 8*(1), 164–212.

Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica*, 97–109.

Samuelson, L. (1997). *Evolutionary Games and Equilibrium Selection*. MIT press.

Samuelson, L. (2001). Analogies, adaptation, and anomalies. *Journal of Economic Theory 97*(2), 320–366.

Sandholm, W. H. (2010). *Population Games and Evolutionary Dynamics*. MIT press.

Sandholm, W. H. (2015). Population games and deterministic evolutionary dynamics. In H. P. Young and S. Zamir (Eds.), *Handbook of Game Theory with Economic Applications*, Volume 4, Chapter 13, pp. 703–778. Elsevier.

Sandholm, W. H. (2020). Evolutionary game theory. In D. Pérez-Castrillo, M. Sotomayor, and F. Castiglione (Eds.), *Complex Social and Behavioral Systems: Game Theory and Agent-Based Models*, pp. 573–608. Springer.

Sanfey, A. G., J. K. Rilling, J. A. Aronson, L. E. Nystrom, and J. D. Cohen (2003). The neural basis of economic decision-making in the ultimatum game. *Science 300*(5626), 1755–1758.

Schelling, T. C. (1956). An essay on bargaining. *The American Economic Review 46*(3), 281–306.

Schlag, K. H. (1998). Why imitate, and if so, how?: A boundedly rational approach to multi-armed bandits. *Journal of Economic Theory 78*(1), 130–156.

Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory 4*(1), 25–55.

Selten, R. (1990). Bounded rationality. *Journal of Institutional and Theoretical Economics 146*(4), 649–658.

Shirata, Y. (2012). The evolution of fairness under an assortative matching rule in the ultimatum game. *International Journal of Game Theory 41*(1), 1–21.

Sigmund, K., C. Hauert, and M. A. Nowak (2001). Reward and punishment. *Proceedings of the National Academy of Sciences 98*(19), 10757–10762.

Skyrms, B. (2014). *Evolution of the Social Contract*. Cambridge University Press.

Slonim, R. and A. E. Roth (1998). Learning in high stakes ultimatum games: An experiment in the Slovak Republic. *Econometrica 66*(3), 569–596.

Smith, V. L. and B. J. Wilson (2018). Equilibrium play in voluntary ultimatum games: Beneficence cannot be extorted. *Games and Economic Behavior 109*, 452–464.

Ståhl, I. (1972). *Bargaining Theory*. Economic Research Institute at the Stockholm School of Economics.

Taylor, P. D. and L. B. Jonker (1978). Evolutionary stable strategies and game dynamics. *Mathematical Biosciences 40*(1), 145–156.

Uriarte, J. R. (2007). A behavioural foundation for models of evolutionary drift. *Journal of Economic Behavior & Organization 63*(3), 497–513.

Van Damme, E. (1991). *Stability and Perfection of Nash Equilibria*. Springer.

Vriend, N. J. (1997). Will reasoning improve learning? *Economics Letters 55*(1), 9–18.

Weibull, J. W. (1995). *Evolutionary Game Theory*. MIT press.

Winter, E. and S. Zamir (2005). An experiment with ultimatum bargaining in a changing environment. *The Japanese Economic Review 56*, 363–385.

Young, H. P. (1993). The evolution of conventions. *Econometrica 61*(1), 57–84.

Young, H. P. (1996). The economics of convention. *Journal of Economic Perspectives 10*(2), 105–122.

Young, H. P. and M. A. Burke (2001). Competition and custom in economic contracts: a case study of illinois agriculture. *American Economic Review 91*(3), 559–573.

Zhang, B. (2013). Social learning in the ultimatum game. *PLoS One 8*(9), e74540.

Zhang, Y., S. Yang, X. Chen, Y. Bai, and G. Xie (2023). Reputation update of responders efficiently promotes the evolution of fairness in the ultimatum game. *Chaos, Solitons & Fractals 169*, 113218.

Zollman, K. J. (2008). Explaining fairness in complex environments. *Politics, Philosophy & Economics 7*(1), 81–97.