Social norms and learning in games



Sam Jindani

Balliol College

A thesis submitted for the degree of
Doctor of Philosophy in Economics

Hilary 2017

Supervisors: H. Peyton Young and Thomas W. L. Norman

Word count: approximately 34,000 words, taking p. 7 as typical page

Contents

Abstract **4**

Acknowledgements **5**

Introduction **6**

1 Duelling **12**
1.1 Introduction 12
1.2 Model 20
1.3 Equilibrium 27
1.4 Discussion 32
1.A Proofs 33

2 Community enforcement using modal actions **41**
2.1 Introduction 41
2.2 Related literature 47
2.3 Repeated matching games with multiple partners 49
2.4 Observable modal actions 53
2.5 Verifiable disclosure 61
2.6 Conclusion 68

3 Learning repeated-game strategies **70**
3.1 Introduction 70
3.2 The learning rule 75
3.3 The result 81
3.4 Alternative specifications 90
3.5 Conclusion 97

Bibliography **99**

Abstract

Duelling
The norm of duelling endured for hundreds of years in Europe. In the United Kingdom it disappeared abruptly in the mid-nineteenth century, whereas in France it declined slowly. I present a simple model of social norms that explains these phenomena. The model predicts that the evolution of norms is characterised by *tipping*, whereby norms can shift suddenly due to shocks, and by a *ratchet effect*, whereby changes in parameters can cause norms to decline gradually. I show that the model can be supported by an equilibrium of a repeated game, with no special assumptions about preferences.

JEL Classification: C70, C72, C73, Z13

Keywords: duelling, social norms, subgame-perfect equilibrium, ratchet effect, tipping

Community enforcement using modal actions
I prove two folk theorems for repeated games with random matching. A large group of players is rematched at random each period, so that players who deviate must be sanctioned by third parties. Previous analyses have either relied on strong assumptions about information transmission, or have been limited to equilibria that are not robust to noise or in which players are indifferent. I use a simple construction based on modal actions to obtain results for strict and robust equilibria.

JEL classification: C70, C72, C73

Keywords: repeated games, random matching, community enforcement, modal actions

Learning repeated-game strategies
The literature on boundedly rational learning has tended to focus on stage-game actions. I present a stochastic learning rule for repeated-game strategies. Players form beliefs about their opponent's strategy based on past actions and best-respond. Occasionally, they make mistakes and experiment, and I show that the equilibrium selected depends on exactly how players make mistakes. Simple specifications of the learning rule yield intuitive selection results: the maxmin, or Rawlsian, outcome; the Nash bargaining solution; the maximum of the sum of payoffs; and a generalisation of risk dominance.

JEL classification: C70, C72, C73

Keywords: repeated games, learning, bounded rationality, stochastic stability

Acknowledgements

I thank my supervisors, Peyton Young and Tom Norman, for their sound guidance. I also thank Yuval Heller, Meg Meyer, and Bill Zame for their insightful comments, as well as seminar participants at the Transatlantic Theory Workshop at Northwestern University and at the Gorman Workshop, the DPhil and Postdoc Workshop, and the Learning, Games and Networks Seminar at the University of Oxford. I gratefully acknowledge financial support from the Economic and Social Research Council and from the University of Oxford Department of Economics.

For their comments and good humour, I am grateful to my fellow students Gregorio Curello, Manuel Fernandez, Kris Hammarbäck, Claudia Herresthal, Josh Lanier, Carlo Pizzinelli, and Anna Sanktjohanser. Finally, I would like to thank my parents.

Introduction

This thesis applies game theory to the emergence and stability of social norms and group conventions. In chapter 1, I consider the norm of duelling and argue that its historical evolution can be accounted for by a simple game-theoretical model. Duelling was prevalent in parts of Europe from the Renaissance up to the late nineteenth century. In the United Kingdom, the practice disappeared abruptly in the mid-nineteenth century. In France, it declined gradually and endured up to the turn of the century. I consider a model in which players choose a more or less costly behaviour, which represents how often they duel, how dangerous the duels they participate in are, and so on. Those who deviate from the average behaviour are subject to exogenous sanctions. I show that this results in *tipping*, as in Schelling 1971 and Young 1998, and in a *ratchet effect*. Tipping occurs when the average behaviour shifts suddenly due to shocks to players' expectations. The ratchet effect is brought on by changes in parameters. For instance if players become less sensitive to sanctions, costly behaviours become unsustainable and the norm will decline. However if players become more sensitive to sanctions, the norm is unchanged. I argue that these dynamics correspond to the historical evolution of duelling in the United Kingdom and in France, respectively. The main assumption of the model is that players are sanctioned for deviating from the average. In order to justify this, I show in an extension that sanctions can be implemented as part of a subgame-perfect equilibrium. I assume that in addition to choosing a behaviour each period, players can sanction one another. I construct an equilibrium in which players sanction not only those who deviate from the average behaviour, but also those

who deviate from the average sanctions.

In chapter 2, I look at how large communities enforce social norms. I consider repeated games with random matching, as introduced by Okuno-Fujiwara and Postlewaite 1995 and Kandori 1992. Each period, players are randomly matched to play a stage game. Because any two players meet infrequently, sanctions against a player who deviated must be carried out by third parties, who need to be informed that the player deviated. This is what Kandori termed the problem of *community enforcement*. I show that if players are matched with multiple partners each period, deviations can be identified with little information, using modal actions. To find out if a player deviated in a given period, one can compare her action with the actions of other players in the same period – specifically, one can check whether she sanctioned someone when others did not, or failed to sanction someone when others did. In contrast, when players are matched with a single opponent each period, they may require a large amount of information to determine whether a given opponent deviated. I prove two folk theorems under different informational assumptions. First, I assume that players can observe their partners' actions and the corresponding *modal actions*. I construct an equilibrium in which players sanction opponents who were non-modal in the recent past. Second, I assume that players can verifiably disclose small amounts of information about their own stage games. I construct an equilibrium in which whenever a player deviates, the partner against whom she deviated discloses the player's action as well as the corresponding actions by the partner's other partners. Then third parties know that the player who deviated was non-modal, and sanction her.

Note that chapters 1 and 2 use a similar mechanism to sustain sanctions. In both chapters, players compare their opponents' actions to those of other players from the same period in order to determine whether they deviated. In chapter

1 players' actions are compared to the average action while in chapter 2 they are compared to the modal action, but the principle is the same. However, the results in each chapter are distinct. The results in chapter 2 are folk theorems for general games, whereas the result in chapter 1 is for a specific setting, and is used to justify an assumption of the model. Moreover, the result in chapter 1 is not an application of the results in chapter 2, since the former is about sustaining particular sanctions whereas the latter are about sustaining particular payoffs. This demonstrates the versatility of the mechanism.[1]

In chapter 3, I consider the problem of learning in repeated games. Much of the boundedly rational learning literature, for instance fictitious play and reinforcement learning, focuses on stage-game actions rather than repeated-game strategies. I present a stochastic learning rule for repeated-game strategies. In the model, players are not perfectly rational, but neither are they wholly myopic. They form a belief about their opponent's repeated-game strategy based on her past actions, and they pick a strategy that is a best response given their belief. Most of the time they play according to this strategy, but occasionally they make mistakes. When a player observes sufficiently many mistakes, she enters a phase of random search. I show that in any stochastically stable state of the resulting process, players' strategies constitute a subgame-perfect equilibrium. Moreover, the payoffs players receive in the equilibrium depend on exactly how they make mistakes. Depending on the specification, the model yields the maxmin, or Rawlsian, payoffs, the Nash bargaining solution, the maximum of the sum of payoffs, or a generalisation of risk dominance. The learning rule is related to procedures by Foster and Young (2003) on the one hand, and by Young (2009) and Pradelski and Young (2012) on the other. Foster and Young consider a learning rule for repeated-game strategies and show conver-

1 I discuss other possible applications in the conclusion to chapter 2.

gence to subgame-perfect equilibrium, but do not obtain equilibrium selection results. Young (2009) and Pradelski and Young (2012) also use phases of random search to make the model tractable, but are concerned with stage-game actions.

These chapters belong to the field of evolutionary game theory. If game theory is the study of strategic interaction, evolutionary game theory could be defined as the study of strategic interaction in populations of agents who are not perfectly rational. In evolutionary game theory, agents make mistakes, form beliefs heuristically, or forget things that happened a long time ago; they might fail to project into the future or to think about how other agents will react to their actions; and so on.

Evolutionary game theory grew out of biology. Maynard Smith and Price (1973) showed how game theory could be used to model the evolution of species. In their framework, games represent the interactions of large populations of individuals. Each individual is endowed with a particular strategy, and the payoff it receives is interpreted as a measure of the individual's fitness, so that strategies that yield high payoffs should tend to become more prevalent in the population. Maynard Smith and Price defined the notion of an *evolutionarily stable strategy* or ESS, namely a strategy that is robust to invasions by other strategies. They used this concept to explain why animals might engage in limited war behaviours, whereby they refrain from inflicting serious wounds in conflict, rather than total war behaviours. Taylor and Jonker (1978) developed a full model of dynamics for this framework. As in Maynard Smith and Price's model, they assume that individuals are endowed with strategies. Individuals are matched with other members of the population to play a game, and their payoffs determine their fitness. Fitnesses in turn determine the evolution of the prevalence of each strategy in the population, according to a system of

differential equations. This model is known as the *replicator dynamics*. Taylor and Jonker showed that, in many cases, evolutionarily stable strategies are also stable points of the replicator dynamics.[2]

This approach was then applied to economics. Players were viewed not as biological individuals with fixed strategies, but as imperfectly rational or imperfectly informed economic agents. In Young's 1993 model, players are drawn at random each period to play a fixed stage game. The players observe a subsample of past outcomes and myopically best-respond given that information. They make mistakes with some small probability. Young showed that in the limit as the mistake probability tends to zero, only certain pure-strategy equilibria, known as *stochastically stable equilibria*, have positive long-run probability. He showed that in 2×2 games with two strict equilibria, the stochastically stable equilibria are the weakly risk-dominant equilibria. In Kandori, Mailath and Rob's 1993 model, players are matched with every other player in each period to play a fixed symmetric 2×2 stage game. If one strategy does better than the other, it becomes more widespread in the population the following period. As in Young's model, players occasionally make mistakes. Kandori, Mailath and Rob showed that this model also selects risk-dominant equilibria as stochastically stable.[3]

Evolutionary game theory is often viewed as a means of selection between equilibria – where deductive equilibrium selection falls short, evolutionary game theory provides a solution. However, it also provides an account of *disequilibrium*; that is, an account of what happens out of equilibrium and how equilibrium might be reached. In an excellent lecture, Arrow (1994) argued that although economists think of their discipline as an exemplar of *methodological*

2 For surveys of early evolutionary game theory see Hofbauer and Sigmund 1988 and Weibull 1995. See also Hammerstein and Leimar 2015.

3 For an overview of evolutionary game theory, see Sandholm 2010.

individualism, this is in fact not the case. Most economic models rely on strong assumptions about beliefs and information, and cannot be reduced to the actions and motivations of individual agents. For instance, in the theory of general equilibrium, agents' individual preferences and production functions are used to determine an aggregate outcome that is stable in some sense, but there is no explanation of how equilibrium is attained. If there are multiple equilibria, the theory does not specify which one we might observe. In fact, it does not even establish that the system will converge to an equilibrium. It is very well to prove that all equilibria are efficient, but in that case one might want to rule out the possibility that agents spend most of their time out of equilibrium. Traditional game theory faces the same problems. In equilibrium, players mutually best-respond and must have correct beliefs about each other's actions. But why should this be the case? Is it reasonable to rule out other outcomes?

In contrast, evolutionary game theory explicitly models what information players have access to, how they form beliefs based on this information, and how they act based on their beliefs. It specifies what happens out of equilibrium, and provides conditions under which equilibrium might be reached. In other words, evolutionary game theory takes methodological individualism seriously. In broad terms, evolutionary models attempt to explain how individual behaviour maps to social outcomes. They attempt to explain how social structures emerge, and what makes some outcomes stable and others unstable. The models in the present thesis fall into this bracket.

1 Duelling

The norm of duelling endured for hundreds of years in Europe. In the United Kingdom it disappeared abruptly in the mid-nineteenth century, whereas in France it declined slowly. I present a simple model of social norms that explains these phenomena. The model predicts that the evolution of norms is characterised by *tipping*, whereby norms can shift suddenly due to shocks, and by a *ratchet effect*, whereby changes in parameters can cause norms to decline gradually. I show that the model can be supported by an equilibrium of a repeated game, with no special assumptions about preferences.

JEL classification: C70, C72, C73, Z13

Keywords: duelling, social norms, subgame-perfect equilibrium, ratchet effect, tipping

1.1 Introduction

Duelling was prevalent in parts of Europe from the Renaissance up to the late nineteenth century. It was generally illegal and decried by institutions such as the Church, yet attempts to extirpate the practice had little impact, or at any rate were unsuccessful, for hundreds of years. It is estimated that more than a third of duels in the early nineteenth century in France ended in the death of one of the duellists,[1] yet people, for the most part, felt compelled to follow the norm.

Contemporary accounts make it clear that social pressure was a significant motivation for duelling: people were concerned about how others would react if they failed to challenge someone to a duel when one was expected, or failed

1 Chesnais 1981, 103.

to accept a duel, or in any way breached the norms surrounding duelling, and this concern was sufficient to drive them to risk their lives. The most striking illustrations of this are the numerous instances of people duelling despite being opposed to the practice. Robert Axelrod, in an influential paper on the evolution of norms, notes how Alexander Hamilton felt that he had no choice but to accept Aaron Burr's challenge.[2] Similarly, Jean Jaurès, a significant socialist figure in French politics and a committed pacifist, once challenged to a duel a political opponent who had insulted him.[3] When a friend asked him not to go through with the duel, he said:

> I've thought about it. I can't let things continue this way. I can feel them all around me, ready to insult me or my wife or my daughter. [...] I feel covered in spit. I want to stop it by an absurd but necessary act.[4]

Until the mid-sixteenth century duellists in France and England predominantly used slashing broadswords and protected themselves with bucklers. Eventually, these were replaced by the use of rapiers, which were far more deadly since they could inflict damage to internal organs whereas broadswords only caused superficial flesh wounds.[5] The fatality rate of duelling spiked and it is thought that as many as ten thousand duellists died in France during the twenty-year reign of Henri IV, between 1589 and 1610.[6] Later, various conventions emerged that limited the risks involved in duelling. Duellists were expected to appoint seconds to organise the duel and conduct negotiations between the two parties; before a duel was formally declared, parties exchanged letters in which they presented their griefs and demands and the seconds attempted to reach a reconciliation.[7] When pistols became popular in the late eighteenth century the distance separating the duellists, which heavily influenced both accuracy and

2 Axelrod 1986, 1095.
3 Nye 1993, 173.
4 Renard 1960, 938.
5 Stone 1965, 242–44.
6 Billacois 1986, 115.
7 Hopton 2007, chap. 3.

lethality, was the subject of strong restrictions. The conventional distance was ten or twelve paces, which might be reduced if the duel was particularly serious, or increased if the converse was true.[8] Duellists who duelled at smaller distances were considered reckless and were often the subjects of social opprobrium. It is interesting to note that although duellists who were considered to have followed conventions were rarely the subject of legal pursuits, this leniency never applied to those who had breached them.[9] The comte de Chatauvillard, the author of a popular duelling manual, said that 'honour might require us to risk our lives, but not to play with them'.[10]

In the United Kingdom, duelling died out suddenly in the mid-nineteenth century – perhaps within a few years in the 1840s and early 1850s.[11] In contrast, in France it declined gradually and endured up to the turn of the century.[12] Swords became popular again in the latter half of the nineteenth century,[13] and, increasingly, the risks involved in duelling were made explicit: duels were announced in advance as being *au premier sang*, to serious wounds, or, rarely, *à la mort*.[14] This had the effect of drastically reducing fatality rates. As mentioned above, it is estimated that more than a third of duels in the early nineteenth century ended in the death of one of the duellists; during the second half of the century the number would be around two percent.[15] Of course these estimates are far from precise, but it is certainly the case that the reduction was so significant that duellists slowly became the subject of ridicule, to the extent that Mark Twain was moved to write in 1880 after visiting France:

> Much as the modern French duel is ridiculed by certain smart people, it is in reality one of the most dangerous institutions of our day.

8 Hopton 2007, 81–83.
9 Hopton 2007, 30–33.
10 Cited in Nye 1993, 142.
11 Hopton 2007, 268.
12 Hopton 2007, 323.
13 Nye 1993, 186.
14 Hopton 2007, 79.
15 Banks 2012, 49–50.

> Since it is always fought in the open air, the combatants are nearly sure to catch cold.[16]

In section 1.2, I present a simple model of social norms that explains why costly norms such as duelling can endure for long periods of time, and why they can sometimes die out rapidly and sometimes decline gradually. Players choose a behaviour on a scale from least costly to most costly, which represents how often they duel or how dangerous the duels they participate in are, and are sanctioned if they deviate from the average behaviour. The further they are from the average behaviour, the higher the sanction. Players therefore best respond by trading off between the cost of the behaviour and the sanction for deviating from the average. Occasionally, players are subject to correlated shocks that affect their choice. The system that results is characterised by *tipping* and by a *ratchet effect*: Most of the time, players choose the average behaviour and the norm is stable. Occasionally, a shock will shift the average and the system will suddenly tip to a new norm. If there is a change in parameters that increases the cost of following the norm relative to the cost of being sanctioned, players may deviate from the average, causing it to decrease. If instead the cost of following the norm decreases relative to the cost of being sanctioned, this will not affect the norm, since players have no incentive to choose a behaviour that is more costly than the average behaviour.

In this framework, the rapid decline of duelling in the United Kingdom can be interpreted as being caused by a shock. In general, a shock could be a particularly unfortunate death that prompts public outcry, an anti-duelling campaign led by the government or by religious authorities, or some other event that might affect people's expectations about the norm. In practice, Hopton (2007, 264–71) associates the decline of duelling in the United Kingdom with three events in particular. Each of these events, or the three of them taken

16 Twain [1880] 1997, 38.

together, can be viewed as a shock that changed people's perceptions of duelling. First, there was significant public outrage over the acquittal of the Earl of Cardigan in 1841. Cardigan was arrested during a duel in which he wounded his opponent. As a peer, he was tried in the House of Lords. This was the first such trial in sixty years, and it aroused keen interest. It was concluded in one day, and Cardigan was acquitted on a technicality. Hopton argues that this and the outrage that followed crystallised public opinion against duelling, causing it to be viewed as an anachronistic and barbaric practice of the privileged. Second, in 1843 a duel between two brothers-in-law, Lieutenant Monroe and Colonel Fawcett, reinforced this sentiment after Fawcett died following the duel. Finally, in 1844 the government introduced legislation aimed at ending duelling in the army. Duelling officers would be cashiered, and widows of duellists would not receive a pension. Hopton argues that these measures not only ended duelling in the army, but also led to the demise of duelling in general. He notes that after they were introduced, 'there were no more than a handful of duels fought in England' (268).

In the case of duelling in France, the gradual decline of the practice can be viewed as being caused by a fall in the impact of social pressure on individuals over the course of the nineteenth century. In the context of the model, this would entail a gradual increase in the cost of the behaviour relative to the cost of being sanctioned and a decline of the norm. A common proposition in sociology is that as societies become more complex and developed, social pressure becomes less effective and informal social control tends to cede place to formal social control, *i.e.* the rule of law.[17] For instance, Merry (1984) provides evidence that gossip can heavily influence individuals' behaviour in small-scale societies but is often ineffective in complex societies. One can make the case that this decline

17 Black 1976, 107–11.

in the impact of social pressure is what happened in France over the course of the nineteenth century, and that this caused the norm of duelling to gradually become unsustainable. It is interesting to note that under this hypothesis, the United Kingdom would have experienced a similar change in the impact of social pressure, which would imply that the decline of duelling in the United Kingdom would have happened in any case and that the shocks that occurred only accelerated the phenomenon.

The key assumption of the model is that players are sanctioned for deviating from the norm. Is this reasonable? In particular, it is usually thought that sanctioning should be costly to the person carrying out the sanction as well as to the person being sanctioned (*e.g.* Coleman 1990, 278–82). But if sanctioning is even slightly costly, why would people do it?

One possible answer is that people who fail to sanction deviators might themselves be sanctioned. This is what Axelrod (1986) termed a *metanorm*. But then why would people sanction those who fail to sanction deviators? This would require a third level of sanctions, which would in turn require a fourth level of sanctions, and so on. We know from the folk theorem for repeated games that such a mechanism is feasible in a game-theoretic sense,[18] however it might seem implausible in practice. Elster (1989) argues that in the context of social norms, people do not keep track of long sequences of deviations. He says that:

> People do not usually frown upon others when they fail to sanction people who fail to sanction people who fail to sanction people who fail to sanction a norm violation. [105]

Elster argues that, as a result, all norms must be supported by preferences of some sort – either preferences that motivate people to follow the norm, or

18 Aumann and Shapley 1976; Rubinstein 1979; Fudenberg and Maskin 1986.

preferences that motivate people to sanction those who deviate from the norm, or both.

I argue that this is not the case and that sanctions can be supported by a simple mechanism that avoids Elster's critique. In section 1.3, I define a repeated game in which players choose whether or not to sanction other players in addition to choosing a behaviour. I propose an equilibrium in which players sanction those who deviated from the average behaviour in the previous period, but also those whose sanctions against other players differed from average sanctions. In such an equilibrium it is in each player's interest to sanction a player who deviates from the norm, because failing to do so would mean failing to sanction someone whom everyone else sanctions, which would trigger sanctions in the following period. And if someone fails to sanction someone who deviates from a norm, it is in everyone's interest to sanction the player who failed to sanction for the same reason. And so on. Yet players do not have to keep track of long sequences of deviations as in Elster's critique; instead, they simply consider whether or not a given player's behaviour and sanctions in the previous period were different from the average. In formal terms, the strategies used by the players are of *memory 1*. It is correct that a norm sustained by social pressure requires agents to sanction not only those who deviate from the norm, but also those who fail to sanction those who deviate from the norm, and those who fail to sanction those who fail to sanction those who deviate from the norm, and so on. However, players do not need to be conscious that that is what they are doing, and the equilibrium I present in section 1.3 demonstrates that these chains of sanctions can exist without players keeping track of who failed to sanction whom.

The idea that behaviours can tip from one norm to another is not new. Indeed, the term 'tipping' goes back to Schelling's 1971 paper on residential

segregation. However, the idea that norms can be subject to a ratchet effect is, to the best of my knowledge, novel. The idea that norms correspond to equilibria has also been explored before. For instance, Young (1998) showed how certain norms can be interpreted as equilibria of coordination games. To take an example, consider the norm according to which a person's shaking her head from left to right signifies 'no', and a person's nodding her head up and down signifies 'yes'. There is no *a priori* reason why these significations should not be reversed – in fact, some cultures have the opposite norm, including in Bulgaria, Greece, and southern Albania. Both norms are instances of coordination equilibria: if everyone else I encounter shakes their head to mean 'no', then it makes sense for me to do the same; conversely, if people use shaking to mean 'yes', then so will I. Viewing norms as coordination equilibria has implications for the way in which they might evolve over time; in particular, Young's theory implies that behaviours should be homogenous within a given culture, but can be wildly diverse from culture to culture – this is referred to as *local conformity / global diversity* – and that norms exhibit tipping. Early versions of these ideas can also be found in Lewis 1969 in the philosophy literature.

However, duelling cannot be viewed as an equilibrium of a coordination game. A person would not feel compelled to duel just because everyone else duelled – instead, it would take either some form of social pressure or some form of internalised preference about duelling. I will refer to duelling as a *costly norm*, and to norms such as the head-shaking norm as *coordination norms*. It is worth noting that, although the present paper focuses on duelling, the model applies to any costly norm. For instance footbinding in China and female genital mutilation are extremely costly practices that are supported by social pressure. Footbinding endured for a millennium in Chinese society, from the eleventh century to the turn of the twentieth century, and, like duelling in the United

Kingdom, it died out abruptly.[19] Note also that costly norms are not limited to socially harmful behaviours: public-good norms are also costly norms, in the sense that they are individually costly even though they are socially beneficial. The model below also applies to these behaviours.

Akerlof (1980) and Axelrod (1986) also look at costly norms and also argue that they can be sustained by sanctions. However both authors assume the existence of sanctions without explaining how they could be sustained. Akerlof posits the existence of an exogenous 'reputation' function. Axelrod assumes that players are 'vengeful' and want to sanction deviators. When discussing metanorms, he assumes that sanctions against people who fail to sanction deviators are also driven by vengefulness. In the sociology literature, Mackie (1996) assumes that players are motivated to imitate one another. It is worth noting that in Akerlof's model players are sanctioned for deviating from the average behaviour as in the model below, but this does not produce a ratchet effect because he looks at a binary set of behaviours, rather than a continuum.

There are also papers that focus on other aspects of norms. Carvalho (2013) shows that veiling in Muslim communities can be understood as a commitment mechanism to limit deviations from religiously prescribed behaviour, and that this can explain the 'new veiling movement' that emerged in the 1970s. Young and Burke (2001) show that a tendency to imitate others' behaviour can explain excess uniformity in cropsharing contracts.

1.2 Model

There is a continuum of players, indexed on $[0, 1]$. Time is discrete, indexed by $t = 1, 2, \ldots$ In period t, each player i chooses a behaviour $x_{it} \in [0, 1]$, where 1

19 See Mackie 1996 for a discussion of the two practices.

is a highly costly behaviour, 0 is uncostly, and 1/2 is intermediate. Specifically, suppose the cost function $c : [0,1] \to \mathbb{R}_+$ is strictly increasing, continuously differentiable and strictly convex. Players whose behaviour is different from the average behaviour are sanctioned, and the sanction a player receives depends on how far her behaviour is from the average behaviour. Specifically, if the average behaviour at t is $\bar{x}_t$, then the sanction incurred by player i is $s(|\bar{x}_t - x_{it}|)$, where $s : [0,1] \to \mathbb{R}_+$ is strictly increasing, continuously differentiable and weakly convex. Player i's total payoff in period t is

$$-\gamma_t c(x_{it}) - s(|\bar{x}_t - x_{it}|),$$

where $\gamma_t \geq 0$ captures the cost of the behaviour relative to the cost of being sanctioned. This *relative cost of the behaviour* follows some known, possibly stochastic process. Most of the results below will not depend on the exact specification of this process. Normalise $c(0) = s(0) = 0$, so that choosing 0 is uncostly and players who pick the average are not sanctioned. Players face a trade-off: on the one hand they want to match what they think the average behaviour will be in order to avoid sanctions; on the other hand they want to choose an uncostly behaviour. If γ_t is large, players will be place relatively more weight on the cost of their behaviour than on potential sanctions. Note that sanctions are exogenous and occur in the same period as the choice of behaviour. In section 1.3, they will be endogenous and will occur in the following period.

I now specify how players act. If in a given period the average is $\bar{x}$ and the relative cost of the behaviour is γ, then player i's best response is

$$b(\bar{x}, \gamma) \coloneqq \arg\max_{x \in [0,1]} -\gamma c(x) - s(|\bar{x} - x|).$$

The fact that c is strictly convex and s is weakly convex ensures that the maximum is unique, so that $b(\bar{x}, \gamma)$ is well defined. In practice, because players act simultaneously, they do not know the average behaviour. I assume players

are *inertial*: in period t, player i best responds to the previous-period average by playing $b(\bar{x}_{t-1}, \gamma_t)$. That is, players act as if the average were unchanged. The initial average $\bar{x}_1$ is arbitrary.

Each period, each player makes a mistake with probability $\lambda \in [0, 1)$. Mistakes represent exogenous shocks to players' expectations about the norm, and are correlated across players according to some known joint distribution. For instance, some event might cause players to believe that the norm will die out or, conversely, become reinforced, and cause them to change their behaviour.

This completes the specification of the model. I now turn to its analysis. For the sake of simplicity, I focus on the case without mistakes. I then discuss qualitatively how the model behaves with mistakes. All proofs are in the appendix.

Note that it is never a best response to choose a behaviour above the average, so $b(\bar{x}, \gamma) \leq \bar{x}$. An immediate consequence is that if there are no mistakes, $\{\bar{x}_t\}$ must be weakly decreasing.

Proposition 1. Suppose there are no mistakes. Then for all t, $\bar{x}_{t+1} \leq \bar{x}_t$.

This implies that the only way $\bar{x}_t$ can increase is through mistakes. If $b(\bar{x}, \gamma) = \bar{x}$, I will say that $\bar{x}$ is *stable* given γ; otherwise, I will say that $\bar{x}$ is *unstable* given γ.

Definition 1. If $b(\bar{x}, \gamma) = \bar{x}$, then $\bar{x}$ is *stable given* γ. If $b(\bar{x}, \gamma) < \bar{x}$, then $\bar{x}$ is *unstable given* γ.

Stability is a key concept. If the average in a given period is stable and there are no mistakes, then all players will play this average as long as it remains stable. Thus stable averages can be viewed as norms. Propositions 2 and 3 below give some characterisation of the set of stable averages.

Proposition 2 states that, given γ, there exists some cutoff average $x^*(\gamma)$ such that any average weakly below $x^*(\gamma)$ is stable and any average above $x^*(\gamma)$ is unstable. In other words, the set of stable averages is $[0, x^*(\gamma)]$. Intuitively, this is because the cost of deviating from the average does not depend on the average, while the benefit from deviating from the average is higher for higher averages (because c is convex). Hence if a given average is stable, then any lower average must also be stable. I will refer to $x^*(\gamma)$ as the *highest stable behaviour given* γ.

Proposition 2. For any relative cost of the behaviour γ, there exists a behaviour $x^*(\gamma)$ such that for all $\bar{x} \leq x^*(\gamma)$, $\bar{x}$ is stable given γ and for all $\bar{x} > x^*(\gamma)$, $\bar{x}$ is unstable given γ.

Proposition 3 establishes that $x^*(\gamma)$ is weakly decreasing in γ. Intuitively, this is because as γ increases, deviating becomes more attractive. Proposition 3 also establishes that $x^*(0) = 1$ and $\lim_{\gamma \to \infty} x^*(\gamma) = 0$, so that any behaviour can be either stable or unstable depending on γ (except 0, which is always stable). Intuitively, when $\gamma = 0$, there is no cost, so all behaviours are stable. When $\gamma \to \infty$, the cost of being sanctioned is irrelevant compared to the cost of the behaviour, so the only stable behaviour is 0.

Proposition 3. The function x^* is weakly decreasing in its argument. Moreover $x^*(0) = 1$ and $\lim_{\gamma \to \infty} x^*(\gamma) = 0$.

Example 1. Suppose $c(x) = x^2/2$ and $s(y) = y$, so that $b(\bar{x}, \gamma)$ solves

$$\max_{x \in [0, \bar{x}]} -\gamma \frac{x^2}{2} - (\bar{x} - x).$$

Suppose $\gamma \neq 0$. Then it is straightforward to show that

$$b(\bar{x}, \gamma) = \begin{cases} 1/\gamma & \text{if } \bar{x} \geq 1/\gamma, \text{ and} \\ \bar{x} & \text{otherwise.} \end{cases}$$

Therefore the highest stable behaviour is $x^*(\gamma) = 1/\gamma$. Note that $x^*(\gamma)$ is decreasing in γ and tends to zero as γ tends to infinity, as stated in proposition 3.

When the current average is stable, players do not deviate from it and the process stays fixed. What happens when the current average is unstable? Propositions 4 and 5 establish that the process will tend to move towards the highest stable behaviour $x^*(\gamma_t)$. Of course, the highest stable behaviour may vary as γ_t varies, so it is not quite as simple as $\bar{x}_t$ converging to $x^*(\gamma_t)$. (In particular, suppose $x^*(\gamma_t)$ were to shift above $\bar{x}_t$. Then $\bar{x}_t$ would be stable and would stay below $x^*(\gamma_t)$ until $x^*(\gamma_t)$ shifts again.) Proposition 4 shows that when the average is unstable, it strictly decreases but does not go below the highest stable average. Proposition 5 shows that when the relative cost is constant, the average converges to the highest stable behaviour.

Proposition 4. Suppose there are no mistakes and that at some period t, $\bar{x}_{t-1} > x^*(\gamma_t)$. Then $\bar{x}_{t-1} > \bar{x}_t \geq x^*(\gamma_t)$.

Proposition 5. Suppose there are no mistakes and the relative cost is constant and equal to γ. Suppose that for some period t', $\bar{x}_{t'} > x^*(\gamma)$. Then $\{\bar{x}_t\} \to x^*(\gamma)$.

Depending on the properties of c and s, convergence to the highest stable behaviour may be fast or slow. The following example illustrates this.

Example 2. In the case where $c(x) = x^2/2$ and $s(y) = y$, recall that when $\bar{x} > 1/\gamma$ we had $b(\bar{x}, \gamma) = 1/\gamma$. Therefore from any unstable average, the process converges instantly to the highest stable average; that is, if $\bar{x}_{t-1} > x^*(\gamma_t)$ then $\bar{x}_t = x^*(\gamma_t)$.

Now suppose $c(x) = x^2/2$ as above, but $s(y) = y + y^2/2$. Sanctions are now strictly convex rather than linear. Then $b(\bar{x}, \gamma)$ is the solution to

$$\max_{x \in [0, \bar{x}]} -\gamma \frac{x^2}{2} - (\bar{x} - x) - \frac{(\bar{x} - x)^2}{2},$$

which implies

$$b(\bar{x}, \gamma) = \begin{cases} \frac{1+\bar{x}}{1+\gamma} & \text{if } \bar{x} \geq 1/\gamma, \text{ and} \\ \bar{x} & \text{otherwise.} \end{cases}$$

The highest stable average is $x^*(\gamma) = 1/\gamma$ as in the previous case. However, the dynamics are different. In the present case the process converges slowly to the highest stable average but never reaches it. To see why, suppose the relative cost is constant and equal to γ. Suppose that the initial average is above $1/\gamma$. Then the process follows a first-order linear difference equation defined by

$$\bar{x}_{t+1} = \frac{\bar{x}_t}{1+\gamma} + \frac{1}{1+\gamma},$$

which slowly converges to $1/\gamma$ but always stays strictly above it.

The difference in the dynamics is to do with the convexity of s. The highest stable average is the average at which the marginal cost of deviating equals the marginal benefit. When s is linear, the cost of an additional marginal deviation is the same as the marginal cost from deviating. But for any behaviour above the highest stable average, the benefit of an additional marginal deviation is strictly greater than this cost, because c is strictly convex. So players will deviate to the highest stable average. In contrast, when s is strictly convex the cost of an additional marginal deviation increases as one moves away from the average.

Finally, consider the case where $c(x) = x^2/2$ and $s(y) = y^2/2$. Then

$$b(\bar{x}, \gamma) = \frac{\bar{x}}{1+\gamma},$$

and therefore the only stable average, for all values of γ, is 0. This is because $s'(0) = 0$. When this is the case and the average is strictly positive, players can always strictly increase their payoff by deviating from the average by some small amount.

When players make mistakes and the relative cost of the behaviour varies over time, the above results imply that the process is characterised by *tipping* and by a *ratchet effect*.

First, suppose mistakes occur as follows: Most of the time, players pick the inertial best response. However, each period a shock occurs with some small probability that shifts players' expectations. This could be any public event that affects how agents think about the norm. The shock causes a significant proportion of players to pick some other behaviour, selected at random. (The new behaviour is the same for all the players who are affected by the shock.) The remaining players pick the inertial best response as normal. Then the process exhibits tipping. In periods when there is no shock, the process behaves as set out in propositions 1 to 5. In particular, if the previous-period average is stable, then the process remains at that average. When there is a shock, however, the average tips suddenly to some new value that depends on the proportion of players who were affected by the shock and the behaviour they shifted to. If the new average is stable, the process will remain there until a new shock occurs or until the new average becomes unstable. In general, depending on the exact specification of mistakes, the process can tip from any stable average to any other stable average.

Second, consider the effect of variation in the relative cost γ_t. For instance, suppose γ_t follows a simple autoregressive process with a small positive trend, so that it may increase or decrease in any given period but will tend to increase overall. Then the process is subject to a ratchet effect. For the sake of simplicity,

suppose there are no mistakes. When γ_t increases, this will tend to reduce the set of stable averages (by proposition 3) and may make the average unstable. If the average becomes unstable, it will tend to converge to the new highest stable average (by propositions 4 and 5). When γ_t increases, this can only expand the set of stable averages. If the current average is stable, it will stay fixed. If the current average stays unstable, it will continue to decrease. This ratchet effect can explain why a norm might decline gradually over time. The system will alternate between phases in which the average decreases and phases in which the average is constant. On average, because γ_t has a positive trend, the average behaviour will tend to decrease. Eventually, the average will converge to zero. If γ_t has a small variance, this will happen slowly.

When players make mistakes and the relative cost varies, then the process is subject to both the above effects. As argued in section 1.1, these two effects can account for the dynamics of duelling in the United Kingdom and in France. The sudden disappearance of the norm in the United Kingdom can be viewed as caused by tipping, while the gradual decline in France can be explained by the ratchet effect.

1.3 Equilibrium

The key assumption in the model of section 1.2 is that players are sanctioned for deviating from the average behaviour. Is this plausible? If it is costly for people to carry out sanctions, why would they do so? As discussed in section 1.1, one possible answer is that people who fail to sanction deviators might also be sanctioned. Elster (1989) argues that this is implausible. Players would not only have to sanction those who deviate from the norm, but also those who fail to sanction those who deviation from the norm, and those who fail to sanction those who fail to sanction those who deviate from the norm, and so on.

I argue that sanctions can be supported by a simple mechanism that avoids this critique. I define a repeated game in which in addition to choosing a behaviour each period, players choose whether or not to sanction other players. I show that for any function s of section 1.2, there is a subgame-perfect equilibrium of the repeated game in which players sanction according to that function. The equilibrium is simple in the sense that it involves memory-1 strategies; that is, strategies that depend only on actions taken in the previous period.

The equilibrium works as follows: instead of sanctioning deviations from equilibrium strategies, players sanction deviations from average actions – both deviations from average behaviour and from average sanctions. Therefore players have an incentive to pick the average behaviour as in section 1.2, but they also have an incentive to sanction players who deviate from the average behaviour, since failing to do so would mean not sanctioning someone whom everyone else sanctions, which would trigger sanctions in the following period. For the same reason, players also have an incentive to sanction players who failed to sanction a deviation, as well as players who failed to sanction players who failed to sanction a deviation, and so on. Yet when picking their action, players just consider what happened in the previous period instead of keeping track of long sequences of deviations.

To avoid complications, suppose that there is a finite population of n players instead of a continuum, where we imagine n to be large. (Formally, we just require $n \geq 4$.) The stage game is as follows: Each player i chooses a behaviour in $[0, 1]$. In addition, each player i can choose to sanction other players. Denote i's behaviour at t by $x_i \in [0, 1]$ as in section 1.2 and i's sanction towards player j by $p_{ij} \in \mathbb{R}_+$, where a high p_{ij} signifies a harsh sanction and 0 signifies no sanctions. (I use the letter 'p' to distinguish the stage-game actions from the sanctions function of section 1.2. One can think of 'p' as standing for

'punishment'.) Denote a typical stage-game action of player i by

$$a_i = (x_i, p_{i1}, p_{i2}, \ldots, p_{i(i-1)}, p_{i(i+1)}, \ldots, p_{in}).$$

Player i's stage-game action set is $A_i \coloneqq [0,1] \times \mathbb{R}_+^{n-1}$. Denote the set of action profiles by $A \coloneqq \times_i A_i$, with typical element a.

Sanctions are costly not only to the player who is sanctioned, but also to the player carrying out the sanction, so that a player's utility depends on her behaviour, on the sanctions she receives, and on the sanctions she carries out. Given an action profile a, denote the average sanction received by player i by $\bar{p}_i$ and the average sanction assigned by i by $\bar{p}_{-i}$; that is,

$$\bar{p}_i \coloneqq \frac{1}{n-1} \sum_{j \neq i} p_{ji}, \text{ and}$$
$$\bar{p}_{-i} \coloneqq \frac{1}{n-1} \sum_{j \neq i} p_{ij}.$$

Then player i's stage-game payoff is

$$u_i(a) \coloneqq -\gamma c(x_i) - \bar{p}_i - \kappa \bar{p}_{-i},$$

where $\kappa \in (0, \frac{n-2}{n-1})$ governs the cost of sanctioning, and c and γ are as in section 1.2. Since $\kappa < 1$, sanctioning is less costly than being sanctioned.

Call the above stage game $\mathscr{G}$. Suppose $\mathscr{G}$ is repeated infinitely many times and suppose players rank payoff sequences according to the discounting criterion, with common discount factor $\delta \in (0,1)$. Call the repeated game $\mathscr{H}$.

Given an action profile a_t in period t, I will denote the average behaviour and average sanctions against player j by $\bar{x}_t$ and $\bar{p}_{jt}$, respectively; that is,

$$\bar{x}_t \coloneqq \sum_i \frac{x_{it}}{n}, \text{ and}$$
$$\bar{p}_{jt} \coloneqq \sum_{i \neq j} \frac{p_{ijt}}{n-1}, \text{ for all } j,$$

where x_{it} is player i's behaviour at t and p_{ijt} is i's sanction against player j at t.

A strategy is of *memory 1* if in every subgame, the action prescribed by the strategy depends only on what happened in the previous period. Proposition 6 states that any sanctions function can be supported by a memory-1 subgame-perfect equilibrium of $\mathscr{H}$.

Proposition 6. For any strictly increasing and weakly convex function $s : [0,1] \to \mathbb{R}_+$, there exists a strict, memory-1 subgame-perfect equilibrium of $\mathscr{H}$ such that for all $t > 1$, i, and $j \neq i$,

$$p_{ijt} = \frac{1}{\delta} s(\tfrac{n}{n-1}|\bar{x}_{t-1} - x_{j(t-1)}|) + \frac{1}{\delta(n-1)} \sum_{k \neq j} |\bar{p}_{k(t-1)} - p_{jk(t-1)}|. \tag{1.1}$$

The sanctions work as follows. Consider the right-hand side of equation 1.1. The first term means that if player j's behaviour at $t-1$ was different from the average behaviour in that period, player i will sanction her at t. Note that the sanction occurs the period following the deviation, whereas in section 1.2, sanctions occurred in the same period. To account for this, the sanction is adjusted by $1/\delta$. Note also that because the population is finite, unlike in section 1.2, individual players can influence averages. Consider a player who deviates in behaviour by Δ. In section 1.2, the difference between that player's behaviour and the average is Δ. Here, however, the deviation causes the average to change by Δ/n, so the difference between the player's behaviour and the average is only $\frac{n-1}{n}\Delta$. To account for this, the sanction in equation 1.1 is adjusted by $\frac{n}{n-1}$. These adjustments imply that a player j who deviates in behaviour at $t-1$ receives a sanction whose present value, viewed from $t-1$, is

$$s(|\textstyle\sum_{i \neq j} \frac{x_{i(t-1)}}{n-1} - x_{j(t-1)}|),$$

which is analogous to the sanctions in section 1.2.

The second term of the right-hand side of equation 1.1 means that a player who fails to sanction someone whom others sanction or who sanctions someone whom others do not sanction will be sanctioned the following period. The quantity

$$\frac{1}{(n-1)}\sum_{k\neq j}|\bar{p}_{k(t-1)} - p_{jk(t-1)}|$$

is player j's average deviation from average sanctions at $t-1$. As with the first term, this average is adjusted by $1/\delta$ to account for the fact that sanctions take place the period following a deviation.

Note that there is no condition on δ. This is possible because sanctions can be arbitrarily large. Alternatively, one could limit sanctions to some interval, which would require placing some restrictions on δ and s. Note also that because individual players can influence averages, any deviation will typically cause everyone to be sanctioned. Suppose a single player deviates in behaviour, thereby reducing the average relative to what it would otherwise have been. Then all the remaining players will have picked an action different to the actual average, and so will be sanctioned the following period. Of course the sanction received by the players who did not deviate will be smaller than the sanction received by the player who deviated, and as n becomes large it tends to zero.

The equilibrium of proposition 6 is memory-1 and therefore avoids Elster's critique. This is because players sanction deviations from average actions rather than deviations from the equilibrium strategy. This means that some deviations may go unsanctioned – for instance if everyone deviates at the same time. However, this does not affect whether or not the strategy profile constitutes an equilibrium, since the concept of equilibrium only considers unilateral deviations by players.

Finally, recall that in section 1.2 it was assumed that players were inertial; that is, they played the best response to the previous-period average $b(\bar{x}_{t-1}, \gamma_t)$.

This is typically not possible in an equilibrium. The reason is that $b(\bar{x}_{t-1}, \gamma_t)$ is a myopic best response, and generally not a mutual best response. If everyone plays $b(\bar{x}_{t-1}, \gamma_t)$ at t, then, depending on the sanctions function, $b(\bar{x}_{t-1}, \gamma_t)$ may not be a best response. In the equilibrium in the proof of proposition 6, players choose the previous-period average $\bar{x}_{t-1}$ if it is stable and the highest stable average $x^*(\gamma_t)$ otherwise. In other words, they choose the closest stable average to the previous-period average. This does not affect the qualitative results about tipping and the ratchet effect.

1.4 Discussion

I have omitted a number of subtler points about the evolution of duelling. For example it is interesting to note that the prevalence of duelling tended to decline during periods of strong central authority, such as in France under Louis XIV.[20] It is also interesting to note that duelling remained prevalent for the longest in France in two professions whose members are particularly susceptible to social pressure: journalism and politics.[21] The pervasiveness of duelling in politics in particular is striking. I have already noted the examples of Hamilton and Jaurès. Abraham Lincoln narrowly avoided a duel when the seconds managed to reconcile him and the other party on the spot of the duel.[22] A total of four Prime Ministers of the United Kingdom participated in duels: William Petty, William Pitt the Younger, George Canning and Arthur Wellesley – of those four Pitt and Wellesley duelled while in office.[23] And finally, one of the last documented duels took place between Gaston Defferre and René Ribière, two French *députés*, in

20 Chesnais 1981, 104–06.
21 Nye 1993, chap. 9.
22 Hopton 2007, 57.
23 Hopton 2007, 233–41.

1967 after the former called the latter an idiot.[24] The interests of politicians are of course directly linked to the public's opinion of them, and therefore it is unsurprising that they should be particularly committed to certain social norms.

I have also ignored the role of other motives beside social pressure in sustaining social norms. Mackie, in his study of footbinding and female genital mutilation (1996), shows that these practices are linked to misinformation about medical facts, and that people often justify them by invoking tradition or religious beliefs.

A richer model might address these and other points. However, the analysis in the present paper should be viewed as exploring one specific aspect of costly social norms, namely how they relate to social pressure and the game-theoretic concept of equilibrium, and what this entails. In particular, I have argued that social pressure can give rise to tipping and to a ratchet effect; that social pressure can be viewed as an equilibrium phenomenon; and that tipping and a ratchet effect together can explain a number of features of the evolution of social norms.

1.A Proofs

Proof of proposition 1. The claim follows immediately from the fact that, for all $\bar{x}$ and γ, $b(\bar{x}, \gamma) \leq \bar{x}$. □

Proof of proposition 2. Fix γ and $\bar{x}$. Recall that the player's problem is

$$\max_{x \in [0,1]} -\gamma c(x) - s(|\bar{x} - x|).$$

24 Hopton 2007, 386.

Since playing above $\bar{x}$ is never optimal, we can reduce the problem to

$$\max_{x \in [0,\bar{x}]} -\gamma c(x) - s(\bar{x} - x).$$

The player's best response is denoted $b(\bar{x}, \gamma)$.

Note that 0 is trivially stable. I claim that if $\bar{x} > 0$, a necessary and sufficient condition for $\bar{x}$ to be stable is

$$-\gamma c'(\bar{x}) + s'(0) \geq 0. \tag{1.2}$$

Intuitively, if the average is $\bar{x}$, then $\gamma c'(\bar{x})$ is the marginal benefit that a player receives from slightly reducing her behaviour relative to $\bar{x}$, while $s'(0)$ is the corresponding marginal increase in sanctions. When inequality 1.2 holds, the sanction is greater than the benefit.

Formally, the argument is as follows: Fix $\bar{x}$. At $x \in [0, \bar{x})$, the derivative of the player's utility with respect to behaviour is $f(x) \coloneqq -\gamma c'(x) + s'(\bar{x} - x)$ Note that f is strictly decreasing because c is strictly convex and s is weakly convex. Intuitively, when x is small relative to $\bar{x}$, the increase in cost from a marginal increase in behaviour is relatively small, while the decrease in sanctions is relatively large. Moreover, $f(\bar{x}) = -\gamma c'(\bar{x}) + s'(0)$.

Suppose inequality 1.2 holds. Then $f(\bar{x}) \geq 0$. Since f is strictly decreasing, $f(x) > 0$ for all $x < \bar{x}$. So any $x < \bar{x}$ is not optimal. Therefore $b(\bar{x}, \gamma) = \bar{x}$. So inequality 1.2 is a sufficient condition for $\bar{x}$ to be stable.

Now suppose inequality 1.2 does not hold. Then $f(\bar{x}) < 0$. Since f is strictly decreasing and continuous, there exists $x < \bar{x}$ that achieves a higher payoff than $\bar{x}$. So $b(\bar{x}, \gamma) < \bar{x}$. So inequality 1.2 is a necessary condition for $\bar{x}$ to be stable.

Using inequality 1.2, we can find a highest stable average. First, I claim that if $\bar{x}$ is stable, then any $\bar{x}' < \bar{x}$ must also be stable. This implies that the set of stable intervals is a subinterval of $[0, 1]$ that includes 0. Intuitively, sanctions

for marginal deviations are the same for high and low averages, but the benefit from deviating is higher for high averages. So if an average is below some stable average, then it must also be stable. Formally, let $\bar{x}$ be a stable average. That is,

$$-\gamma c'(\bar{x}) + s'(0) \geq 0.$$

Let $\bar{x}' < \bar{x}$. Since c is strictly convex, we have

$$-\gamma c'(\bar{x}') + s'(0) \geq 0.$$

So $\bar{x}'$ is stable.

Now consider the supremum of the set of stable averages. (It exists because the set is bounded.) Denote it by $x^*(\gamma)$. I claim that $x^*(\gamma)$ is stable. Because the set of stable averages is a subinterval of $[0, 1]$ that includes 0, any $\bar{x} < x^*(\gamma)$ is stable. That is, for any $\bar{x} < x^*(\gamma)$,

$$-\gamma c'(\bar{x}) + s'(0) \geq 0.$$

Therefore, since c' is continuous, it must be the case that

$$-\gamma c'(x^*(\gamma)) + s'(0) \geq 0.$$

So $x^*(\gamma)$ is stable.

Moreover, since $x^*(\gamma)$ was defined as the supremum of the set of stable averages, any $\bar{x} > x^*(\gamma)$ must be unstable. Therefore there exists an $x^*(\gamma)$ such that any $\bar{x} \leq x^*(\gamma)$ is stable and any $\bar{x} > x^*(\gamma)$ is unstable, as required.

□

Proof of proposition 3. First, I show that $x^*(\gamma)$ is weakly decreasing in γ. Fix γ and γ' such that $\gamma' > \gamma$. Suppose that $\bar{x}$ is stable given γ'; that is,

$$-\gamma' c'(\bar{x}) + s'(0) \geq 0.$$

Then

$$-\gamma c'(\bar{x}) + s'(0) \geq 0,$$

so $\bar{x}$ is also stable given γ. Therefore the set of stable behaviours given γ' is a subset of the set of stable behaviours given γ. But these sets are $[0, x^*(\gamma')]$ and $[0, x^*(\gamma)]$, respectively. Therefore $x^*(\gamma') \leq x^*(\gamma)$. So x^* is weakly decreasing in its argument.

Second, I show that $x^*(0) = 1$. Suppose $\gamma = 0$. Then for any average $\bar{x}$, inequality 1.2 reduces to $s'(0) \geq 0$, which holds since s is strictly increasing. Therefore all averages are stable when $\gamma = 0$. Therefore $x^*(0) = 1$.

Third, I show that $\lim_{\gamma \to \infty} x^*(\gamma) = 0$. Consider any average $\bar{x} > 0$. This average is stable if and only if

$$-\gamma c'(\bar{x}) + s'(0) \geq 0.$$

That is, $\bar{x}$ is stable if and only if

$$\gamma \leq \frac{s'(0)}{c'(\bar{x})}.$$

Therefore for any $\bar{x}$, it is possible to find some γ such that $\bar{x}$ is unstable (since s and c are strictly increasing). Since the set of stable behaviours given γ is $[0, x^*(\gamma)]$, it must be the case that

$$\lim_{\gamma \to \infty} x^*(\gamma) = 0.$$

□

Proof of proposition 4. First, I show that if $\bar{x}_{t-1}$ is unstable, $\bar{x}_t < \bar{x}_{t-1}$. In the model, each player chooses $b(\bar{x}_{t-1}, \gamma_t)$ at t. Therefore $\bar{x}_t = b(\bar{x}_{t-1}, \gamma_t)$. Suppose $\bar{x}_{t-1}$ is unstable. Then $b(\bar{x}_{t-1}, \gamma_t) < \bar{x}_{t-1}$. So $\bar{x}_t < \bar{x}_{t-1}$.

Second, I show that if $\bar{x}_{t-1}$ is unstable, $\bar{x}_t \geq x^*(\gamma_t)$. This amounts to showing that for any $\bar{x}$ and γ such that $\bar{x} > x^*(\gamma)$, $b(\bar{x}, \gamma) \geq x^*(\gamma)$.

Since $x^*(\gamma)$ is stable,

$$-\gamma c'(x^*(\gamma)) + s'(0) \geq 0.$$

Since s is weakly convex, this implies that

$$-\gamma c'(x^*(\gamma)) + s'(\bar{x} - x^*(\gamma)) \geq 0. \tag{1.3}$$

But the left-hand side of inequality 1.3 is the derivative of a player's utility at $x^*(\gamma)$ when the average is $\bar{x} > x^*(\gamma)$ and the relative cost is γ. Since this derivative is decreasing in the player's behaviour, it must be strictly positive at all $x < x^*(\gamma)$. So no $x < x^*(\gamma)$ is optimal. Therefore $b(\bar{x}, \gamma) \geq x^*(\gamma)$.

Intuitively, if a player can't gain from deviating to a lower behaviour when the average is $x^*(\gamma)$, then $x^*(\gamma)$ must still be preferable when the average is higher, because of the convexity of the sanctions function.

□

Proof of proposition 5. Fix γ and $\bar{x}_{t'} > x^*(\gamma)$ for some t'. Proposition 4 implies that $\{\bar{x}_t\}_{t>t'}$ is bounded above and below by $\bar{x}_{t'}$ and $x^*(\gamma)$, respectively. Moreover, proposition 1 implies that it is weakly decreasing. So $\{\bar{x}_t\}$ converges to some $\ell \in [x^*(\gamma), \bar{x}_t]$. Suppose $\ell \neq x^*(\gamma)$. I show that this implies a contradiction.

Since $b(\bar{x}, \gamma)$ is continuous in $\bar{x}$, we have

$$\lim_{t\to\infty} b(\bar{x}_t, \gamma) = b(\ell, \gamma).$$

In addition, since $\ell \in [x^*(\gamma), \bar{x}_t]$ and by assumption $\ell \neq x^*(\gamma)$, it must be the case that $\ell > x^*(\gamma)$. That is, ℓ is unstable. So $b(\ell, \gamma) < \ell$. Therefore

$$\lim_{t\to\infty} b(\bar{x}_t, \gamma) < \ell. \tag{1.4}$$

But $b(\bar{x}_t, \gamma) = \bar{x}_{t+1}$, so

$$\begin{aligned}\lim_{t\to\infty} b(\bar{x}_t, \gamma) &= \lim_{t\to\infty} x_{t+1} \\ &= \ell,\end{aligned}$$

which contradicts inequality 1.4. □

Proof of proposition 6. As discussed in section 1.3, it is not feasible for players to act inertially, since the inertial best response may not be a mutual best response. I consider an equilibrium in which players choose the previous-period average $\bar{x}_{t-1}$ if it is stable and the highest stable average $x^*(\gamma_t)$ otherwise. That is, they choose $\min\{\bar{x}_{t-1}, x^*(\gamma_t)\}$.

Consider the following strategy for player i:

> In period 0, play $x_{i0} = 0$ and, for all $j \neq i$, $p_{ij0} = 0$. In period $t > 1$, play $\min\{\bar{x}_{t-1}, x^*(\gamma_t)\}$; for all $j \neq i$, play
>
> $$p_{ijt} = \frac{1}{\delta} s(\tfrac{n}{n-1}|\bar{x}_{t-1} - x_{j(t-1)}|) + \frac{1}{\delta(n-1)} \sum_{k\neq j} |\bar{p}_{k(t-1)} - p_{jk(t-1)}|.$$

Note that the strategy is memory-1 and that the equation determining the sanctions is the same as equation 1.1. Suppose every player $j \neq i$ uses this strategy. I claim that it is a strict best response for i to use it too.

First, note that i will not deviate in period 0, since this yields no short-term benefits and results in sanctions in the following period.

Second consider a one-shot deviation in behaviour in some period $t \geq 1$. (Suppose i does not deviate in sanctions.) All the other players will be playing some behaviour $\bar{x}$. By construction, $\bar{x}$ is stable. Clearly, i will not deviate to a behaviour above $\bar{x}$, so suppose i deviates to some $x < \bar{x}$. Then at t her cost from the behaviour is $-\gamma c(x)$ instead of $-\gamma c(\bar{x})$. The average behaviour at t is

$$\frac{(n-1)\bar{x} + x}{n},$$

so in period $t+1$, the average sanction received by i is

$$\frac{1}{\delta} s\left(\tfrac{n}{n-1}\left|\tfrac{(n-1)\bar{x}+x}{n} - x\right|\right),$$

which simplifies to

$$\frac{1}{\delta} s(\bar{x} - x).$$

The present value of this sanction viewed at t is $s(\bar{x} - x)$. But since $\bar{x}$ is stable, it solves

$$\max_{x \in [0,\bar{x}]} -\gamma c(x) - s(\bar{x} - x).$$

So deviating is not a best response.

Third, consider a one-shot deviation in sanctions in period $t > 0$. (Suppose i does not deviate in behaviour.) As before, we can assume that i only deviates to less costly sanctions. For each $j \neq i$, let $\bar{p}_j$ be the sanction chosen by each other player against j at t. Define i's average deviation

$$\Delta \coloneqq \frac{1}{(n-1)} \sum_{k \neq i} |\bar{p}_k - p_{ikt}|.$$

At t, i's stage-game payoffs increase by $\kappa\Delta$. The average sanction at t against $k \neq i$ is

$$\frac{(n-2)\bar{p}_k + p_{ikt}}{n-1}.$$

So in period $t+1$, the average sanction received by i is

$$\frac{1}{\delta(n-1)} \sum_{k \neq i} \left| \frac{(n-2)\bar{p}_k + p_{ikt}}{n-1} - p_{ikt} \right|,$$

which simplifies to

$$\frac{n-2}{\delta(n-1)} \Delta.$$

The present value of this sanction viewed at t is $\frac{n-2}{n-1}\Delta$. Since $\kappa < \frac{n-2}{n-1}$, deviating is not a best response.

Finally, since sanctions increase linearly in both deviations in behaviour and deviations in sanctions, the arguments above imply that i has no incentive to deviate in both at the same time.

Therefore it is a strict, memory-1 subgame-perfect equilibrium for all players to use the above strategy. □

2 Community enforcement using modal actions

I prove two folk theorems for repeated games with random matching. A large group of players is rematched at random each period, so that players who deviate must be sanctioned by third parties. Previous analyses have either relied on strong assumptions about information transmission, or have been limited to equilibria that are not robust to noise or in which players are indifferent. I use a simple construction based on modal actions to obtain results for strict and robust equilibria.

JEL classification: C70, C72, C73

Keywords: repeated games, random matching, community enforcement, modal actions

2.1 Introduction

The folk theorem for repeated games can be seen as a formal solution to the problem of cooperation between individuals.[1] For instance, it implies that players in a repeated prisoner's dilemma can cooperate, despite incentives to cheat one another, by threatening credible sanctions if one of them deviates. More generally, it implies that in many repeated games efficient outcomes can be enforced by an equilibrium if players are sufficiently patient. These results hold for standard repeated games, in which players interact with the same partners every period. However in many economic settings, agents interact with different partners at different times and interactions with any given partner may be

1 Aumann and Shapley 1976; Rubinstein 1979; Fudenberg and Maskin 1986.

too infrequent for the standard folk theorem to apply. Can cooperation still be sustained in such settings?

Following Okuno-Fujiwara and Postlewaite (1995) and Kandori (1992), I model this situation as a *repeated game with random matching*. Each period, players are randomly matched to play a stage game. Because any two players meet infrequently, sanctions against players who deviate must be carried out by third parties. This creates a problem of information transmission: If a player deviates, how will her partners in future periods know that she should be sanctioned? And if one of her future partners fails to sanction her, how will the partner's future partners know that she, in turn, should be sanctioned? And so on. This is what Kandori termed the problem of *community enforcement*. In a standard repeated game, the problem does not arise, because players observe every action taken by every player; in a repeated game with random matching, this would be implausible. The question is how much information transmission is necessary to sustain cooperation.

I prove two folk theorems for repeated games with random matching under two mild informational assumptions. Previous analyses have either

1. relied on strong assumptions about information transmission;
2. have only applied to certain types of stage game, such as the prisoner's dilemma;
3. or have been limited to equilibria that are not robust to noise or in which players are indifferent.

Kandori's seminal paper presents two approaches. First, he shows that cooperation can be sustained in the prisoner's dilemma without information transmission, but at the price of instability: in his equilibrium, players cooperate as long everyone they encounter cooperates; but as soon as one of their oppon-

ents defects against them, they defect in all future periods. This is known as a *contagion equilibrium*, because defection spreads from player to player through the population. This type of equilibrium is only feasible in certain stage games such as the prisoner's dilemma, where sanctions correspond to a stage-game Nash equilibrium. Moreover, the equilibrium, as Kandori notes, is not robust to noise, in the sense that if a single player makes a mistake and defects, the whole population starts to defect and never returns to cooperation.

Kandori's second folk theorem applies to to games that satisfy a relatively unrestrictive assumption (essentially, it should be possible to reward those who sanction deviators) and is robust to noise. However it relies on strong assumptions about information transmission. Kandori assumes that there is an exogenous information clearinghouse that recognises deviations and assigns publicly observable 'labels' to players who deviate. Given this assumption, he is able to prove a robust folk theorem. In his equilibrium, the labels denote the number of periods a player should be sanctioned for. If a player deviates, her label is updated to a positive integer. Players who are matched with her in the ensuing periods observe this label and know that they are supposed to sanction her; if they fail to do so their label is updated and they will be sanctioned in turn. In his conclusion, Kandori notes: 'Perhaps the most important question which is unanswered by the present paper concerns the way in which the information transmission postulated in our model is implemented' (77). Okuno-Fujiwara and Postlewaite make a similar assumption in their paper (with 'statuses' instead of 'labels').

Most of the literature on repeated games with random matching has followed the first of these approaches in assuming little or no information transmission between players. These results tend only to apply to games like the prisoner's dilemma or to be limited to non-robust or weak equilibria (see the

literature review in section 2.2). I take the second approach. I show that mild information transmission is sufficient to derive a folk theorem for general games and for strict and robust equilibria.

The results are obtained using a simple construction: In a standard repeated game with random matching, players are paired with a single partner each period. I assume that players are matched with multiple partners in each period, and play the same stage game with each. This drastically changes the nature of the game. In a standard repeated matching game without multiple matching, players may require a large amount of information to determine whether a partner deviated in a previous period. For instance, in the prisoner's dilemma a player who defected in a given period may have deviated, or may have been sanctioning a player who deviated. To find out, one needs to determine whether the player she was matched with had deviated previously, which may require determining whether the players that player was matched with in previous periods deviated, and so on. In general, the amount of information required increases over time and is unbounded. In contrast, when players are matched with multiple partners, one can determine whether or not a player deviated in a given period just by looking at her actions and the actions of other players in that period – specifically, it suffices to check whether a player sanctioned someone when others did not, or failed to sanction someone when others did. That is, one can compare a player's actions with the corresponding *modal actions*. If they differ, then one can infer that the player deviated.[2]

I implement this in two different ways. In section 2.4, I assume that play-

2 It is worth noting that there exist bounded-memory folk theorems for standard repeated games (*i.e.* games without random matching) in which the amount of information required by players does not increase over time. See for instance Barlo, Carmona and Sabourian 2016. However the strategies involved are complex, requiring players to set aside periods to play 'signalling sequences' at regular intervals to identify deviators. Moreover, it is not clear that the results would extend to the random matching case. I discuss this more in section 2.6.

ers can observe partners' actions in recent periods as well as the corresponding modal actions. I show that this is sufficient to obtain a folk theorem for games that satisfy Kandori's assumption about rewarding sanctioners. In the equilibrium, players sanction partners who were non-modal in the recent past. Then it is in each player's interest not to deviate on the equilibrium path, because a player who does so will be non-modal and will be sanctioned. It is also in each player's interest to sanction a player who deviates, because failing to do so would mean failing to sanction someone whom others sanction, which would trigger sanctions in the following periods. The amount of information available to players does not increase over time. In section 2.5, I assume that players can verifiably disclose small amounts of information about their own stage games. I construct an equilibrium in which whenever a player deviates, the partner against whom she deviated discloses the player's action as well as the corresponding actions by the partner's other partners. Then third parties know that the player who deviated was non-modal, and sanction her. To make the equilibrium strict, players who report deviators are rewarded. The result holds for general stage games, not just those that satisfy Kandori's assumption about rewarding sanctioners. Both of the results are for strict and robust sequential equilibria. Depending on the exact setting, the assumptions of one or the other may be more appropriate.

I use a strong notion of robustness. Consider a sequential equilibrium in which each player has the same continuation payoff at every period on the equilibrium path. Kandori says that an equilibrium is *globally stable* if, from any history, players' expected continuation payoffs eventually converge to their equilibrium-path payoffs. This means that the impact of any deviation on play and on payoffs should disappear over time. Kandori's labelling equilibrium satisfies this property but his contagion equilibrium does not. There are two

problems with global stability as a notion of robustness. First, it does not rule out the possibility that payoffs converge very slowly. Suppose players are subject to trembles, in the sense that they make mistakes each period with some probability. Then even if the probability of making a mistake is small, it is possible in a globally stable equilibrium for payoffs to be far from equilibrium payoffs in a large proportion of periods. Second, global stability does not say anything about whose payoffs are affected by a deviation. To see why this is a problem, consider the following contagion equilibrium for the prisoner's dilemma, due to Ellison (1994). Suppose players have access to public randomisation devices. As in Kandori's contagion equilibrium, players cooperate on the equilibrium path and start defecting as soon as they are matched with a partner who defects. However, players return to cooperating as soon as the public randomisation device takes a certain value. The value is chosen to set the expected duration of the defection phase, which ensures that the strategies constitute an equilibrium. This equilibrium satisfies global stability, since play eventually returns to cooperation with probability one, however a deviation by a single player causes everyone to defect for a stretch of time. In particular, as Ellison notes, if there is a single 'crazy' player who always defects, this will cause defection throughout the population in every period.

To account for these problems, I define *robustness* as follows: In a robust equilibrium, a player's continuation payoff in any given period can only differ from her equilibrium-path payoffs if either

1. the player deviated in the recent past;
2. one of her partners deviated in the recent past; or
3. multiple players deviated in some period in the recent past.

From any history, players' payoffs converge to their equilibrium-path values after some fixed number of periods, which rules out excessively slow conver-

gence. Only deviators and those matched with deviators can have non-equilibrium payoffs, unless multiple players deviated. The caveat about multiple deviations is to address the fact that in the equilibria of sections 2.4 and 2.5, modal actions can be affected by multiple deviations, which can result in players who did not deviate being sanctioned. The equilibria I construct in sections 2.4 and 2.5 satisfy this definition of robustness. In fact, so does Kandori's labelling equilibrium. Both Kandori's and Ellison's contagion equilibria do not.

2.2 Related literature

As discussed in the preceding section, the seminal papers in the literature are Okuno-Fujiwara and Postlewaite 1995 and Kandori 1992. Another early contribution is Harrington 1995, which considers a model in which time is continuous and players meet at random intervals.

Ellison (1994) extends Kandori's result for the prisoner's dilemma, showing that contagion equilibria can be made robust to noise in a certain sense. As mentioned above, he shows that when players have access to a public randomisation device, contagion equilibria can be made globally stable. He also proves a result for the case in which there is no public randomisation, using a different notion of robustness. He considers the game with trembles and says that an equilibrium is robust if, in the limit as discounting and the probability of making mistakes vanish, players' payoffs converge to the payoffs from cooperation. He constructs an equilibrium that satisfies this notion of robustness, in which defections are spread out over time, between periods of cooperation. (He says that we might see 'players cheating on every third Friday but cooperating on all other days' [581].) This equilibrium does not satisfy global stability, since sanctions never stop. Moreover, if many deviations occur there may be sanctions in a large proportion of periods. As with Kandori's global stability,

Ellison's definition of robustness does not say anything about whose payoffs are affected by deviations. As discussed above, his equilibrium with public randomisation, because it is based on a contagion equilibrium, involves all players being affected by a deviation. The same is true of the equilibrium without public randomisation.

Takahashi (2010) considers the prisoner's dilemma and assumes that players can observe their own past actions and their partners' past actions. The equilibria he constructs are related to concepts in the imperfect private monitoring literature, and in particular to belief-free equilibria. As in that literature, the strategies involved are complex and players are always indifferent between sanctioning and not sanctioning following a deviation. His equilibria are not guaranteed to be globally stable (and *a fortiori* are not guaranteed to be robust in the sense discussed above).

Deb (2012), also drawing on the imperfect private monitoring literature, obtains a folk theorem for general games. She assumes matched players can engage in cheap talk before each stage game, which allows them to identify one another. Since information is cheap talk, players prove their identity by playing 'signatures' that depend on their and their partner's last interaction. As with Takahashi's equilibria, the strategies are complex and involve players being indifferent.

Finally, Deb and González-Díaz (2014) have strict equilibria for games that satisfy the one-sided incentive problem (a class that includes the prisoner's dilemma but is otherwise restrictive). They assume no information transmission. The equilibria are similar to Kandori's and Ellison's contagion equilibria, but with an added 'trust-building' phase at the beginning of the game. The equilibria are not robust to noise (although the authors conjecture that they could be made robust in the sense considered by Ellison).

2.3 Repeated matching games with multiple partners

The setup of the model broadly follows Kandori 1992, except that players are matched with multiple partners each period.

There is an even number of players, represented by the set $N \coloneqq \{1, 2, \ldots, 2n\}$, where $n \geq 3$. The population is divided into two subpopulations $N_1 \coloneqq \{1, 2, \ldots, n\}$ and $N_2 \coloneqq \{n+1, n+2, \ldots, 2n\}$. I will refer to players in N_1 as type-1 players and to players in N_2 as type-2 players.[3] In each period each type-1 player is matched with s type-2 players, where $s \geq 3$ is a factor of n, to play a stage game $\mathscr{G}$.[4]

Let A_1 and A_2 be the action sets of type-1 and type-2 players in $\mathscr{G}$, respectively, and let g_1 and g_2 be their payoff functions. Suppose A_1 and A_2 are finite. Let $A \coloneqq A_1 \times A_2$, with typical element a. Let V denote the payoff space of $\mathscr{G}$; that is, V is the convex hull of

$$\{v \in \mathbb{R}^2 : v = (g_1(a), g_2(a)) \text{ for some } a \in A\}.$$

As in Kandori's model, I rule out mixed actions. The only way in which this restricts the results is that, in general, the minmax payoffs without mixed actions may be higher than the minmax payoffs when mixed actions are allowed. It is straightforward to show that the results still hold when one allows mixed actions, provided one also allows either observability of mixed actions or public randomisation.

Each period, matches are realised according to some stochastic rule. Let $\mu(i, t)$ be the set of player i's realised matches in period t. The results in section 2.4 hold for any matching rule. For instance, each player might be more

3 We could also have a model in which there were no subpopulations and players were randomly assigned a type each period in addition to being assigned partners.

4 The results also hold for $s = 2$, but the analysis is slightly more complicated. The requirement that s be a factor of n guarantees that matchings are possible.

likely to be matched with some players than others, or matchings in period t might depend on matchings in period $t-1$. In fact, matches could be deterministic. (Although matches cannot depend on players' actions.) In section 2.5, I will assume *uniform random matching* (the probability of any two matches in a given period is equal, and matches are independent across periods).

Players evaluate infinite sequences of payoffs using the discounting criterion with shared discount factor $\delta \in (0,1)$. Because players receive s stage-game payoffs per period, I define repeated-game payoffs to be discounted sums of *average* stage-game payoffs. Specifically, suppose player i's average stage-game payoff in each period t is v_i^t. Let $\mathbf{v}_i = (v_i^t)_{t=1}^{\infty}$ be the resulting infinite sequence of average payoffs. Then define i's average discounted payoff to be

$$u_i(\mathbf{v}_i) \coloneqq (1-\delta)\sum_{t=1}^{\infty} \delta^{t-1} v_i^t.$$

Since each v_i^t is an average, we have $u_i(\mathbf{v}_i) \in V$. Let $\underline{v}_1$ be the minmax payoff for type-1 players and let m_2 be one of the corresponding minmaxing actions of type-2 players. Formally,

$$\underline{v}_1 \coloneqq \min_{a_2} \max_{a_1} g_1(a_1, a_2), \text{ and}$$

$$m_2 \in \arg\min_{a_2} \max_{a_1} g_1(a_1, a_2).$$

Define $\underline{v}_2$ and m_1 analogously. The set of feasible and strictly individually rational payoff pairs is

$$V^* \coloneqq \{v \in V : v \gg (\underline{v}_1, \underline{v}_2)\}.$$

Assume V^* is non-empty. Finally, it will be convenient to define the mutual minmaxing action profile $m \coloneqq (m_1, m_2)$, the mutual minmaxing payoffs $w_1 \coloneqq g_1(m)$ and $w_2 \coloneqq g_2(m)$, and the maximum payoffs $\bar{v}_1 \coloneqq \max_a g_1(a)$ and $\bar{v}_2 \coloneqq \max_a g_2(a)$.

This repeated matching game represents situations in which agents interact with different partners at different times. By way of illustration, consider the following example. Suppose the players are traders. Each period, each trader meets other traders at random and they can barter. There is always some mutually beneficial exchange – say one trader comes from a town that produces apples and the other comes from a town that produces bananas. However, each trader can cheat the other (perhaps by sending a crate of rotten apples or overripe bananas). The question is whether or not trade can develop despite the fact that traders can cheat, even if each pair of traders does not meet regularly enough for the standard folk theorem to apply. These interactions can be modelled by the prisoner's dilemma in figure 2.1. Each player has a good that is worth 1 to them, but that the other player can sell for 2. Playing C means giving the good to the other player and playing D means cheating. If players exchange goods they both get 2; if they both cheat they both get 1; and if one cheats but the other does not, the player who cheated gets 3 while the other player gets 0. The prisoner's dilemma, of course, is a particularly important application for the standard folk theorem. In addition to trade, it can be used to model the tragedy of the commons, collusion between firms, and any number of situations in which agents must trust each other to cooperate. One could also imagine that interactions between traders could be modelled by other stage games. The results below will hold for a range of stage games, including the prisoner's dilemma.

If the repeated matching game described above were a game with perfect information, it would be straightforward to adapt Fudenberg and Maskin's (1986) folk theorem to apply. However, since players are unlikely to be able to observe all actions in a large population, it would seem reasonable to restrict players' information. The question is how much information is necessary to sus-

	C	D
C	$2, 2$	$0, 3$
D	$3, 0$	$1, 1$

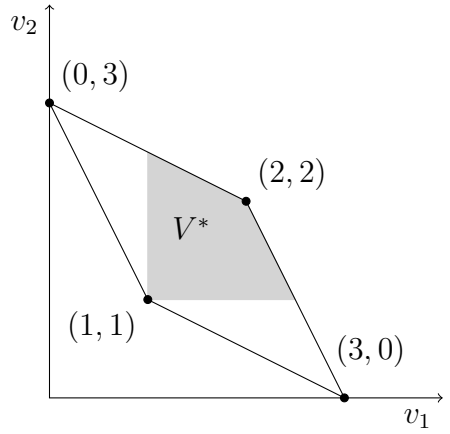


Figure 2.1: Game matrix and payoff space for the prisoner's dilemma.

tain a folk theorem. Kandori assumes that deviators can be labelled as such by an exogenous information clearinghouse and that players observe these labels. Given this assumption, he is able to prove a relatively general and robust folk theorem. In what follows, I show that Kandori's result can be obtained without assuming the existence of a public labelling mechanism, using the fact the players are matched with multiple opponents each period. I prove two results under different assumptions about information transmission. In section 2.4, I assume players can observe the modal actions corresponding to their partners' actions; in section 2.5, I assume they can verifiably disclose information about their own stage games.

I will focus on equilibria that are robust to noise. In a robust equilibrium:

1. following a deviation, payoffs converge to their equilibrium values after some fixed number of periods, and
2. unless multiple deviations occur, only the payoffs of deviators and those matched with deviators are affected by a deviation.

Formally, I define robustness as follows:

Definition 1. A sequential equilibrium with equilibrium-path payoffs $v \in V^*$ is *robust* if there is some T such that for any period t and any player i of type

ℓ, i's expected continuation payoff at t can only differ from v_ℓ if either

1. player i deviated in the past T periods;
2. some player $j \in \mu(i,t)$ deviated in the past T periods; or
3. there were multiple deviations in at least one of the last T periods.

Kandori uses a weaker concept, which he calls 'global stability'. A globally stable equilibrium is one in which payoffs eventually converge to their equilibrium values following a deviation. However global stability does not place any restrictions on the speed of convergence following a deviation. In contrast, under the notion of robustness defined above, convergence must occur in finite, bounded time. Moreover, global stability does not restrict whose payoffs can be affected by a deviation.

In addition to robustness, I will focus on strict equilibria.

2.4 Observable modal actions

Before turning to the result, I introduce some notation.

If i was matched with j in period t, let a_{ij}^t be the action chosen by i in the stage game with j at t. Let $\boldsymbol{\alpha}_i^t$ be the s-tuple of actions taken by i at t; that is,

$$\boldsymbol{\alpha}_i^t \coloneqq (a_{ij}^t)_{j \in \mu(i,t)}.$$

Let θ_i^t be the set of the most frequent actions played against i at t. For example, suppose the stage game is the prisoner's dilemma of figure 2.1 and suppose $s = 4$. If three of the players i was partnered with at t played C against i and one played D, then $\theta_i^t = \{C\}$. If two of the players picked C against i and two picked D, then $\theta_i^t = \{C, D\}$. I will refer to θ_i^t as the set of *modal actions* against i at t.

Denote the s-tuple of modal sets corresponding to $\boldsymbol{\alpha}_i^t$ by $\boldsymbol{\vartheta}_i^t$; that is

$$\boldsymbol{\vartheta}_i^t \coloneqq (\theta_j^t)_{j \in \mu(i,t)}.$$

In the equilibrium, players will compare $\boldsymbol{\alpha}_i^t$ to $\boldsymbol{\vartheta}_i^t$ to determine whether i deviated at t. Specifically:

Definition 2. Player i was *non-modal* at t if there exists some j such that $j \in \mu(i,t)$ and $a_{ij}^t \notin \theta_j^t$. Otherwise, i was *modal* at t.

Then, in particular, a player who failed to sanction someone whom others sanctioned will be non-modal; so will a player who sanctioned someone whom others did not sanction. Note that determining whether or not some player i was modal at t requires knowing exactly $\boldsymbol{\alpha}_i^t$ and $\boldsymbol{\vartheta}_i^t$. One can think of $\boldsymbol{\vartheta}_i^t$ as specifying the actions i should have played in order to be modal.

In the present section, I assume that players have information about their partners' actions in the recent past as well as the corresponding modal actions. This ensures that players can determine whether or not their partners were modal. To be precise, in period t player i knows who she is matched with and the actions she and each current partner took in the previous p periods; that is, she knows $\mu(i,t)$ and $(\boldsymbol{\alpha}_j^\tau)_{\tau=t-1}^{t-p}$ for each $j \in \mu(i,t)$ and for $j = i$. Player i also knows, for each current partner and herself, the modal actions from the previous p periods corresponding to her or the partner's actions; that is, $(\boldsymbol{\vartheta}_j^\tau)_{\tau=t-1}^{t-p}$ for each $j \in \mu(i,t)$ and for $j = i$. She has no other information. I will refer to this as a *repeated matching game with multiple matching and observable modal actions*. The integer p can be thought of as the players' memory.

Suppose we want players to receive some payoff pair $v \in V^*$. I construct an equilibrium in which players sanction those who were non-modal in previous periods. This mechanism not only disincentivises players from deviating on the equilibrium path, but also incentivises them to sanction those who deviate. A

player who deviates on the equilibrium path will be playing a non-modal action, which means that she will be sanctioned in ensuing periods. A player who fails to sanction someone who played a non-modal action will herself be playing a non-modal action, so she will be sanctioned too.

A complication is that under this strategy, there are subgames in which players would expect to sanction their partners for a number of periods in a row (for instance a subgame following a period in which a large number of players were non-modal). In the standard two-player folk theorem, following a deviation there is a phase of sanctions in which players minmax each other. The potential gains from deviating in the sanctions phase are limited by the fact that each player is being minmaxed. In the random matching setting, however, this does not work. The problem is as follows: Suppose sanctions are carried out by mutual minmaxing. Consider a player who believes that a large fraction of other players deviated in the previous period, but who herself did not deviate. If the player is matched with a partner who deviated, her gain from deviating is limited, as in the standard folk theorem. If, however, the player is matched with a partner who did not deviate, then her gain from deviating could potentially be large, which causes the argument to break down.

To address this problem, Kandori assumes that sanctioners can be rewarded for carrying out sanctions. More precisely, he assumes that there exist actions r_1 and r_2 that allow deviators to 'repent' and that reward sanctioners. In his equilibrium, when a deviator is matched with a non-deviator the deviator plays the 'repenting' action and the non-deviator plays the minmaxing action. I make the same assumption:

Assumption 1. There exists $r \in A$ such that

$$g_1(m_1, r_2) > g_1(m) \geq g_1(r_1, m_2), \text{ and}$$
$$g_2(r_1, m_2) > g_2(m) \geq g_2(m_1, r_2).$$

Assumption 1 is exactly A1 in Kandori 1992. Note that it is satisfied by the prisoner's dilemma. In a given period where the type-1 player must be sanctioned, players will pick (r_1, m_2); they will pick (m_1, r_2) when a type-2 player must be sanctioned. This ensures that players prefer to sanction than to be sanctioned.

Theorem 1. Under assumption 1, every point $v \in V^*$ in the repeated matching game with multiple matching and observable modal actions is sustained by a strict, robust sequential equilibrium if δ and p are large enough.

Proof. Assume there exists some $a \in A$ that achieves the designated payoff pair v. If there is no such action profile one can assume public randomisations are available or appeal to the techniques in Fudenberg and Maskin 1991.[5]

Suppose players i and j are matched in period t. Let ℓ be i's type. Fix an integer $q \leq p$. The strategy is as follows:

> If $t = 1$, player i sets $a_{ij}^t = a_\ell$. If $t > 1$ and j was non-modal at some t' in the last q periods but i was not, or if both were non-modal at some t' and t'' in the last q periods, i sets $a_{ij}^t = m_\ell$; if i was non-modal at some t' in the past q periods but j was not, i sets $a_{ij}^t = r_\ell$; otherwise, i sets $a_{ij}^t = a_\ell$.

The idea is that a player matched with someone who was non-modal should sanction her by playing m_ℓ. A player who was non-modal should 'repent' by

5 This would require relaxing robustness slightly, since one could no longer guarantee that a particular payoff pair would be received exactly in every period in the no-sanctions phase. Instead, one could require that payoffs always be arbitrarily close to the equilibrium payoffs when players aren't being sanctioned.

playing r_ℓ, unless she is matched with another player who was non-modal. Otherwise players pick equilibrium-path actions. The integer q is the length of sanctions.

I now show that it is a strict sequential equilibrium for all players to follow the above strategy for δ large enough and for appropriate q. The proof is similar to the proof of theorem 2 in Kandori 1992. There are two types of subgames to check.

First, consider a player i of type 1 and a subgame beginning in period t such that i was non-modal in at least one of the q periods preceding t. Player i's incentive to deviate is greatest when she was non-modal in period $t-1$, so suppose that is the case. If player i does not deviate at t, she gets either w_1 or $g_1(r_1, m_2)$ in each of her stage games in periods t to $t+q-1$, depending on who she is matched with and her beliefs about who is non-modal. Denote the expected average of her stage-game payoffs in period τ by $x(\tau)$. From $t+q$ on, she gets v_1 for certain in each stage game in each period. In total, her (average) expected continuation payoff from not deviating is

$$(1-\delta)\left(x(t)+\delta x(t+1)+\cdots+\delta^{q-1}x(t+q-1)\right)+\delta^q v_1.$$

If i deviates at t, she gets at most $\underline{v}_1$ at t in each stage game. (She can potentially deviate in all of her stage games.) From $t+1$ to $t+q-1$, her expected payoffs are the same as her expected payoffs in the previous case (since, by assumption, her actions do not affect who she is matched with). In period $t+q$, she gets $g_1(r_1, m_2)$ in every stage game since no other players will have deviated from t on under the assumption of equilibrium. Thereafter she gets v_1. So her expected continuation payoff from deviating is at most

$$(1-\delta)\left(\underline{v}_1+\delta x(t+1)+\cdots+\delta^{q-1}x(t+q-1)+\delta^q g_1(r_1, m_2)\right)+\delta^{q+1}v_1.$$

Therefore the payoff from not deviating minus the payoff from deviating is at least

$$(1-\delta)\left(x(t)-\underline{v}_1-\delta^q g_1(r_1,m_2)\right)+(\delta^q-\delta^{q+1})v_1.$$

Simplifying, we get that

$$(1-\delta^q)g_1(r_1,m_2)+\delta^q v_1-\underline{v}_1>0 \tag{2.1}$$

implies that i has no incentive to deviate at t. (Recall that $x(t)$ is an average of w_1's and $g_1(r_1,m_2)$'s, so that $x(t)\geq g_1(r_1,m_2)$.)

Second, consider a player i of type 1 and a subgame beginning in period t such that i was modal in all of the q periods preceding t. If i does not deviate at t, her payoffs between t and $t+q-1$ are, for any belief she might hold, at least $\min\{g_1(m_1,r_2),v_1\}$ in each stage game. From $t+q$ on, she gets v_1. So her expected continuation payoff if she does not deviate is at least

$$(1-\delta^q)\min\{g_1(m_1,r_2),v_1\}+\delta^q v_1.$$

If she deviates, she gets at most $\bar{v}_1$ at t in each stage game and at most w_1 between $t+1$ and $t+q$. From $t+q+1$ on she gets v_1. So her continuation payoff from deviating is at most

$$(1-\delta)\bar{v}_1+\delta(1-\delta^q)w_1+\delta^{q+1}v_1.$$

Therefore if

$$(1-\delta^q)\min\{g_1(m_1,r_2),v_1\}+\delta^q v_1>(1-\delta)\bar{v}_1+\delta(1-\delta^q)w_1+\delta^{q+1}v_1 \tag{2.2}$$

then it is not a best response for i to deviate at t.

Inequalities 2.1 and 2.2 are, *mutatis mutandis*, the same as inequalities 3 and 4 in Kandori 1992. Analogous conditions ensure that type-2 players have no incentive to deviate either. As Kandori argues, there exists $\underline{\delta}$ such that for

all $\delta \geq \underline{\delta}$ there exists q such that inequalities 2.1 and 2.2 and the corresponding inequalities for type-2 players hold. This requires p to be large enough. Thus if δ is large enough and p is large enough relative to δ there exists a strict sequential equilibrium that supports the payoffs v.

The equilibrium is robust: If some player i of type ℓ has a payoff in some stage game in some period that differs from v_ℓ, then it must be the case that either i was non-modal in the past q periods or that at least one of i's partners was non-modal in the past q periods. This can only be the case if either i or one of her partners deviated in the past q periods, or if enough deviations occurred in a single period to change modal actions relative to what they otherwise would have been. □

How reasonable are the informational assumptions? Note that players consider at most $2ps(s+1)$ separate pieces of information in each period.[6] In the equilibrium above, p just needs to be large enough to allow deviators to be sanctioned for sufficiently many periods. If both payoffs in v are significantly larger than the minmax payoffs, as would be the case for the efficient outcome of the prisoner's dilemma, then v could be sustained with a relatively low p. Moreover, there is no restriction on s, other than it be weakly greater than 3 and a factor of n. That said, it may still seem as if the informational requirements are quite strong. A player has to know not only her actions and the actions of her partners, but also the modal actions corresponding to her and her partners' actions. In particular, this implies that she must have some information about

6 There are s actions in each $\boldsymbol{\alpha}_i^t$, and each period players consider p such s-tuples for $s+1$ agents, namely themselves and the players they are currently matched with. In addition, for each action considered, the players also consider the corresponding modal action. So the total number of actions considered, including redundancies, is $2ps(s+1)$. If one considers that players need to know the individual actions that determine the modal actions, then the total is $ps^2(s+1)$. (For each of their and their partners' actions in a given period, players need to know s actions – namely the action and the $s-1$ corresponding actions. So the total is $s(s+1)$ times s times p.)

her partners' partners' partners' actions. There are two points worth noting.

First, as discussed in section 2.1, the amount of information required is significantly less than would be necessary in a game without multiple matching. For example, suppose that the stage game is the prisoner's dilemma of figure 2.1 and that players are matched with a single opponent in each period. Consider a strict equilibrium in which players sanction deviators by defecting. Consider some player i trying to determine whether her partner in period t deviated at $t-1$. Let $j = \mu(i,t)$ be i's partner. If j played D at $t-1$, this could be because j deviated, but it could also be because the player she was matched with at $t-1$, namely $\mu(j,t-1)$, deviated at $t-2$. Say $\mu(j,t-1)$ played D at $t-2$. Then in order to check whether $\mu(j,t-1)$ deviated at $t-2$, player i needs to know about the past actions of the player $\mu(j,t-1)$ was matched with at $t-2$. And so on. The amount of information required increases as time passes. In fact, in some subgames players would have to know actions that occurred in the first period of the game in order to detect a deviation. (Of course, there may be equilibria in which players do not care whether their opponents deviated or were sanctioning a deviator. Contagion equilibria are an example. In games without random matching, there are also equilibria in which deviators can be identified with a bounded memory, as in Barlo, Carmona and Sabourian 2016. Both have problems, as discussed in section 2.1.)

In contrast, in the model with multiple matching, a player trying to determine whether her partner deviated in the previous period just needs to look at a subset of the actions of that period. In particular, the amount of information needed in each period does not increase over time. An interesting conceptual point is that this is possible because players don't actually check whether their partners deviated from their equilibrium strategies. Rather they check whether they deviated relative to modal behaviour. This means that deviations can go

unsanctioned if enough players deviate to change the modal action. Similarly, a player may be sanctioned despite not having deviated if enough other players deviate. This, of course, is not a problem because equilibria only require considering unilateral deviations.

The second point is that all a player really needs to know is whether she or her partners were non-modal in the past p periods. In particular, one could imagine that there is some mechanism whereby players become informed of non-modal players, rather than having to know all of their partners' past actions and checking whether they were modal. I consider one such mechanism in the following section.

2.5 Verifiable disclosure

I now assume that players can publicly communicate small amounts of verifiable information. I construct an equilibrium in which players disclose all their partners' actions against them whenever one of their partners deviates. This informs all third parties that the partner was non-modal and allows them to sanction her. For the equilibrium to be strict, I need to ensure that players who correctly report a deviation are rewarded.

This construction is only feasible because of multiple matching. If players were just matched with one partner each period, the amount of information required to prove that a player deviated would be unbounded and would not be available to a single player. So allowing verifiable disclosure would not help us, unless players systematically disclosed every action from every one of their stage games.

The informational assumptions are as follows. Players now only have direct information about the stage games they were involved in. Specifically, for every stage game player i has played, she knows her opponent's identity, her own

action, her opponent's action, and the period in which the stage game took place. However, players can verifiably disclose information at no cost. At the end of each period, after the stage games have taken place, players can disclose actions from their stage games. Once an action is disclosed, every other player knows the action, the period in which it was played, the identity of the player who played it, and the identity of the player against whom it was played. Players remember actions for p periods after they are disclosed. As in section 2.4, p can be thought of as the players' memory. I will refer to this as a *repeated matching game with multiple matching and verifiable disclosure*. In the proposed equilibrium, actions will only be disclosed when a deviation occurs.

For the sake of convenience, I assume that if players disclose an action from a particular period, they must disclose all other actions by their partners from that period. Let $\boldsymbol{\beta}_i^t$ denote the s-tuple of actions by player i's partners against i at t; that is,

$$\boldsymbol{\beta}_i^t \coloneqq (a_{ji}^t)_{j \in \mu(i,t)}.$$

Recall that θ_i^t is the set of modal actions against i at t.

Definition 3. Player i was *non-modal* in $\boldsymbol{\beta}_j^t$, for some t and $j \neq i$, if $i \in \mu(j,t)$ and $a_{ij}^t \notin \theta_j^t$. Player i was *revealed non-modal* at t against j if i was non-modal in $\boldsymbol{\beta}_j^t$ and j disclosed $\boldsymbol{\beta}_j^t$.

As before, we are interested in strict and robust sequential equilibria. The proposed equilibrium is straightforward: on the equilibrium path players choose actions that yield the desired payoffs; if some player i deviates against j in period t, j discloses $\boldsymbol{\beta}_j^t$; this will reveal that i's action was different from the other actions played against j and third parties will sanction her in the following periods. In addition, players who correctly report a deviation are rewarded.

There are two caveats. The first is that the proposed equilibrium is not quite strict. Players are not disincentivised from disclosing information when no deviation occurred (or from disclosing their own deviations). However such revelations are irrelevant to the game, in the sense that they do not affect anyone's actions or payoffs. I say that the equilibrium is *quasi-strict.*[7]

Definition 4. A sequential equilibrium is *quasi-strict* if, for any subgame, any alternative best response by a player in the subgame does not affect any player's subsequent actions or payoffs.

The second caveat is that not all payoffs in the payoff space V^* can be sustained by the equilibrium. This is because players who report a deviation need to be rewarded with a slightly increased payoff. Therefore a player's equilibrium payoff cannot be the highest possible payoff for that player. Recall that $\bar{v}_1$ and $\bar{v}_2$ are the maximum payoffs for type-1 and type-2 players, respectively. Define

$$\hat{V} \coloneqq \{v \in V : (\underline{v}_1, \underline{v}_2) \ll v \ll (\bar{v}_2, \bar{v}_2)\}.$$

Note that $\hat{V} \subseteq V^*$. In fact, $\hat{V}$ is either equal to V^* or is V^* with part of the efficient frontier removed. In the prisoner's dilemma, $\hat{V} = V^*$. Our folk theorem will be for every point in $\hat{V}$ rather than V^*.[8]

I also assume that matching is uniform random – that is, the probability of any two matches in a given period is equal, and matches are independent across periods – and that n is large. This is to rule out situations in which a player might not want to report a partner who deviated because she is likely to

7 It turns out that it is more difficult than one might expect to disincentivise people from revealing irrelevant information. It is possible to do so at the price of additional assumptions (for instance if the minmax payoffs are obtained by a strict stage-game Nash equilibrium, as in the prisoner's dilemma, or if revealing information is slightly costly).

8 This reward is analogous to the rewards in Fudenberg and Maskin's (1986) folk theorem. Except that they do not require the equilibrium to be robust, so the rewards do not affect the set of payoffs that can be sustained.

be rematched with her in the future and would then have to sanction her. With uniform random matching, the probability of any two players being matched in a given period tends to zero as n tends to infinity. Although the result holds for any matching rule with this property, uniform random matching simplifies notation in the proof.

Assumption 1 about rewarding sanctioners is no longer needed. This is because deviations are revealed publicly, which allows players to coordinate and means that there is no need to sanction multiple players at once. In the equilibrium below, when a player deviates all previous deviations are forgiven and if multiple players deviate in a given period, none of them are sanctioned. (This is similar to the standard folk theorem for games without random matching.) So theorem 2 applies to general games.

Theorem 2. Any point $v \in \hat{V}$ in the repeated matching game with multiple matching and verifiable disclosure is sustained by a quasi-strict, robust sequential equilibrium if δ, p, and n are large enough and if matching is uniform random.

Proof. Let $a \in A$ be an action profile that achieves the designated payoff pair v. As in the proof of theorem 1, if no such action profile exists one can appeal to the techniques in Fudenberg and Maskin 1991 or allow public randomisation. Using the same rationale, assume that there exists $a^* \in A$ such that a^* yields payoffs $(v_1+\varepsilon, v_2+\varepsilon)$, for some $\varepsilon > 0$. Note that this is possible because $v \in \hat{V}$. This action profile will be used to reward players who report deviators.

Fix an integer $q \leq p$. The strategy is similar to the one in the proof of theorem 1, except that players disclose deviations, that players who disclose deviations are rewarded, and that there is no 'repent' action.

There are two types of phase. Phase A is the cooperation phase and phase B_{ij} is for when player i should be sanctioned and player j rewarded. Start in

phase A. From any phase, whenever some player i is revealed non-modal against some player j, start phase B_{ij}. If multiple players are revealed non-modal in the same period, go to A. If in A and no player is revealed non-modal, stay in A. If in B_{ij} and no players are revealed non-modal after q periods spent in B_{ij}, go to A. Actions in each phase are as follows:

(A) Play a.

(B_{ij}) In stage games that involve i, play m. In stage games that involve j but not i, play a^*. In all other stage games, play a.

In both types of phase, if at least one of player i's partners is non-modal against her, she discloses $\boldsymbol{\beta}_i^t$; otherwise, she does not disclose anything.

I argue that it is a quasi-strict sequential equilibrium for all players to follow this strategy, under uniform random matching, for δ and n large enough, and for appropriate q.

First, consider a player i of type 1 in phase A. If i does not deviate, her continuation payoff is v_1. (As before, recall that repeated-game payoffs are normalised by s. The player gets v_1 in each of her stage games.) If she deviates she gets at most

$$(1-\delta)\bar{v}_1 + \delta(1-\delta^q)w_1 + \delta^{q+1}v_1.$$

So

$$v_1 > (1-\delta)\bar{v}_1 + \delta(1-\delta^q)w_1 + \delta^{q+1}v_1 \tag{2.3}$$

implies that i has no incentive to deviate.

Second, consider a player i of type 1 in phase B_{ij}, for some $j \neq i$. She has the greatest incentive to deviate when there are q periods remaining in the sanctions phase, so suppose that is the case. If i does not deviate, she gets

$$(1-\delta^q)w_1 + \delta^q v_1.$$

If she deviates, she gets

$$(1-\delta)\underline{v}_1 + \delta(1-\delta^q)w_1 + \delta^{q+1}v_1.$$

So i has no incentive to deviate provided

$$(1-\delta^q)w_1 + \delta^q v_1 > (1-\delta)\underline{v}_1 + \delta(1-\delta^q)w_1 + \delta^{q+1}v_1,$$

which simplifies to

$$(1-\delta^q)w_1 + \delta^q v_1 > \underline{v}_1. \tag{2.4}$$

Third, consider a player i of type 1 in phase B_{jk}, for some $j \neq i$ and $k \neq i, j$. Suppose i is not matched with j. (Suppose she is not matched with k either, since this would just increase her payoff from not deviating.) Then for n large enough, given uniform random matching, i's problem is close to the problem of a player in phase A. Formally, given q and δ, we can choose n large enough that inequality 2.3 implies that i has no incentive to deviate.

Now suppose i is matched with j in period t; that is, $j \in \mu(i,t)$. If i does not deviate, she gets w_1 at t in her stage game with j and v_1 from each of her other stage games. (Again, we can suppose she is not matched with k.) Thereafter, if n is large enough she will expect to get a payoff arbitrarily close to v_1 in every stage game. So for n large enough her expected continuation payoff is arbitrarily close to

$$(1-\delta)\left(\frac{w_1}{s} + \frac{s-1}{s}v_1\right) + \delta v_1.$$

If she deviates she gets $\underline{v}_1$ at t in her stage game with j at at most $\bar{v}_1$ in her other stage games, after which she is sanctioned. So her continuation payoff is at most

$$(1-\delta)\left(\frac{\underline{v}_1}{s} + \frac{s-1}{s}\bar{v}_1\right) + \delta(1-\delta^q)w_1 + \delta^{q+1}v_1.$$

So for n large enough,

$$(1-\delta)\left(\frac{w_1}{s}+\frac{s-1}{s}v_1\right)+\delta v_1 > (1-\delta)\left(\frac{\underline{v}_1}{s}+\frac{s-1}{s}\bar{v}_1\right) + \delta(1-\delta^q)w_1+\delta^{q+1}v_1 \quad (2.5)$$

implies that she has no incentive to deviate. It is straightforward to show that inequalities 2.3 and 2.4 together imply inequality 2.5.

Finally the problem of player i in phase B_{ji} for $j \neq i$ is analogous to the previous case, except that i has even less incentive to deviate, because she is being rewarded.

Therefore players have no incentive to deviate in the stage games provided inequalities 2.3 and 2.4 hold. But these are equivalent to inequalities 5 and 7, respectively, in Fudenberg and Maskin 1986. Therefore it is possible to choose δ large enough and q given δ such that inequalities 2.3 and 2.4 hold.

We also need to show that players have a strict incentive to disclose information when someone deviates. The argument is as follows. Suppose i is matched with j in period t and j is non-modal in $\boldsymbol{\beta}_i^t$. (To be precise, let us also suppose that j is the only non-modal player in $\boldsymbol{\beta}_i^t$ and that i did not deviate.) Let ℓ be i's type. If i is never rematched with j, disclosing $\boldsymbol{\beta}_i^t$ strictly increases i's payoff in the following p periods, since she receives $v_1+\epsilon$ in each of her stage games. In periods in which i is rematched with j, disclosing $\boldsymbol{\beta}_i^t$ may decrease her payoffs, since she will have to sanction j; however, given uniform random matching we can find n large enough that disclosing $\boldsymbol{\beta}_i^t$ still increases i's continuation payoff in expectation.

Therefore it is possible to choose δ large enough and p and n large enough relative to δ such that there exists a sequential equilibrium that supports the payoffs v.

The equilibrium is quasi-strict rather than strict because there is no disin-

centive against a player disclosing information when no one deviated, when the player herself deviated, or when multiple partners deviated. As in section 2.4, it is straightforward to check that the equilibrium is robust. □

2.6 Conclusion

The analysis above demonstrates how cooperation can be sustained in random matching games under mild assumptions about information transmission. The result relies on the use of *modal actions*. When players are matched with multiple partners each period, deviations in a given period can be identified with information from that period alone. This significantly reduces the informational burden on players. In contrast to many existing results about repeated games with random matching, the folk theorems in the present paper are robust to noise and involve strict equilibria.

It is worth noting that the idea of using modal actions to detect deviations can be applied to settings other than random matching games. Consider the following model of intergenerational transfers due to Bhaskar (1998). One player is born each period, and each player lives for two periods. In the first period of a player's life, when she is 'young', she receives an endowment. The 'old' player in a given period does not have an endowment, but she can receive transfers from the young player of the following generation. (As in an unfunded pension system.) The socially optimal outcome is for young players to transfer half of their endowment, so that consumption is smoothed over time. Samuelson (1958) noted that this outcome could be supported by a Nash equilibrium. Bhaskar shows that it cannot be supported by a subgame-perfect equilibrium if players have limited memory: players who deviate must be sanctioned by the following generation but cannot be distinguished from players who sanction a deviator; as a result, any candidate equilibrium unravels by backward induction.

However, if multiple people are born each period one can determine whether a player deviated by comparing her actions to those of other players born in the same period, so limited memory is not a problem. Conceptually, random matching games and models of overlapping generations are similar. In both cases, players who observe a deviation directly are unable to carry out sanctions, so the problem is to have sanctions carried out by third parties. Modal actions provide a solution to this problem by allowing effective information transmission about which players to sanction.

Modal actions can also be applied to standard repeated games with bounded memory. As mentioned in section 2.1, existing bounded-memory folk theorems tend to involve complex strategies. The key problem is to construct strategies that are such that deviations can be detected and the identity of the deviator revealed. Barlo, Carmona and Sabourian 2009 call such strategies 'confusion-proof'. In Barlo, Carmona and Sabourian 2016, players play specific sequences of actions, which the authors call 'signalling sequences', at regular intervals. These serve as reminders of the play path and tell players who should be sanctioned. Players need to be incentivised to play the signalling sequences, which renders the strategies complex. Moreover, the sequences need to be spaced far apart to ensure that they are not too costly, so that the required memory is potentially large. In contrast, modal actions can be used to derive a simple folk theorem in pure strategies, provided the stage game is symmetric. If the game is symmetric and players strategies are symmetric, then any deviating player will be non-modal, so deviations can be easily detected even with a bounded memory. In this example, as in the overlapping generations model and in the case of repeated games with random matching, the key conceptual point is that modal actions allow deviations to be succinctly encoded using information about actions from the period in which they occurred.

3 Learning repeated-game strategies

The literature on boundedly rational learning has tended to focus on stage-game actions. I present a stochastic learning rule for repeated-game strategies. Players form beliefs about their opponent's strategy based on past actions and best-respond. Occasionally, they make mistakes and experiment, and I show that the equilibrium selected depends on exactly how players make mistakes. Simple specifications of the learning rule yield intuitive selection results: the maxmin, or Rawlsian, outcome; the Nash bargaining solution; the maximum of the sum of payoffs; and a generalisation of risk dominance.

JEL classification: C70, C72, C73

Keywords: repeated games, learning, bounded rationality, stochastic stability

3.1 Introduction

Much of the boundedly rational learning literature, for instance fictitious play and reinforcement learning, focuses on stage-game actions rather than repeated-game strategies, and therefore must be thought of as being concerned with situations in which players are randomly rematched in each period, or other settings in which repeated-game effects are neutralised. I present a model of boundedly rational learning in two-player, infinitely repeated games in which players learn repeated-game strategies.[1]

The model provides sharp selection results. For any stage game, the learning rule selects a subgame-perfect equilibrium of the corresponding repeated

1 I consider two-player games for clarity of exposition. The extension to n-player games is straightforward.

game, in the sense that all stochastically stable states involve players choosing strategies that constitute a subgame-perfect equilibrium.[2] The payoffs received by each player in the equilibrium depend on exactly how players make mistakes. Simple specifications of the learning rule yield intuitive selection results: the maxmin, or Rawlsian, outcome; the Nash bargaining solution; the maximum of the sum of payoffs, sometimes known as the utilitarian outcome; and a generalisation of risk dominance. For instance, in the prisoner's dilemma and the battle of the sexes games of figure 3.1, the version of the rule that yields the maxmin outcome selects the payoff profiles $(3, 3)$ and $(1.5, 1.5)$, respectively. (In fact, so does the version that yields the Nash bargaining solution.) This means that the rule leads to (C, C) being played in all periods on the equilibrium path in the prisoner's dilemma, and to (B, B) and (F, F) being played approximately alternately in the battle of the sexes (depending on the discount factor). These results are achieved under plausible assumptions about the players: they form beliefs based on evidence; they are rational with high probability, in the sense of optimising given beliefs; and the learning rules are uncoupled.[3]

The model works as follows: Players form beliefs about their opponent's strategy based on recent histories of play. They reject beliefs when their opponent plays something that conflicts with their belief for too many periods. When they hold a belief they usually best respond to it, but sometimes make mistakes. After rejecting a belief, they become uncertain and spend a few periods forming a new belief based on their opponent's recent actions. During this time they play more randomly. Under this learning rule, states that do not

2 A *stochastically stable* state of a stochastic process is, roughly speaking, one that recurs with non-negligible frequency in the long run as noise vanishes. I give a precise definition of the concept below.

3 A learning rule is *uncoupled* if it does not require the player to have knowledge of her opponent's payoff function to implement it; it is *completely uncoupled* if it is uncoupled and does not require knowledge of the opponent's past actions (Young 2009). The rule in the present paper is uncoupled but not completely uncoupled.

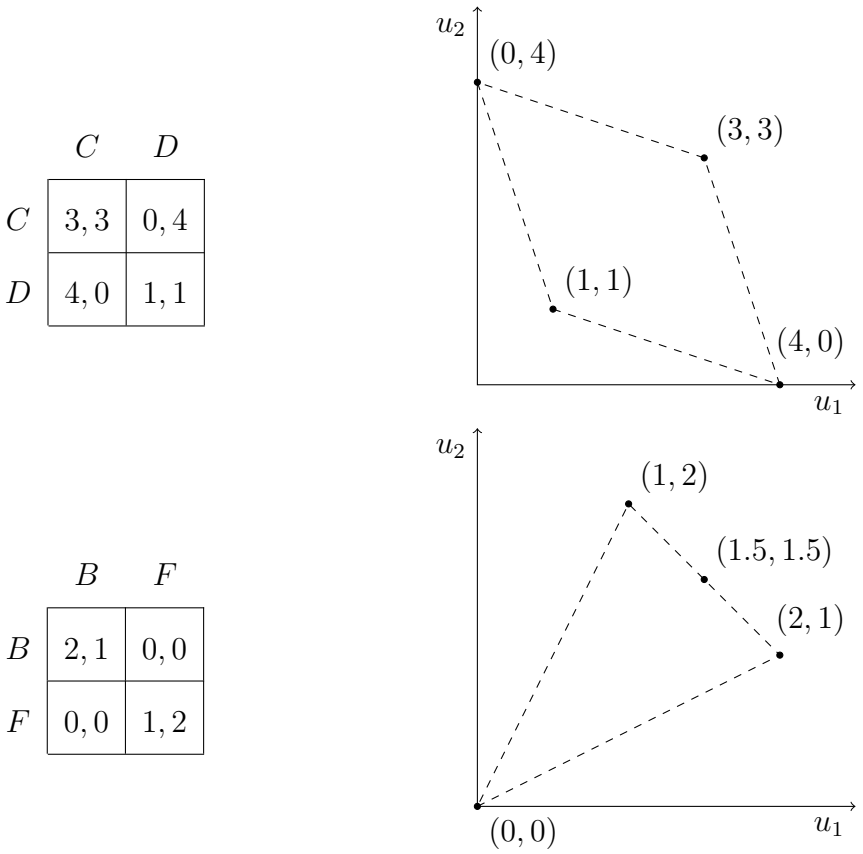


Figure 3.1: Game matrices and payoff spaces for the prisoner's dilemma (top) and the battle of the sexes.

correspond to a subgame-perfect equilibrium will tend to be unstable because they involve one or both players having a wrong belief. The assumption that players tend to play more randomly when they are uncertain is key in making the model tractable, since it ensures that any equilibrium can be easily reached from a state in which either player is uncertain.

The payoffs in the stable states are determined by the way in which players make mistakes. In sections 3.2 and 3.3, I consider a baseline specification of the model in which the probability of an agent's making a mistake in a given state is decreasing in the expected discounted utility received in that state. That is, I assume that a more dissatisfied player is more likely to make mistakes. I show that under this assumption, any stochastically stable state achieves maxmin

payoffs. This is because in a state in which one of the players expects a low utility, that player will make mistakes relatively frequently, rendering the state unstable. The use of stochastic stability, in addition to making the model tractable, is key for this result. Stochastic stability involves taking the limit as noise vanishes, which implies that mistakes by the player with highest utility are negligible compared to mistakes by the worse-off player. Therefore the utility of the worse-off player is what determines the stability of a state.

In section 3.4, I consider variants to this baseline specification. If the probability of making a mistake is decreasing in the number of mistakes the player has made in the past – that is, if players who recently made mistakes are less likely to make make further mistakes – then both players' mistakes matter for determining the stochastically stable states. I show that this can be used to achieve the maximum of the sum of payoffs or the Nash bargaining solution. If the probability of a player's making a mistake depends on her current stage-game payoff rather than her expected discounted utility, then the results change drastically. I present a specification that selects a generalisation of risk dominance.

The learning rule is related to procedures by Young (2009) and by Pradelski and Young (2012). These rules share the feature that players can be in two different modes: one mode in which they play according to a fixed strategy, and one mode of random search, in which they experiment with different strategies. In Young 2009 and Pradelski and Young 2012, the random search mode is triggered by low payoffs, whereas in the present model it is brought about by the opponent's play not conforming to the player's beliefs. Pradelski and Young summarise this as *fast and slow search*. They note that this type of rule has parallels in other fields, such as computer science, and in nature. For instance bees move from flower to flower in a small neighbourhood as long

as they find sufficient nectar; but if the nectar yield is low, they search for another neighbourhood further away. This is known as a 'near-far strategy'. (See Pradelski and Young 2012, 885.) In both papers, the authors consider convergence to equilibria of the stage game and completely uncoupled learning rules, whereas I consider convergence to repeated-game equilibria and rules that are simply uncoupled. Thus the rules have different interpretations – the stage-game rules can be viewed as models of large-population settings in which players have little information about each other, while the present rule is more suited to small populations.

Although the focus of much of the learning literature has been on stage-game actions, there are exceptions. First, models of Bayesian learning, following Kalai and Lehrer (1993), feature players that have priors over their opponent's strategy space, which they update as the game is played. The approach in the present paper differs by relaxing perfect rationality for near rationality, which allows me to obtain convergence to subgame-perfect equilibrium rather than Nash equilibrium and to arrive at selection results. Second, there are evolutionary models in which repeated games are themselves played repeatedly and the repeated-game payoff of a player using a given strategy determines that strategy's prevalence in the next iteration of the repeated game. An early example is Axelrod and Hamilton 1981. Finally, the approach that is closest to the present one is that of Foster and Young (2003), which also uses stochastic stability, and in which players are permanently matched to play a unique repeated game, form beliefs about their opponent's strategy given past play, and are near rational given those beliefs. Foster and Young show that in their model any stochastically stable state must involve players' strategies constituting a subgame-perfect equilibrium, but do not obtain equilibrium selection results. I rule out mixed strategies, which simplifies the analysis considerably, since it

allows players to count mistakes rather than infer them probabilistically. (And it does not affect the payoff space, since we are dealing with repeated games.)

The rest of this paper is organised as follows: In section 3.2, I present the baseline specification of learning rule. In section 3.3, I state and prove the selection result for that specification. Finally, in section 3.4, I consider alternative specifications.

3.2 The learning rule

Let $\mathscr{G}$ be a two-player game with finite action sets A_1 and A_2 and payoff functions u_1 and u_2. Let a_1 and a_2 denote typical elements of A_1 and A_2, respectively. Define $A \coloneqq A_1 \times A_2$, with typical element a. Suppose players do not use mixed actions. This assumption buys us considerable tractability, as will be made clear below. The game $\mathscr{G}$ is repeated infinitely many times and players rank infinite payoff sequences according to the discounting criterion; that is, player i assigns utility $U_i(\mathbf{v}_i)$ to any sequence of payoffs $\mathbf{v}_i = (v_i^t)_{t=1}^{\infty}$, where

$$U_i(\mathbf{v}_i) \coloneqq (1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}v_i^t,$$

where $\delta \in (0,1)$ is a constant shared by both players. Call the repeated game $\mathscr{H}$.

Players follow a particular learning rule, which I will call R_M^{ε}, where ε is a small positive number that captures players' propensity for mistakes and 'M' stands for 'maxmin'. The rule works as follows: In a given period, player i is in one of two *moods*. In the *certain* mood, she has a belief about her opponent's strategy and plays according to a strategy that is a best response, but sometimes makes mistakes; in the *uncertain* mood, she is unsure of her opponent's strategy and attempts to guess it, and, because she is unsure, she is

more prone to playing randomly.[4]

Players' beliefs and strategies are restricted to being of memory k; one can think of k as the *complexity* of players' strategies. Fix k to be some arbitrary strictly positive integer for the rest of the paper.

Definition 1. For any subgame of $\mathscr{H}$, the *ℓ-history* at that subgame is the ℓ-tuple of the action profiles played in the ℓ most recent periods.

Definition 2. A strategy is of *memory k* if for any two subgames with the same k-history, the strategy prescribes the same action.

Restricting strategies to being of memory k is necessary in order to define the process generated by the learning rule as a finite Markov chain: transition probabilities must depend on a finite history, otherwise the number of states would be infinite. Let S_i be the set of memory-k strategies for player i, with typical element s_i.

I first describe how players act in each of the moods, before explaining how transitions between them occur. The two moods are:

> *Certain* $c(s_i, s_j)$: Player i has a belief about j's behaviour, which is a strategy $s_j \in S_j$. She also has a strategy $s_i \in S_i$ that is a best response to s_j in every subgame of $\mathscr{H}$.[5] In each period, i makes a mistake with probability $\varepsilon^{f(U_i)}$, where $\varepsilon \in (0,1)$, f is a strictly positive and strictly increasing function, and U_i is i's predicted continuation payoff i given s_i, s_j, and the current k-history. If a mistake is made, i plays an action from

4 Moods are simply states of the automaton that determines a player's behaviour. I use the term to distinguish from the states of the Markov chain that will be defined below, following Young 2009.

5 Note that there exists a best response to s_j that belongs to S_i.

A_i at random with uniform probability.[6] If no mistake is made, i plays according to s_i.

Uncertain $u(s_i)$: Each period, i plays an action from A_i according to a fully mixed distribution $p(s_i, h)$, where h is the current k-history.

In the certain mood, the higher the player's continuation payoff, the less likely she is to make mistakes. In other words, the more dissatisfied a player is, the more likely she is to make a mistake. This will imply that equilibria in which at least one player receives low utility will tend to be unstable. The distribution p in the uncertain mood is unimportant – for instance we could imagine that $p(s_i, \cdot)$ is close to s_i. What matters is that it is fully mixed. Then for ε small enough, mistakes in the uncertain state are arbitrarily more likely than mistakes in the certain state.

I now describe how transitions between moods occur. Fix strictly positive integers m and τ for the rest of the paper. In broad terms, a player in the certain mood looks back m periods to check her beliefs. If her opponent deviated τ or fewer times relative to her belief in those m periods, she stays certain; otherwise, she becomes uncertain with positive probability. A player in the uncertain mood looks back m periods and with positive probability forms a new belief based on her opponent's actions in those periods. Thus m can be thought of as the players' *memory*, while τ is their *tolerance* for discrepancies between actions and their beliefs.

In order to determine whether an action in a given period was a deviation, players need to know what happened in the k periods prior to that. Therefore we need m to be larger than k. The larger m, the easier it is for players to identify wrong beliefs. If m is larger than k but not by much, then players may

6 I use the uniform distribution for ease of exposition. It suffices that the distribution is fully mixed.

forget deviations too soon to notice that their belief is wrong. It turns out that

$$m > 2k + \tau$$

is sufficient for the results below. We also need to assume

$$\tau > 2k.$$

This means that players should not be too sensitive to mistakes.

I assume that players include their own deviations in the count for transitioning to the uncertain state. This is not critical to the model but simplifies analysis. It is reasonable if one considers that players realise that their mistakes will tend to cause their opponent to change strategy whenever the opponent has a belief that is roughly correct.

I will say that an m-history is *inconsistent* with a strategy profile if it contains more than τ deviations by either player:

Definition 3. Let h be an m-history. Let h' be the $(m-k)$-history consisting of the $m-k$ most recent action profiles in h. Let $s_1 \in S_1$ and $s_2 \in S_2$. Given h, count the deviations from s_1 and s_2 in h'. If the number of deviations is strictly greater than τ, h is *inconsistent* with the strategies s_1 and s_2; otherwise, h is *consistent* with s_1 and s_2. If the number is zero, h is *fully consistent* with s_1 and s_2.

Players will start to reject beliefs when the history is inconsistent with their beliefs, in the sense defined above. In particular, if a player's belief is correct and her opponent makes $\tau + 1$ mistakes in a row, she will reject her belief with positive probability. Note that if mixed actions were allowed, players could not count mistakes. Instead, they would have to infer them probabilistically, which would significantly complicate analysis (as in Foster and Young 2003).

Transitions occur as follows:

Certain $c(s_i, s_j)$: As long as the most recent m-history is consistent with s_i and s_j, player i stays certain. Otherwise i goes to $u(s_i)$ with probability $b \in (0, 1)$ each period.

Uncertain $u(s_i)$: Each period, a transition occurs with probability $b' \in (0, 1)$. If a transition occurs, pick a belief $s_j \in S_j$ according to a distribution $q(s_i, h)$, where h is the most recent m-history. Pick one of the best responses to s_j in S_i at random with uniform probability. Call the chosen strategy s_i. Player i goes to $c(s_i, s_j)$. Otherwise, i stays uncertain.

The probabilities b and b' are arbitrary. The distribution $q(s_i, h)$ can be any distribution provided that the probability of choosing the correct belief when one exists is strictly positive. For example, $q(\cdot, h)$ could be such that the player is more likely to choose strategies that could have generated the opponent's actions in h, or could have generated actions similar to those ones.

Looking ahead, note that there are two ways in which a given pair of certain moods might be unstable: First, if the players' strategies do not constitute a subgame-perfect equilibrium – that is, if at least one of the players has an incorrect belief – then at some point one of the players is likely to become uncertain as her opponent plays actions that conflict with her belief. Second, if at least one player has a low expected utility then she is likely to make mistakes, which will eventually make her opponent become uncertain. Therefore a stable state will correspond to a subgame-perfect equilibrium in which payoffs are maxmin.

In the learning rule, the probability of player i's making a mistake depends on her predicted continuation payoff U_i. It might seem more natural for it to depend on the utility loss from deviating, *i.e.* the level U_i minus the utility from deviating. However, this would render the model intractable. Moreover, it is not clear that it would change the result: the off-equilibrium paths of any stable

equilibrium would involve many periods of sanctions, but, since the probability of making a mistake would still depend on the utility from not deviating, it seems plausible that the equilibrium payoffs would still be maxmin.

Finally, suppose the function f is $\mathbb{R} \to \mathbb{R}^{++}$ and is continuous. Define the maximum and minimum payoffs in $\mathscr{G}$:

$$\bar{v} \coloneqq \max_{i \in \{1,2\},\, a \in A} u_i(a), \text{ and}$$

$$\underline{v} \coloneqq \min_{i \in \{1,2\},\, a \in A} u_i(a).$$

Assume the range of f is small given τ, in the sense that

$$(\tau + 1) f(\underline{v}) > \tau f(\bar{v}). \tag{3.1}$$

(Recall that f is strictly increasing.) Thus a high utility reduces the probability of making a mistake, but not too much. To see what this implies, consider a given sequence of mistakes. The total probability of this sequence depends on both the number of mistakes in the sequence and the continuation payoffs of the players at the moment of making each mistake, and this probability is decreasing in both the number and the payoffs. Inequality 3.1 implies that the former effect dominates the latter. This will ensure that states that don't correspond to subgame-perfect equilibria are unstable relative to ones that do, even if they are states in which players expect high utility.

This completes the definition of the learning rule R_M^ε.

Suppose both players use R_M^ε. This determines the transition matrix of a Markov chain for which a state is a pair of moods, one for each player, and an m-history. Let M_i be the set of moods i can be in. Let $Z \coloneqq M_1 \times M_2 \times A^m$ be the set of states, with typical element z. Denote the transition matrix of the Markov chain by P_M^ε.

3.3 The result

I now state the main result formally. In broad terms, I show that provided k and δ are large enough, the stochastically stable states of the process P_M^ε must involve (i) players' beliefs being correct, (ii) players' strategies constituting a subgame-perfect equilibrium, and (iii) players receiving discounted payoffs arbitrarily close to a maxmin payoff profile in every period.

At this point, it will be convenient to introduce some notation. Let V be the convex hull of the set of feasible payoff pairs in $\mathscr{G}$. Define the minmax payoffs for each player as usual, except that mixed actions are disallowed. That is, player i's minmax payoff is

$$w_i \coloneqq \min_{a_j \in A_j} \max_{a_i \in A_i} u_i(a_i, a_j).$$

Let V^* be the set of feasible and strictly individually rational payoff pairs; namely

$$V^* \coloneqq \{v \in V : v \gg (w_1, w_2)\}.$$

Assume that $\mathscr{G}$ is such that V^* is non-empty. Let x be the *maxmin payoff*; that is

$$x \coloneqq \max_{v \in V^*} \min_{i=1,2} v_i.$$

Note that x is a payoff, rather than a payoff pair.

When ε is strictly positive, the process P_M^ε is irreducible and aperiodic and therefore has a unique stationary distribution, μ^ε. The approach used in the stochastic stability literature is to consider the limit of this distribution as ε becomes small, namely

$$\mu^* \coloneqq \lim_{\varepsilon \to 0} \mu^\varepsilon,$$

as μ^* is generally more tractable than μ^ε.[7] The *stochastically stable* states of

7 The seminal paper in game theory is Foster and Young 1990. Ellison 2000 provides a good overview of the type of analysis used here.

P_M^ε are defined to be the states that have positive probability in μ^*. Since μ^ε corresponds to the long-run frequency distribution of P_M^ε, the stochastically stable states are the states that are observed in the long-run with non-negligible frequency when ε is small.

Let Ω be the set of recurrent classes of P_M^0, with typical element ω. Note that Ω is a set of subsets of Z. I will say that $\omega \in \Omega$ is a *limit set* of P_M^ε. Note that if at least one state of a limit set is stochastically stable, so are the others; I will say that a limit set is stochastically stable if it contains at least one stochastically stable state. Note, moreover, that within a given limit set, players must always have the same strategies and beliefs, since any transition between moods happens via the uncertain mood, and, as we will see below, any state in which at least one player is uncertain is not in any limit set. I will say that players use *constant strategies* within a limit set.

Proposition 1. In any stochastically stable limit set of P_M^ε, players' beliefs are correct and players use constant strategies that constitute a subgame-perfect equilibrium of $\mathscr{H}$. Moreover, for any $\lambda > 0$, the continuation payoff of each player in every state of any stochastically stable limit set is at least $x - \lambda$, provided δ is large enough and k is large enough given δ.

Thus the learning rule selects a subgame-perfect equilibrium that yields maxmin payoffs.

The condition on δ and k is to ensure that there exists an equilibrium that yields payoffs close enough to a maxmin payoff pair. If there exists an equilibrium in which the minimum payoff is exactly x, then there is no need for approximations: in any stochastically stable equilibrium each player must receive at least x. In the battle of the sexes of figure 3.1, it is always a subgame-perfect equilibrium to play (B, B) and (F, F) alternately, and if δ is large, players' payoffs are approximately 1.5 in each period. There is no condition

on k, since alternating only requires memory 1. In the prisoner's dilemma, we need δ to be large to ensure that cooperating is an equilibrium. There is no condition on k since the grim trigger strategy requires memory 1. In other games, we may require a condition on k to allow a phase of sanctions, as in Fudenberg and Maskin's (1986) folk theorem.

The fact that noise is taken to zero plays a key role in the result. It implies that in any given period, the probability of the player with the highest payoff making a mistake is insignificant relative to the probability of the player with the lowest payoff making one. Therefore when evaluating the stability of a payoff pair it is the payoff of the worse-off player that matters, and states that maximise this will be more stable.

The proof proceeds by a series of lemmas. Throughout, suppose that δ is large enough to guarantee that at least one subgame-perfect equilibrium exists. (Mixed actions are disallowed, so there may not be a stage-game Nash equilibrium.)

First, I show that no state in which a player is uncertain can belong to a limit set of the Markov chain.

Lemma 1. Any state in which at least one of the players is uncertain is not in a limit set of P_M^ε.

Proof. First suppose that both players are uncertain. With positive probability under P_M^0, both players play in a way that is compatible with some subgame-perfect equilibrium s for m periods and then switch to the certain mood with strategies and beliefs corresponding to s. From this new state, the players stay certain, so the initial state is not in a recurrent class of P_M^0.

Next, suppose one player is uncertain and the other is certain. With positive probability under P_M^0, the uncertain player remains uncertain and plays actions that conflict with the certain player's belief for sufficiently long to make

the history inconsistent with the belief. Then with positive probability the certain player becomes uncertain. Note that we require $m > \tau + k$ to guarantee that the uncertain player's actions count as mistakes up to τ periods in the future. Then we are in the case of the previous paragraph, so the initial state is not in a recurrent class of P_M^0. □

Next, I show that every subgame-perfect equilibrium corresponds to a limit set.

Lemma 2. Let s be a subgame-perfect equilibrium. Then there exists $\omega \in \Omega$ such that in every state of ω, players are certain with strategies s and correct beliefs, and the history is fully consistent with s.

Proof. Consider a state $z \in Z$ such that both players are certain with strategies s and correct beliefs, and the history is fully consistent with s. (Recall that an m-history is fully consistent with a strategy profile if there were no deviations from the strategy profile in the past $m - k$ periods.) Following z, no player leaves the certain mood under R_M^0 and the history stays fully consistent with s. Either z is in a recurrent class of P_M^0 or the process eventually reaches a recurrent class of P_M^0, so the claim follows. □

Lemma 2 states that every subgame-perfect equilibrium corresponds to at least one limit set, but it is worth noting that a given subgame-perfect equilibrium may correspond to multiple limit sets. For instance, in the prisoner's dilemma the grim trigger equilibrium corresponds to two limit sets: one in which players cooperate and one in which players defect.

There may also be limit sets in which beliefs are incorrect, for instance when the discrepancies occur on histories that are not reached in the set, or if the interval between each mistake is large compared to the memory of the

process. In lemma 4, we will see that such limit sets are less stable than those in which beliefs are correct.

Before turning to lemma 3, I introduce some further notation. (This follows Ellison 2000.) Let $z, z' \in Z$. The transition $z \to z'$ is *possible* if for all $\varepsilon > 0$, $P_M^\varepsilon(z, z') > 0$. If a transition $z \to z'$ is possible, define its *cost* to be the unique real number $c(z, z')$ such that

$$\lim_{\varepsilon \to 0} \frac{P_M^\varepsilon(z, z')}{\varepsilon^{c(z,z')}}$$

exists and is strictly positive. A finite sequence of states $(z_1, z_2, \ldots, z_n)$ is a *path* from z_1 to z_n if, for all $t < n$, $z_t \to z_{t+1}$ is possible. Its cost is the sum of the pairwise costs, namely

$$c(z_1, z_2, \ldots, z_n) \coloneqq \sum_{t=1}^{n-1} c(z_t, z_{t+1}).$$

Let $\omega, \omega' \in \Omega$. A sequence $(z_1, z_2, \ldots, z_n)$ is a path from ω to ω' if it is a path from z_1 to z_n and $z_1 \in \omega$ and $z_n \in \omega'$. Let $Q(\omega, \omega')$ be the set of all paths from ω to ω'. Define the *resistance* of the transition from ω to ω' to be

$$r(\omega, \omega') \coloneqq \min_{q \in Q(\omega, \omega')} c(q).$$

The resistance $r(\omega, \omega')$ is a measure of how likely the transition from ω to ω' is relative to the size of ε; the higher $r(\omega, \omega')$, the less likely the transition. As $\varepsilon \to 0$, transitions that cost 1 become arbitrarily more likely than transitions that cost 2, say. Define the minimum outgoing resistance or *radius* of ω to be

$$r(\omega) \coloneqq \min_{\omega' \neq \omega} r(\omega, \omega').$$

The radius $r(\omega)$ is a measure of how difficult it is to exit ω.

In general, one can determine the stochastically stable limit sets by analysing the resistances of the various transitions. In this model, the analysis is simplified by the following fact:

Lemma 3. Let $\omega, \omega' \in \Omega$. Then

$$r(\omega, \omega') = r(\omega).$$

Proof. By definition, $r(\omega, \omega') \geq r(\omega)$. I show that ω' can be reached by a path that costs $r(\omega)$. Let z' be some state in ω'; denote its strategy profile by s' and its m-history by h'. Consider any path from ω to some other recurrent class that costs $r(\omega)$. This path must pass through a state in which at least one player is uncertain. Start by copying the path until such a state is reached. Call the uncertain player i. Then the following events occur without further cost: i stays in uncertain mood and plays in such a way as to make j switch from certain to uncertain; for m periods, both players play in such a way that the history is h'; in the following period, they become certain with the strategies and beliefs of z'. Then we have reached ω'. Thus $r(\omega, \omega') \leq r(\omega)$, and therefore $r(\omega, \omega') = r(\omega)$. □

Lemma 3 means that it is as easy to go from ω to ω' as it is to go from ω to ω'', in the sense that both transitions have the same resistance. This is because any transition out of ω involves passing through a state in which at least one player is uncertain, and from any such state there is a path to any limit set that has zero resistance. Lemma 3 will allow us to determine the stochastically stable states from the radiuses alone.

Let $\Omega^* \subseteq \Omega$ be the set of limit sets such that in every state players are certain with correct beliefs and the history is fully consistent with the strategies. Lemma 2 implies that Ω^* is non-empty (provided at least one subgame-perfect equilibrium exists).

I now show that limit sets in which beliefs are incorrect are less stable, in the sense of having a lower radius, than limit sets in which beliefs are correct.

Lemma 4. Let $\omega \in \Omega^*$. Let $\omega' \in \Omega$ be a recurrent class in which at least one belief is incorrect. Then $r(\omega) > r(\omega')$.

Proof. First, I argue that $r(\omega) \geq (\tau + 1)f(\underline{v})$. Denote the strategies of ω by s_1 and s_2. The beliefs are correct, and so are equal to the strategies. For a change of mood to occur, it must be the case that the m-history is inconsistent with s_1 and s_2. Since the history is fully consistent in every state of ω by assumption, the players must make at least $\tau + 1$ mistakes in total, each of which costs at least $f(\underline{v})$. Thus $r(\omega) \geq (\tau + 1)f(\underline{v})$.

Next, I argue that $r(\omega') \leq \tau f(\bar{v})$. Again, denote the strategies by s_1 and s_2. Let s'_j be i's belief about j's strategy. At least one of the players has a wrong belief by assumption; suppose that this is true of i, *i.e.* $s'_j \neq s_j$. I claim that it is sufficient for the players to make τ mistakes in total for i to switch to the uncertain mood: pick one of the k-histories for which s'_j and s_j differ, and suppose that in the next k periods i and j make mistakes in such a way as to arrive at that history – this takes $\ell \leq 2k$ mistakes (remember that $\tau > 2k$ so $\ell < \tau$); in the next period, suppose both i and j do not make mistakes, so that j plays an action that conflicts with s'_j; at this stage, the m-history contains at least $\ell + 1$ mistakes from i's point of view; thereafter suppose the players make mistakes each period until the m-history is inconsistent for i, at which point i switches to uncertain. Then a total of $\ell' \leq \tau$ mistakes have been made, each costing less than $f(\bar{v})$, so $r(\omega') \leq \tau f(\bar{v})$. Moreover, this takes at most $k + 1 + \lceil \tau/2 \rceil$ periods, so $m > 2k + \tau$ is sufficient to ensure that the history is in fact inconsistent for i in the last period.

It then follows from inequality 3.1 that $r(\omega) > r(\omega')$. □

Thus limit sets in which beliefs are incorrect are less stable than those in which beliefs are correct. Intuitively, this is because a player with an incorrect

belief will tend to observe deviations when her opponent isn't making mistakes, which reduces the total number of mistakes required to leave a limit set. Note the role of inequality 3.1 and the assumptions that $\tau > 2k$ and $m > 2k + \tau$ in the proof. Inequality 3.1 implies that the number of mistakes is more important in determining the cost of a transition than the payoffs of the players making the mistakes, so that a limit set with high payoffs but incorrect beliefs has a lower radius than a limit set with low payoffs but correct beliefs. The condition $\tau > 2k$ ensures that it is possible to reach a history in which the incorrect belief diverges from the strategy without adding extra mistakes. And the condition $m > 2k + \tau$ ensures that players recognise mistakes sufficiently far in the past.

Next, I show that any limit set in which players' payoffs do not attain the maxmin payoff x is unstable, in the sense that for appropriate δ and k there exists another limit set that has a higher radius.

Lemma 5. Fix $\lambda > 0$ and let $z \in Z$ be a state in a limit set $\omega \in \Omega^*$ in which at least one player's continuation payoff is weakly less than $x - \lambda$. Then there exist $\underline{\delta}$ and a function $\underline{k}(\delta)$ such that $\delta \geq \underline{\delta}$ and $k \geq \underline{k}(\delta)$ implies that there exists $\omega' \in \Omega^*$ such that $r(\omega') > r(\omega)$.

Proof. Consider the following path out of ω: start in z, and suppose the player with the minimum continuation payoff makes a mistake – this costs weakly less than $f(x-\lambda)$; in each of the next τ periods, the player with the minimum payoff makes a total of τ mistakes, which costs weakly less than $\tau f(x)$, at which point the players become uncertain. Thus

$$r(\omega) \leq \tau f(x) + f(x - \lambda).$$

Note that this requires $m > k + \tau$.

Next, I claim that there exist $\underline{\delta}$ and a function $\underline{k}(\delta)$ such that $\delta \geq \underline{\delta}$ and $k \geq \underline{k}(\delta)$ implies that there exists a subgame-perfect equilibrium s of $\mathscr{H}$ in

which both players' continuation payoffs in every subgame are weakly greater than $x - \eta$, where η is defined to satisfy

$$(\tau + 1)f(x - \eta) > \tau f(x) + f(x - \lambda). \tag{3.2}$$

The proof of the existence of such an equilibrium is along the lines of standard folk-theorem arguments for two-player games, taking account of the fact that strategies are constrained by k. See Fudenberg and Maskin 1986; 1991. (The limit on δ relative to k is necessary because the number of periods of sanctions following a deviation may depend on δ.)

Let ω' be a limit set in which players' strategies and beliefs are s. Then $r(\omega') \geq (\tau + 1)f(x - \eta)$, so by inequality 3.2 we have

$$r(\omega') > r(\omega),$$

as required. □

Roughly speaking, lemma 5 implies that by choosing appropriate k and δ, we can ensure that in every state of the limit set in Ω^* with maximal radius, players receive payoffs arbitrarily close to a maxmin payoff pair. The final step of the proof is to show that limit sets with non-maximal radiuses are not stochastically stable. I use the following theorem:

Theorem (Ellison 2000)**.** Let $\omega, \omega' \in \Omega$. If $r(\omega) > r(\omega', \omega)$, then ω' is not stochastically stable.[8]

Intuitively, Ellison's theorem rules out a given limit set ω' as stochastically stable if there is a second limit set ω that can be easily reached from ω' and that is hard to exit. One might have supposed that $r(\omega, \omega') > r(\omega', \omega)$ would be sufficient to rule out ω', however this is not the case because it might be easy

8 This is theorem 3 in Ellison 2000. One can also use the main theorem, although doing so requires slightly more notation.

to go from ω to other limit sets and easy to reach ω' from other sets, say. It turns out that $r(\omega) > r(\omega', \omega)$ is sufficient to rule out ω'.

In our case, $r(\omega) > r(\omega', \omega)$ is equivalent to $r(\omega) > r(\omega')$ by lemma 3. So limit sets with non-maximal radiuses are not stochastically stable. Then proposition 1 follows.

3.4 Alternative specifications

The selection result in section 3.3 is determined by the probability of making a mistake when players are certain, namely $\varepsilon^{f(U_i)}$. By changing the specification of this probability, one can obtain different selection results.

3.4.1 Maximum of the sum of payoffs

Suppose the probability of a player's making a mistake, in addition to being decreasing in her utility, is decreasing in the number of mistakes she has made in the past. Then one would expect that both players' utilities would contribute to determining the stochastically stable states, since least-cost transitions from one limit set to the other would involve mistakes by both players. I show how this can be used to select a payoff profile that maximises the sum of payoffs rather than a maxmin payoff profile. In section 3.4.2, I show how it can be used to select the Nash bargaining solution.

Consider the following variant of the rule R_M^ε, which I will call R_U^ε. ('U' for 'utilitarian'.) If player i has made n mistakes in the past $m-k$ periods, she makes another one with probability $\varepsilon^{f(U_i,n)}$, where $f : \mathbb{R} \times \{0, 1, \ldots, m-k\} \to \mathbb{R}$ is defined by

$$f(U_i, n) \coloneqq \alpha + \beta U_i + n,$$

where α and β are strictly positive constants that satisfy

$$\alpha > \tau(\beta\bar{v} + \tau) - (\tau + 1)\beta\underline{v}, \text{ and}$$
$$\beta < \frac{1}{\bar{v} - \underline{v}}.$$

The motivation for these conditions is that they guarantee that f satisfies

$$(\tau + 1)f(\underline{v}, 0) > \tau f(\bar{v}, \tau), \text{ and} \tag{3.3}$$

$$\forall n \in \{0, 1, \ldots, \tau - 1\}, f(\bar{v}, n) < f(\underline{v}, n + 1), \tag{3.4}$$

respectively. Assume in addition that $\alpha > -\beta\underline{v}$ to ensure that f is always strictly positive and therefore that the probability of making a mistake is always strictly less than one for $\varepsilon < 1$.

Inequality 3.3 is analogous to inequality 3.1 in section 3.2 and inequality 3.4 means that although a high utility and past mistakes both make further mistakes less likely, the effect of past mistakes dominates the effect of utility. This will mean that the least-resistance path out of a limit set will involve both players making approximately the same number of mistakes.

Let P_U^ε be the process that results from both players using R_U^ε.

Proposition 2. In any stochastically stable limit set of P_U^ε, players' beliefs are correct and players use constant strategies that constitute a subgame-perfect equilbrium of $\mathscr{H}$. Moreover, let

$$y \coloneqq \max_{v \in V^*} v_1 + v_2.$$

Then for any $\lambda > 0$, the sum of continuation payoffs in every state of any stochastically stable limit set is at least $y - \lambda$, provided δ is large enough and k is large enough given δ.[9]

9 It is possible that y does not exist, in which case one can use the supremum instead of the maximum. The same comment applies to proposition 3 below.

The proof is similar to the proof of proposition 1, except that, because of inequality 3.4, any least-cost path out of a limit set must now involve both players making the same number of mistakes (or one player making exactly one more mistake than the other if the total number of mistakes required is odd). Each mistake costs $\alpha + \beta U_i + n$, for some n and U_i, and the n terms add up to a constant number, so the total cost of the path is determined by the sum of all the U_i. (Note the role of the assumption that f is linear.)

3.4.2 Nash bargaining solution

Suppose that, as in section 3.4.1, the probability of making a mistake depends on whether or not the agent has already made a mistake. Now, though, suppose f is defined as

$$f(U_i, n) \coloneqq \alpha + \beta \ln U_i + n,$$

where n is the number of mistakes already made by i in the past $m - k$ periods and U_i is the continuation payoff anticipated by i. The constants α and β are again strictly positive and defined to make inequalities 3.3 and 3.4 hold and to ensure that the probability of making a mistake is strictly less than one for $\varepsilon < 1$. Assume payoffs are normalised to be strictly positive, so that f is well-defined.

Call this rule R_N^ε, and call the resulting process P_N^ε. I claim that R_N^ε selects the Nash bargaining solution of $\mathscr{G}$.[10]

Proposition 3. In any stochastically stable limit set of P_N^ε, players' beliefs are correct and players use constant strategies that constitute a subgame-perfect

10 Recall that the Nash bargaining solution of a game is defined to be

$$\max_{v \in V^*} v_1 v_2,$$

after Nash 1950.

	A	B
A	a_{11}, b_{11}	a_{12}, b_{12}
B	a_{21}, b_{21}	a_{22}, b_{22}

Figure 3.2: Payoff matrix of a generic 2×2 game.

equilbrium of $\mathscr{H}$. Moreover, define

$$y \coloneqq \max_{v \in V^*} v_1 v_2.$$

Then for any $\lambda > 0$, the product of continuation payoffs in every state of any stochastically stable limit set is at least $y - \lambda$, provided δ is large enough and k is large enough given δ.

The argument is similar to the argument in section 3.4.1 except that the sums become products.

3.4.3 Risk dominance

Finally, I consider the case in which the probability of making a mistake depends on the player's current stage-game payoff rather than her continuation payoff. This changes the results radically. I present a specification that leads to a generalisation of risk dominance.

Before presenting the learning rule, I define risk dominance and introduce a generalisation of the concept. Consider a 2×2 coordination game with payoff matrix given in figure 3.2. By coordination game, I mean that (A, A) and (B, B) are Nash equilibria; that is, the inequalities

$$a_{11} \geq a_{21}, b_{11} \geq b_{12}, a_{22} \geq a_{12}, \text{ and } b_{22} \geq b_{21}$$

hold.

The equilibrium (A, A) is *risk dominant* if

$$(a_{11} - a_{21})(b_{11} - b_{12}) \geq (a_{22} - a_{12})(b_{22} - b_{21}),$$

and (B, B) is risk dominant if the reverse inequality holds. In words, an equilibrium is risk dominant if it maximises the product of the losses from unilateral deviation.[11]

The usual generalisation of risk dominance to larger games is to say that a pure Nash equilibrium is risk dominant if it pairwise risk dominates every other pure Nash equilibrium. Kandori, Mailath and Rob (1993) and Young (1993) showed how stochastic evolutionary models could select the risk dominant equilibrium in 2×2 games. Young (1993) also showed why a pairwise risk-dominant equilibrium might not be selected in larger games.

For 2×2 games, one can motivate risk dominance in terms of probabilities of errors: If deviations from best responses are less likely the higher the cost of the mistake, then $a_{11} - a_{21}$ and $b_{11} - b_{12}$ are measures of the stability of (A, A). If mistakes are viewed as in some way equally likely to be made by both agents, irrespective of, say, cardinal differences in payoffs, then

$$(a_{11} - a_{21})(b_{11} - b_{12})$$

is a measure of the stability of (A, A). Similarly, $(a_{22} - a_{12})(b_{22} - b_{21})$ is a measure of the stability of (B, B).

Given this interpretation of risk dominance, consider the following generalisation of the concept to larger games. Let us return to our general two-player game $\mathscr{G}$, and let N be the set of its pure Nash equilibria. Suppose N is non-empty. For any pure-strategy Nash equilibrium a, define the *stability* of a to

11 The concept of risk dominance was introduced by Harsanyi and Selten (1988). This discussion draws on Young 1998, section 4.1.

be

$$s(a) \coloneqq \left(u_1(a) - \max_{a_1' \neq a_1} u_1(a_1', a_2)\right)\left(u_2(a) - \max_{a_2' \neq a_2} u_2(a_1, a_2')\right).$$

Note that, in the 2×2 game of figure 3.2,

$$s((A,A)) = (a_{11} - a_{21})(b_{11} - b_{12}), \text{ and}$$
$$s((B,B)) = (a_{22} - a_{12})(b_{22} - b_{21}),$$

so that an equilibrium in such a game is risk dominant if and only if it has maximal stability.

The intuition for this notion of stability is as follows: Suppose deviations from best responses are less likely the costlier the deviation, and moreover that, for any given player, the probability of a very costly mistake is negligible compared to the probability of a less costly mistake. Then in any equilibrium a, when player 1 makes mistakes she will almost always make mistakes to $\arg\max_{a_1' \neq a_1} u(a_1', a_2)$, and we can ignore other deviations. Therefore the first term in $s(a)$ is a measure of the stability of a with respect to player 1's mistakes. Similarly, player 2 will tend to make mistakes to $\arg\max_{a_2' \neq a_2} u(a_1, a_2')$, so the second term in $s(a)$ is a measure of the stability of a with respect to player 2's mistakes. If, in addition, both players are in some way equally likely to make mistakes, then $s(a)$ is a measure of the stability of a.

I present an intuitive specification of the learning rule for which the stochastically stable outcome is to repeatedly play the Nash equilibria of the stage game that have maximal stability.

Suppose that, as in sections 3.4.1 and 3.4.2, the probability of making a mistake depends on whether or not the player has already made a mistake. Now, additionally, the probability of i's making a mistake in a given period depends not on discounted future payoffs, but on current payoffs: Let a_j^t be the action i expects j to play at t given her belief s_j' and the current k-history. Let

a_i^t be the action prescribed by s_i given the current k-history. Following s_i yields $u_i(a_i^t, a_j^t)$ in period t. Let a_i^* be an action that would maximise i's stage-game payoff in period t among actions other than a_i^t; that is,

$$a_i^* \in \arg\max_{a_i \neq a_i^t} u_i(a_i, a_j^t).$$

Finally, let g_i be the difference in period-t payoff between playing a_i^t and a_i^*:

$$g_i = u_i(a_i^t, a_j^t) - u_i(a_i^*, a_j^t).$$

If a_i^t is a best response in the stage game, then g_i is weakly positive. Now, define the probability of making a mistake in period t to be $f(g_i, n)$, where

$$f(g_i, n) = \begin{cases} \alpha + \beta \ln g_i + n & \text{if } g_i > 0, \\ \gamma & \text{if } g_i = 0, \text{ and} \\ \lambda & \text{otherwise,} \end{cases}$$

where n is the number of mistakes made in the past $m - k$ periods, as in sections 3.4.1 and 3.4.2, and α, β, γ and λ are strictly positive constants. Suppose the constants α and β are chosen to satisfy conditions analogous to inequalities 3.3 and 3.4; suppose α is large enough to ensure that mistakes are always more likely when g_i is negative; and suppose α, γ and λ are large enough that the probability of making a mistake is always less than one for $\varepsilon < 1$. Finally, suppose $\gamma > \lambda$.

Call this specification of the learning rule R_{RD}^{ε}. As above, let P_{RD}^{ε} be the process that results from both players acting according to R_{RD}^{ε}.

Proposition 4. In any stochastically stable limit set of P_{RD}^{ε}, players' beliefs are correct and players use constant strategies that constitute a subgame-perfect equilbrium of $\mathscr{H}$. Moreover, the strategies involve playing a stage-game Nash equilibrium with maximal stability in every state of the limit set.

The proof of proposition 4 is along similar lines to the proofs of the previous propositions. A Nash equilibrium with high stability will lead to fewer mistakes than one with low stability, which implies that in any stochastically stable state players' actions will correspond to Nash equilibria with maximal stability.

Note that there is no condition on k and δ in this variant. This is because a set of stage-game Nash equilibria played in sequence is always a subgame-perfect equilibrium of the repeated game.

3.5 Conclusion

I have shown how to construct a stochastic learning rule that results in players playing a subgame-perfect equilibrium of the repeated game most of the time in the long run. The payoff profiles selected depend on how players make mistakes. Different specifications yield the maxmin, or Rawlsian, outcome, the maximum of the sum of payoffs, the Nash bargaining solution, and a generalisation of risk dominance.

It is worth noting that the maxmin outcome, the maximum of the sum of payoffs and the Nash bargaining solution all involve maximising some symmetric social welfare function that is increasing in the players' utilities. As a result, they are all Pareto efficient. A risk-dominant equilibrium, however, may be inefficient. The key difference between the learning rule that leads to risk dominance and the three others is that in the former, players' mistakes are myopic; that is, they are determined by stage-game payoffs. In the latter, they are determined by repeated-game payoffs. This is intuitive: if players are greedy and tend to make mistakes when they see a short-term gain, the process will be pushed towards action profiles that avoid short-term gains, even if they are inefficient; if players make mistakes when they are discontent and expect a low utility, equilibria in which players have low utility will be unstable, and the

process will favour efficient outcomes. In practice, both types of mistake may coexist, each pushing the process towards different outcomes.

Finally, the model is made tractable by the use of 'fast and slow search'. Players use a fixed strategy most of the time, but sometimes, when their beliefs are proven to be incorrect, they fall into phases of more random search. This means that the stochastically stable states of the process can be determined by looking at the radiuses of the limit sets, which simplifies analysis considerably.

Bibliography

Akerlof, G. A. 1980. 'A theory of social custom, of which unemployment may be one consequence'. *Quarterly Journal of Economics* 94: 749–775.

Arrow, K. J. 1994. 'Methodological individualism and social knowledge'. *American Economic Review* 84: 1–9.

Aumann, R. J., and L. S. Shapley. 1976. 'Long-term competition: A game-theoretic analysis'. Working paper.

Axelrod, R. 1986. 'An evolutionary approach to norms'. *American Political Science Review* 80: 1095–1111.

Axelrod, R., and W. D. Hamilton. 1981. 'The evolution of cooperation'. *Science* 211: 1390–1396.

Banks, S. 2012. *Duels and duelling.* Oxford: Shire.

Barlo, M., G. Carmona and H. Sabourian. 2009. 'Repeated games with one-memory'. *Journal of Economic Theory* 144: 312–336.

———. 2016. 'Bounded memory Folk Theorem'. *Journal of Economic Theory* 163: 728–774.

Bhaskar, V. 1998. 'Informational constraints and the overlapping generations model: Folk and anti-folk theorems'. *Review of Economic Studies* 65: 135–149.

Billacois, F. 1986. *Le duel dans la société française des XVIe–XVIIe siècles.* Paris: École des Hautes Études en Sciences Sociales.

Black, D. 1976. *The behaviour of law.* New York: Academic Press.

Carvalho, J.-P. 2013. 'Veiling'. *Quarterly Journal of Economics* 128: 337–370.

Chesnais, J.-C. 1981. *Histoire de la violence.* Paris: Robert Laffont.

Coleman, J. S. 1990. *Foundations of social theory.* Cambridge, MA: Harvard University Press.

Deb, J. 2012. 'Cooperation and community responsibility: A folk theorem for repeated random matching games'. Working paper.

Deb, J., and J. González-Díaz. 2014. 'Enforcing social norms: Trust-building and community enforcement'. Working paper.

Ellison, G. 1994. 'Cooperation in the prisoner's dilemma with anonymous random matching'. *Review of Economic Studies* 61: 567–588.

———. 2000. 'Basins of attraction, long-run stochastic stability, and the speed of step-by-step evolution'. *Review of Economic Studies* 67: 17–45.

Elster, J. 1989. 'Social norms and economic theory'. *Journal of Economic Perspectives* 3: 99–117.

Foster, D. P., and H. P. Young. 1990. 'Stochastic evolutionary game dynamics'. *Theoretical Population Biology* 38: 219–232.

———. 2003. 'Learning, hypothesis testing, and Nash equilibrium'. *Games and Economic Behavior* 45: 73–96.

Fudenberg, D., and E. Maskin. 1986. 'The Folk Theorem in repeated games with discounting or with incomplete information'. *Econometrica* 54: 533–554.

Fudenberg, D., and E. Maskin. 1991. 'On the dispensability of public randomization in discounted repeated games'. *Journal of Economic Theory* 53: 428–438.

Hammerstein, P., and O. Leimar. 2015. 'Evolutionary game theory in biology'. In *Handbook of game theory with economic applications,* vol. 4*,* edited by H. P. Young and S. Zamir, 575–617. Amsterdam: North-Holland.

Harrington, J. E. 1995. 'Cooperation in a one-shot prisoners' dilemma'. *Games and Economic Behavior* 8: 364–377.

Harsanyi, J. C., and R. Selten. 1988. *A general theory of equilibrium selection in games.* Cambridge, MA: MIT Press.

Hofbauer, J., and K. Sigmund. 1988. *The theory of evolution and dynamical systems.* Cambridge: Cambridge University Press.

Hopton, R. 2007. *Pistols at dawn: A history of duelling.* London: Piatkus.

Kalai, E., and E. Lehrer. 1993. 'Rational learning leads to Nash equilibrium'. *Econometrica* 61: 1019–1045.

Kandori, M. 1992. 'Social norms and community enforcement'. *Review of Economic Studies* 59: 63–80.

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

Lewis, D. 1969. *Convention: A philosophical study.* Cambridge, MA: Harvard University Press.

Mackie, G. 1996. 'Ending footbinding and infibulation: A convention account'. *American Sociological Review* 61: 999–1017.

Maynard Smith, J., and G. R. Price. 1973. 'The logic of animal conflict'. *Nature* 246: 15–18.

Merry, S. E. 1984. 'Rethinking gossip and scandal'. In *Toward a general theory of social control,* edited by D. Black, vol. 1. London: Academic Press.

Nash, J. 1950. 'The bargaining problem'. *Econometrica* 18: 155–162.

Nye, R. A. 1993. *Masculinity and male codes of honor in modern France.* Oxford: Oxford University Press.

Okuno-Fujiwara, M., and A. Postlewaite. 1995. 'Social norms and random matching games'. *Games and Economic Behavior* 9: 79–109.

Pradelski, B. S. R., and H. P. Young. 2012. 'Learning efficient Nash equilibria in distributed systems'. *Games and Economic Behavior* 75: 882–897.

Renard, J. 1960. *Journal.* Paris: Gallimard, « Bibliothèque de la Pléiade ».

Rubinstein, A. 1979. 'Equilibrium in supergames with the overtaking criterion'. *Journal of Economic Theory* 21: 1–9.

Samuelson, P. A. 1958. 'An exact consumption-loan model of interest with or without the social contrivance of money'. *Journal of Political Economy* 66: 467–482.

Sandholm, W. H. 2010. *Population games and evolutionary dynamics.* Cambridge, MA: MIT Press.

Schelling, T. C. 1971. 'Dynamic models of segregation'. *Journal of Mathematical Sociology* 1: 143–186.

Stone, L. 1965. *The crisis of the aristocracy, 1558–1641.* Oxford: Oxford University Press.

Takahashi, S. 2010. 'Community enforcement when players observe partners' past play'. *Journal of Economic Theory* 145: 42–62.

Taylor, P. D., and L. B. Jonker. 1978. 'Evolutionarily stable strategies and game dynamics'. *Mathematical Biosciences* 40: 145–156.

Twain, M. (1880) 1997. *A tramp abroad.* London: Penguin.

Weibull, J. W. 1995. *Evolutionary game theory.* Cambridge, MA: MIT Press.

Young, H. P. 1993. 'The evolution of conventions'. *Econometrica* 61: 57–84.

———. 1998. *Individual strategy and social structure.* Princeton: Princeton University Press.

———. 2009. 'Learning by trial and error'. *Games and Economic Behavior* 65: 626–643.

Young, H. P., and M. A. Burke. 2001. 'Competition and custom in economic contracts: A case study of Illinois agriculture'. *American Economic Review* 91: 559–573.