

THEORIES OF LEARNING IN ECONOMICS

Daniel Sgroi

Nuffield College, University of Oxford

Trinity Term 2000



For my wife and family

ABSTRACT

How should we model learning behaviour in economic agents? This thesis addresses this question in two distinct ways. In the first set of chapters the assumption is that agents learn through the observation of others. They use Bayesian updating which together with specific informational assumptions can generate the problem known as *herding* with the potential for significant welfare losses. In the final set of chapters the agent is instead modelled as learning by example. Here the agent cannot learn by observing others, but has a pool of experience to fall back on. This allows us to examine how an economic agent will perform if he sees a particular economic situation (or game) for the first time, but has experience of playing related games. The tool used to capture the notion of learning through example is a neural network. Throughout the thesis the central theme is that economic agents will naturally use as much information as they can to help them make decisions. In many cases this should mean they take into consideration others' actions or their own experiences in similar but non-identical situations. Learning throughout the thesis will be rational or bounded-rational in the sense that either the best possible way to learn will be utilized (so players achieve full rational play, for example, through Bayesian updating), or a suitable *local error-minimizing algorithm* will be developed (for example, a rule of thumb which optimizes play in a subclass of games, but not in the overall set of possible games). Several themes permeate the whole thesis, including the scope for firms or planners to manipulate the information that is used by agents for their own ends, the role of rules of thumb, and the realism of current theories of learning in economics.

ACKNOWLEDGEMENTS

I should like to start by thanking the staff, fellowship and students at Nuffield for four enjoyable years as an MPhil and then DPhil student. My funding came from the Economic and Social Research Council, who also funded my period as a master's student, so I am doubly grateful. Some of this thesis appeared in the *Nuffield College Discussion Papers in Economics*, a version of chapter 3 is forthcoming in *Games and Economic Behavior* and I have presented parts at the *Seminar in Group Theory* at St. John's College, Oxford, the *Gorman Workshop in Economics* and the *Experimental Economics Workshop*, at the Economics Department in Oxford. I travelled to Amsterdam for the *Young Economist's Conference* and Bilbao for the *First World Congress of the Game Theory Society* to present some of the work in this thesis, so thanks to the many organisers and other contributors for a chance to get some valuable feedback.

Paul Klemperer, my university supervisor was of course invaluable and provided many useful comments throughout the thesis. Peter Sorensen introduced me to herding, and showed me how much work there was still to do. For help, advice and numerous comments I would like to thank Michael Bacharach, Simon Board, Ernesto Dal Bo, Huw Dixon, Jorg Oechssler, Sophocles Mavroeidis, Meg Meyer, John Morgan, Lucy White, and my college advisor, Hyun Shin.

My only co-author in any of this thesis, Daniel J. Zizzo, was heavily involved in the work detailed in chapters 7 and 8 of the thesis. I would also like to thank Daniel Zizzo for many useful comments relating to chapter 6. David Gill, Marc Chapman and my wife, Rosemary, provided me with invaluable research assistance for chapter 6. I am also very grateful to Chris Wallace and David P. Myatt for answering numerous format-related

questions (this was my first substantial piece of work produced with LaTeX), and Jurgen Doornik for the excellent OxEdit! Neural network simulations in chapter 8 were made using the TLearn software package.

I must thank Gerald, my brother, and his family, Gill, Hannah and Gregory for just being themselves. A special thanks must go to my two favourite women: Antonietta, my mother and Rosemary, my wife. Rosemary read much of this thesis and for that I am eternally grateful. Finally, Alfio, Ada and Concetto, I hope you realize how important a part you played even though you left this earth when I was only a child.

MAIN ABBREVIATIONS

0SD	0-Level Strict Dominance
1SD	1-Level Strict Dominance
2SD	2-Level Strict Dominance
BR	Best Response (or reply)
DTM	Deterministic Turing Machine
ESS	Evolutionary Stable Strategy
GI	Game Harmony Index
GMA	Global Error-Minimizing Algorithm
IID	Independent and Identically Distributed
LMA	Local Error-Minimizing Algorithm
MPD	Maximum Payoff Dominance
MSE	Mean-Squared Error
NDTM	Non-Deterministic Turing Machine
NE	Nash Equilibrium (or equilibria)
NNG	Nearest Neighbour Algorithm
OSP	One-Step Property
PBE	Perfect Bayesian Equilibrium
PNE	Pure Nash Equilibrium (or equilibria)
PSPD	Pure Sum of Payoff Dominance
RMS	Root-Mean Squared

CONTENTS

1	Introduction	1
1.1	Motivation	1
1.2	Learning by Observation	3
1.3	Learning by Example	5
1.4	Overview of the Thesis	6
2	The Herd Externality	10
2.1	Introduction	10
2.2	The Basic Model	12
2.3	The Continuity of Action Spaces	24
2.4	Stochastic Environments	30
2.5	Conclusions	35
3	Optimizing Information in the Herd	37
3.1	Introduction	37
3.2	The Model	39
3.3	Payoffs and Consumer welfare	44
3.4	The Social Planner	49
3.5	The Firm	56
3.6	Conclusions	63
4	Herding in Time and Choice	73
4.1	Introduction	73

CONTENTS

vii

4.2	The Precision of Signals	73
4.3	Models of Discrete Choice	75
4.4	Choosing from a Continuum	79
4.5	A Critical Appraisal: Addressing Some Concerns	82
5	Models of Endogenous-Time Herding	85
5.1	Introduction	85
5.2	Introducing a Simple Model	87
5.3	The Full Model: Incorporating Irreversibility	90
5.4	Will Information Revelation Help?	96
5.5	Conclusions	109
6	An Experiment in Endogenous Time	118
6.1	Introduction	118
6.2	The Experimental Literature	119
6.3	Experiment Design	121
6.4	Optimal Behaviour	123
6.5	Results	125
6.6	An Alternative Condition	130
6.7	Conclusions	133
7	Neural Networks and Bounded-Rational Learning	147
7.1	Introduction	147
7.2	Neural Networks and Learning	152
7.3	The Model	158
7.4	The Neural Network Player	161
7.5	Inadequate Learning	166
7.6	Algorithm Complexity and Intractability	170
7.7	Rules of Thumb	175
7.8	Conclusions	178

CONTENTS

viii

8 Neural Networks in Practice	181
8.1 Introduction	181
8.2 Alternatives to Nash	189
8.3 Game Features	198
8.4 Multiple Equilibria	205
8.5 Games with No Equilibria	213
8.6 Conclusions	217
9 Conclusion	220
9.1 Imperfect Learning	220
9.2 Future Work	221

CHAPTER 1

INTRODUCTION

1.1 MOTIVATION

Why do the French drive on the right-hand side of the road, while the British drive on the left? Why do teenagers wear jeans while their fathers wear chinos? Why is one restaurant the place to be, while another is out of fashion? Why were bell-bottomed jeans ever fashionable? How can a technology have risen to prominence when most people have always known it to be inferior? All of these questions can be answered in many different ways, with reference to many different theories in economics. However, *herding theory*, a branch of information economics developed simultaneously by Banerjee (1992) and Bikchandani, Hirshleifer, and Welch (1992) can answer all of these questions of *localized conformity*, and many more, in a unified and intuitive way.

How can a good investment opportunity be missed by two firms with good information about the profits to be made? Why would a group of decision-makers delay making a choice despite the costs of doing so? Even this second set of questions fall within the range of herding theory.

Finally, consider the following third set of questions. Should a firm with a new product release it to the entire market on the same day or pre-release to a select set of customers first? Should movie premieres be made available to members of the press and high profile celebrities or should films simply be released world-wide on the same day? Should a firm's product be released on the same day in several markets or be released sequentially in

different markets? Is it optimal for every country in the European Community to conduct separate drug trials or should they pool and have one drug trial? Is it sensible for the government to slowly release new drug treatments to doctors or should there be a core release to a select group of doctors first? Once again these issues can be resolved with the careful application of herding theory.

In all of these cases the unifying feature is the role of learning through the observation of others. When taking an economic decision over a set of choices, an agent might well benefit from observing the choices made by others. In a sense the agent is learning about the payoff by discovering some part of the information held by other agents, inferable from their actions. An agent may know the value of each choice with certainty, or there may be no scope for observation. However, in many real world contexts the agent can learn from others. In a game-theoretic sense the agent will benefit from observing the choice of strategy employed by others who are playing identical games. Jackson and Kalai (1997) call such games *recurring games*, and it is clear that when making decisions relating to our first set of questions, there is plenty of scope to benefit from the observation of others' actions. The third set of questions are similar but suggest an alternative approach. Given that sensible agents will try to gain from observation, perhaps a firm, or government can gain by limiting the scope for such observation. The first part of this thesis deals with the first and third sets of questions. In the first instance we see that the role of observation is to increase the information available to followers, but at the same time early incorrect decisions can generate considerable welfare losses for the entire group of decision-makers as later decision-makers can but assume that such decisions were good. The resulting chain reaction, called the herd, produces an externality attributable to early decision-makers whose choices affect not just their own welfare, but those of later movers, named the *herd externality* by Banerjee (1992).

The second set of questions requires a little more thought. Here we are dealing with agents' ability to decide not what to do, but when to act. Nevertheless we can still model their actions with reference to herding theory, by considering the choice of time to act to be part of the choice of strategy. However, it is now the case that the precise timing

of one agent's decision can reveal much useful information to anyone else who is facing the same choice. The second part of this thesis looks at such questions. Once again we note the potential for a form of the herd externality to arise. In this case the impact may be a complete failure to make a decision at all, or the correct decision may be made, but after considerable delay. The classic example comes in the theory of irreversible investment under uncertainty, where an investor may well delay his decision in order to try to increase his understanding of the payoffs of different projects by observing the actions of other decision-makers.¹

1.2 LEARNING BY OBSERVATION

The herding or informational cascade literature can be said to be a response to two needs within economics. On a small scale there is a need to understand what Bikchandani, Hirshleifer, and Welch (1992) call *localized conformity*: why certain types of behaviour seems to rule in one area but not another; there is a need to understand the development, life and ending of a fad or cultural change. With the aid of herding theory we can explain why there may be a clustering of demand around one good when others are available, and even why a lower quality good may survive in the marketplace while a higher quality product may fail. In other words we can explain herds and also explain socially inefficient clustering.

A second, grander, more daring aim concerns not specific instances of herd behaviour, but the process of *decision making under uncertainty* itself. At the turn of the century it eventually became accepted that goods may have no implicit value, but rather their value is a function of demand (willingness to pay) and supply (cost). Similarly, it may eventually be accepted that choice under uncertainty in a world of many heterogeneous agents cannot be made in isolation of the marketplace. Observing others reveals information; to ignore this possibility and enter the marketplace with a rigid believe in the optimal bun-

¹For consistency, throughout this thesis the convention will be that agents or players will be referred to as "he"; this is meant as a gender-neutral pronoun.

market is therefore inefficient whenever there is doubt about the value of an item and there exists the possibility that not all information will become publicly known. In general equilibrium in an Arrow-Debreu world, trades are made at the beginning of time, and uncertainty is dealt with via contingent commodity trading. Radner (1982) introduced the idea of Bayesian updating given prices in a general equilibrium under uncertainty setting, and assuming rational expectations he found that the equilibrium would be informationally efficient. However, much of the recent theoretical microeconomic literature has focused on the potential for asymmetric information to produce inefficient outcomes. Therefore, the herding literature can be seen as the beginnings of an attempt to consider economic activity in a world of uncertainty, when others' actions are informative about the quality of an item, but allowing for the possibility that some private information may be lost. At its ambitious height, this may represent the origin of an alternative theory of decision making under uncertainty, evident in Bikhchandani *et al*'s claim to be able to explain fads, fashion, customs and cultural change all in terms of herding. With this in mind, the herding literature can also be seen as part of research into the processing of information, which includes evolutionary game theory, and the literature on complexity and bounded rationality.

The early 1990s saw a growing interest in the herding phenomenon. In particular two papers initiated the current research output, Banerjee (1992) and Bikhchandani, Hirshleifer, and Welch (1992). Banerjee focused on what he called the herd externality: the loss of valuable information to later individuals in a sequence of decision-making agents. Bikhchandani *et al* referred to informational cascades in which individuals, despite acting perfectly rationally, would be trapped. Both restricted themselves to sequences of individuals in unchanging environments, with finite support for private signals and specific belief processes. This thesis examines the Bikhchandani *et al* type of model in depth, concentrating on the seminal papers that defined the problem, and briefly looking at some advances in the literature, in particular Lee (1993) where the importance of continuity of the action-space was first noted, and the possibility of non-stationary environments of the type suggested by Moscarini, Ottaviani, and Smith (1996). Two seminal papers,

Chamley and Gale (1994) and Gul and Lundholm (1995), examined the scope for endogeneity in the ordering process. These two papers differ in much the same way that Bikhchandani *et al* and Lee differ, based upon the continuity of action spaces, and we see that even in an endogenous-timing model inefficiency derives from an incomplete action-space. Finally, Jackson and Kalai (1997) formalize the class of games for which the herd externality is likely to apply as *recurring games*.

This thesis adds to this body of research by extending herding theory in several directions. Examples include: the scope for the manipulation of observational learning by firms and governments; an examination of the role of exogenous information in endogenous-time herding models; and a first experiment testing the importance of observational learning in an endogenous-time herding model in the laboratory.

1.3 LEARNING BY EXAMPLE

What if the previous series of observed games were actually played by the agent himself? In effect, rather than considering a sequence of players, what if a single player plays the game many times? This case of course refers to the well-researched case of the *repeated game*. Alternatively, we might consider how many times the player is willing to repeat the game, and perhaps try different strategies to test the outcome. Then we travel the well-trodden road of *optimal experimentation*. This can be neatly linked to the herding literature, as in Smith and Sorensen (1997b). Finally, we might wonder what potential for learning might exist if the player begins by selecting a random action or follows an *ad hoc* decision-rule and then simply observes the outcome, changing his action in response to the success or failure of his previous set of choices. This is the question asked in the *evolutionary games* literature.

The final part of this thesis moves away from all of these questions to probe a similar but less well understood problem in the theory of learning: what happens if a player repeatedly faces a series of completely different games? In evolutionary game theory this provides no time to converge to any particular steady state, or long-run behaviour. It

is not usefully seen as a repeated game, as the game changes every period. Finally, this is hardly a form of experimentation since the player will never observe the same basic game twice. This thesis argues that despite this we do not need to assume that the player will simply make the correct choice every time or else simply chose randomly. We can construct a theory of learning, not of how to play a particular game, but rather, of how to play games in general. More specifically, the thesis examines the idea that a player can learn to recognize a Nash equilibrium in a previously unseen game, through the experience of having played similar, but non-identical, games. The tool used here is not Bayesian learning as in the herding models in the first two parts of the thesis, but a neural network, which usefully models the building of a set of decision rules based on earlier decision-making. Put more succinctly, it is a good way of capturing the idea of learning by example.

A neural network coupled with a biologically plausible model of learning by example known as backpropagation produces some surprising results. A player having observed a sequence of example games with a Nash equilibrium highlighted each time, will eventually be able to pick Nash equilibria in never before seen games with a high degree of success. Furthermore, the success rate while well above that achievable through random guessing is bounded by the difficulty of the learning process, and what results is a new way of modelling bounded-rational play.

1.4 OVERVIEW OF THE THESIS

This thesis is divided neatly into three parts. All three parts deal in slightly different ways with the question of how to model learning, and what results obtain. Part I focuses on sequential herding. Chapter 2 examines several seminal herding papers, focusing on the nature of the herd externality, and some of the basic results in this chapter will be cited in later chapters. Chapter 3 deals with a new question: how can firms or social planners exploit the existence of the herd externality? This is answered by allowing the structure of decision-making to be manipulated by the firm or social planner, in particular through

the use of *guinea pigs*, agents who are forced to make their decisions early, and so provide extra information for those who decide later.

Part II of the thesis moves into the realm of endogenous timing. Here the problem of what to do is joined by the question of when to do it! Once more learning from others is the focus of attention, and a new form of the herd externality obtains. Chapter 4 examines this alternative form of externality, with reference to the seminal papers in this field. Chapter 4 also sets the scene for the work in chapter 5, by suggesting a series of issues which we might like to address concerning the models of chapter 4. In particular, chapter 5 once again looks at the scope for manipulating the herd externality through the provision of extra information. Chapter 6 presents some new experimental evidence which highlights the importance of herding considerations, even in the more complex world of endogenous-time herding, where players need to decide when to act as well as how to act.

Part III switches the emphasis to an examination of learning by example. Here the focus is on learning *how to play*, rather than learning *what to play*, and the tool to be used is not herding theory, but instead a neural network model of decision-making. The last chapter of part II began to turn the focus away from modelling agents as purely rational to a discussion of rules of thumb. There is however some difficulty in differentiating between rules of thumb which approach optimal behaviour and may in fact reproduce the outcome of optimal behaviour in a subclass of games, and purely *ad hoc* rules of thumb mainly derived from the casual observation of behaviour. The third part of the thesis details an attempt to employ one of the most widely used models of human behaviour in the biological and neurological sciences, the neural network, to develop the idea of a rule of thumb which is close to optimal. Chapter 7 describes the functioning of a neural network designed to play 3×3 normal form games: by training the network to isolate the Nash equilibrium strategies in a set of examples, the network develops a generalized ability to find Nash equilibrium strategies in games it has never seen before. It is argued that while networks will learn, they will not do so perfectly, so they represent a feasible means of modelling bounded-rational play in games. Chapter 7 can be seen as presenting

a hypothesis that networks will play in a bounded way focusing on *local error-minimizing algorithms* which capture the notion of a rule of thumb that is close to optimal in a subset of the space of possible games. This is tested using a simulated neural network in chapter 8. Chapter 8 concludes by arguing that the network will stop learning when it is satisfied it has done well enough, and will often learn until it has developed a general method for playing games which results in the selection of Nash equilibria about 60% of the time.

Finally, chapter 9 provides a general conclusion, and details some possible further applications of the ideas and methods developed within the thesis.

PART I
SEQUENTIAL HERDING

CHAPTER 2

THE HERD EXTERNALITY

2.1 INTRODUCTION

If an agent can benefit from observing the actions of his predecessors, how can this ever be welfare-reducing? This chapter briefly reviews the major seminal papers that addressed this question, and lists the main results in herding theory, many of which will be used and cited in later chapters.

2.1.1 An Example of a Herd

Let us begin with a simple example of a herd. Consider two restaurants, A and B. Now consider a stream of agents arriving sequentially at the doors of the two restaurants. They all have a private signal (perhaps a newspaper review or information about the restaurants passed on from a friend). They consider all private information to be of the same quality - their own signal is not *a priori* any better than anyone else's. They can add to the information contained within their signal by observing the action (not the signal) of their predecessors. The first agent has only his (informative) signal to guide him, so will go where his signal indicates, say into restaurant A. The next agent has both his private signal and also the public information relating to the action of the first agent, which in this case perfectly reveals his signal - it clearly indicated that restaurant A was superior to restaurant B. Let us assume that agent 2 also has a signal favouring restaurant A. He will therefore also enter restaurant A. The third agent arrives, observes

the actions of the first two agents. Let us consider what happens if his signal suggests that restaurant B is superior. He has observed the actions of his two predecessors, and can infer that the first mover had a signal favouring restaurant A. The second mover is difficult: he may also have had a signal favouring restaurant A, which would account for his actions. However, he may have had a signal favouring restaurant B, but this would render his net information neutral - so he might still go for restaurant A. To deal with this we assume that a player will go for his own information if his net information is neutral.² The third player can now infer two signals suggesting that restaurant A is superior. This swamps his own signal and he will (rationally) go for restaurant A. He is the first to be trapped in what Bikchandani, Hirshleifer, and Welch (1992) call an informational cascade. With no new information, agent 4 will of course now also enter restaurant A even if his signal suggests that he should not, as will agent 5, and so on, all subsequent agents herding into restaurant A.

It is of course perfectly possible that restaurant B is in fact superior. Consider, for example, defining superiority with reference to the number of positive reviews. It might be that restaurant A was considered superior by 40% of reviewers and restaurant B by 60%, but a predominance of those who had read reviews by those favouring restaurant A early on might well produce a cascade in its favour. Note that there is a huge loss of information which would eventually reveal the superiority of restaurant B if signals and not actions were observable. This is our first glimpse of the *herd externality*.

2.1.2 Overview of the Chapter

This chapter sets out to examine the role of the herd externality as laid out by the two seminal papers of Banerjee (1992) and Bikchandani, Hirshleifer, and Welch (1992). There have been several recent papers which develop the herding literature in various new directions, and Gale (1996) provides a survey of many of these papers. The role of this chapter is not to simply to act as a rival survey, rather it confines itself to only a few very

²This was the assumption in Banerjee (1992) and is also justified by the experimental evidence in Anderson and Holt (1997), as detailed in chapter 6.

papers, and examines these in considerable depth, the aim being to cite definitions and results which will be used in the later chapters. Lee (1993) and Moscarini, Ottaviani, and Smith (1996) develop herding in two interesting directions, and these will be examined here. Lee (1993) examines when the herd externality applies in a more general context, and Moscarini, Ottaviani, and Smith (1996) examines the role of a stochastic state of the world. In this case while agents may benefit from observing the actions of their predecessors, the state is changing, so earlier information becomes less relevant.

2.1.3 A Note on Terminology

Herding theory is blessed with multiple seminal papers, many of which developed ideas independently. However, an unfortunate side-effect is the emergence of three separate sets of terminology. Banerjee (1992) uses the term *herd*; Bikchandani, Hirshleifer, and Welch (1992) use the alternative *informational cascade*. This thesis will use the two terms interchangeably.

More recently Jackson and Kalai (1997) have responded to the question of terminology: can herding models be rightly regarded as a part of game theory, since the *players* are not involved in any strategic interaction in a strict sense? Jackson and Kalai (1997) replied by defining a class of games in which herding or cascading takes place as *recurring games*. Jackson and Kalai (1997) provide formal definitions of such general games, but intuitively, a recurring game describes a repeated game where the players change at each repetition, but the actions of the previous players can be observed by later players. The examples given earlier, and the models discussed and developed in the first part of this thesis can each be thought of as recurring games in the Jackson and Kalai (1997) sense.

2.2 THE BASIC MODEL

This section outlines the basic binary model put forward by Banerjee (1992) and, in particular, Bikchandani, Hirshleifer, and Welch (1992). We begin with a situation in which agents receive a payoff of $V = 1$ from adopting a good quality technology, but

receive $V = 0$ from a low quality technology. There is an assumed cost of investment, $C = 0.5$, which makes the decision non-trivial. Initially agents must decide in a strict sequence in an exogenously determined order which is common knowledge.

The model is based around the assumption that agents receive a private signal, and observe the *actions* of those earlier in the sequence, but cannot observe the *signals* of their predecessors. Earlier actions are common knowledge to all individuals who come later. The main difference between the binary model examined here and a more general model concerns the nature of the private signals. In the binary model there are only two possible signals, that quality is high or low, so agent i 's signal $X_i \in \{H, L\}$. In the general model, examined later in this section, this is relaxed and a range of signals is considered. Each model implicitly assumes finite support for private signals and a specific belief process which involves movements in 'large' jumps; both of these assumptions are relaxed in Smith and Sorensen (1996) where a *smooth*, more general belief process, is examined.

2.2.1 The Model with Binary Signals

Consider an exogenously determined strict sequence of N individuals. Each individual is faced with the non-trivial problem of deciding whether to accept some behaviour, adopt a technology, or follow a fad. The cost of acceptance is set to $C = 0.5$ for all individuals. The gain, V , is the same for all with respective probabilities:

$$V = \begin{cases} 1 & \text{probability } 0.5 \\ 0 & \text{probability } 0.5 \end{cases} \quad (2.1)$$

Individuals differ in only two ways: first by their position in the sequence; second by the signal, X_i , they receive concerning the quality of the behaviour or technology. In this section the signal is assumed to take one of only two possible values H or L , so $X_i \in \{H, L\}$. H suggests $V = 1$, and L suggests $V = 0$. The signals are influenced by the true nature of the technology, and are hence informative, so if $V = 1$ then the signal

probabilities are:

$$X_i = \begin{cases} H & \text{probability } p_i \\ L & \text{probability } 1 - p_i \end{cases} \quad (2.2)$$

If $V = 0$ then the signal probabilities are:

$$X_i = \begin{cases} H & \text{probability } 1 - p_i \\ L & \text{probability } p_i \end{cases} \quad (2.3)$$

It is further assumed that $p_i > 0.5$ and $p_i = p$ for all i (i.e. identically distributed signals). The quality of agents' signals is the same. Denote σ as the posterior probability that $V = 1$. The expected value of accepting the technology is: $E[V] = (\sigma)1 + (1 - \sigma)0 = \sigma$. The decision-making process is closed with the following assumption.

Assumption 1. *An individual, indifferent between acceptance and rejection, i.e. for whom $E[V] = C = 0.5$, adopts the technology with probability 0.5.*

Banerjee (1992) assumes that the individual will give precedence to his own signal if he finds himself indifferent; this assumption is designed to minimize the possibility of herding. More generally Banerjee (1992) favours tie-breaking rules designed to reduce the probability of herding. His approach is to show that even in the face of such assumptions herding still occurs. However, the indifference condition considered here has the advantage of allowing herd probabilities to be developed more objectively, giving no added weight to, or against, the prospect of a herd. The rule might also be considered to be more natural, though Banerjee's approach is useful since it shows that herding is robust to the tie-breaking rule adopted by Bikchandani, Hirshleifer, and Welch (1992).

Consider the possible chain of events. The *first agent* will adopt if his signal is H and reject if his signal is L . The *second agent* can infer the signal of the first agent from his action. He will then adopt if his signal is H having observed adoption by the first

agent. If he observed rejection but received the H signal then he is indifferent and adopts with probability 0.5. If he receives the L signal and the first person rejected, then he too will reject. If the first person accepted then again he would be indifferent and so would adopt with probability 0.5. The *third agent* is the first to face the possibility of herding behaviour. If he observed an adoption and a rejection then he is in the same situation as the first individual and so his signal determines his choice. If he observed two adoptions then he will also adopt regardless of his signal since he knows that the first agent's signal is H and the second agent's signal is also H with probability > 0.5 , so the weight of evidence is in favour of adoption. This initiates an *up cascade*: the *forth agent* will also adopt as will the fifth, etc. Similarly if the third agent observes that both previous choices were rejections then he too will reject, initiating a *down cascade*. We can therefore define an *informational cascade* in the context of this model.

Definition 1. *Informational Cascade (Bikchandani, Hirshleifer, and Welch, 1992). An informational cascade occurs if an individual's action does not depend upon his private information signal. An individual, having observed the actions of those ahead of him in a sequence, who follows the behaviour of the preceding individual, without regard to his own information, is said to be in a cascade.*

The process is described in terms of a stylized probability tree for the case when $V = 1$, in figure 1. A symmetric diagram for the case when $V = 0$ would have p replaced with $1 - p$, and $1 - p$ replaced with p throughout. Examining figure 1 we find two routes to an up cascade, the first route occurring if both two agents receive H signals; this occurs with probability p^2 . The second route requires the first agent to have received an H signal and the second an L signal, but then chooses to adopt via the tie-breaking rule; this occurs with probability $0.5p(1 - p)$. Finally, note that this is all conditional on $V = 1$, which occurs with prior probability 0.5. So we have the conditional probability:

$$\Pr [\text{up cascade} \mid V = 1] = p^2 + 0.5 [p(1 - p)] = 0.5(p + p^2) \quad (2.4)$$

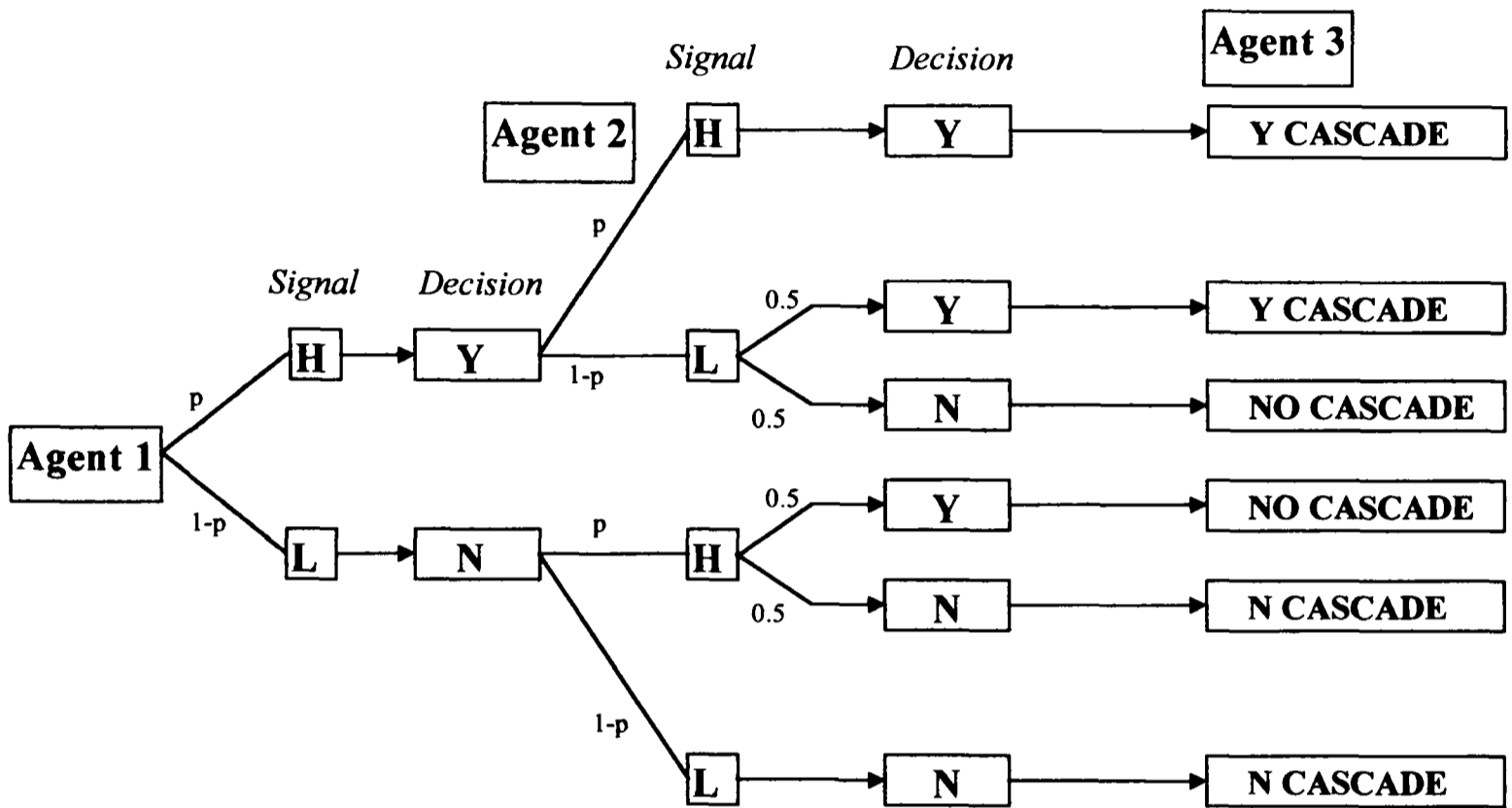


FIGURE 1: A Schematic Illustration of a Cascade, $V=1$

Considering the case when $V = 0$ we again have two routes for an up cascade, with respective probabilities $(1 - p)^2$ and $0.5(1 - p)p$, resulting in the conditional probability:

$$\Pr [\text{up cascade} \mid V = 0] = (1 - p)^2 + 0.5 [(1 - p)p] \quad (2.5)$$

So we have the unconditional probability of an up cascade as:

$$\begin{aligned} \Pr [\text{up cascade}] &= 0.5 [p^2 + 0.5p(1 - p)] + 0.5 [(1 - p)^2 + 0.5p(1 - p)] \\ &= 0.5 [1 - p + p^2] \end{aligned} \quad (2.6)$$

By symmetry, the case for a down cascade yields the same unconditional probability, so:

$$\Pr [\text{down cascade}] = 0.5 [1 - p + p^2] \quad (2.7)$$

Similarly, from observing figure 1, there are two routes which correspond to no cascade, with respective probabilities $0.5p(1 - p)$ and $0.5(1 - p)p$, which yields the unconditional probability:

$$\Pr [\text{no cascade}] = p - p^2 \quad (2.8)$$

Generalizing to an even number of n individuals, we have:

$$\Pr [\text{up cascade}] = 0.5 [1 - (p - p^2)^{n/2}] \quad (2.9)$$

$$\Pr [\text{down cascade}] = 0.5 [1 - (p - p^2)^{n/2}] \quad (2.10)$$

$$\Pr [\text{no cascade}] = (p - p^2)^{n/2} \quad (2.11)$$

The probability of being in an up cascade after four individuals is the probability of being in an up cascade after two individuals, plus the probability of being in no cascade after two individuals multiplied by the probability of being in an up cascade after another two individuals. Extended to n even individuals the chance of being in an up cascade occurs at every even person, so in general we need to check the probability of being in an up cascade after 2, 3, 4, ..., n individuals and sum these. As $p \rightarrow 0.5$ it is likely to take longer for a cascade to start. So the more informative are signals the sooner are cascades likely to start. Furthermore, the probability of not being in a cascade falls exponentially in n . Note that when $p = 0.5 + \varepsilon$, with ε arbitrarily small, and, $n = 10$, then $\Pr[\text{no cascade}] < 0.1$. It is already becoming clear that cascades are likely to occur given a large enough n .

It is also possible to find the probabilities of ending up in the correct cascade, i.e. the up cascade if $V = 1$ and the down cascade if $V = 0$. These are simply conditional probabilities and can be visualized again with reference to figure 1. Consider the true value to be $V = 1$, then after two individuals:

$$\Pr [\text{up cascade} \mid V = 1] = 0.5(p + p^2) \quad (2.12)$$

$$\Pr [\text{down cascade} \mid V = 1] = 0.5(p - 1)(p - 2) \quad (2.13)$$

$$\Pr [\text{no cascade} \mid V = 1] = p - p^2 \quad (2.14)$$

Generalizing equations 2.12 to 2.14 to the case after an even number of individuals, n :

$$\Pr [\text{up cascade} \mid V = 1] = \frac{(p^2 + p) [1 - (p - p^2)^{n/2}]}{2(1 - p + p^2)} \quad (2.15)$$

$$\Pr [\text{down cascade} \mid V = 1] = (p - p^2)^{n/2} \quad (2.16)$$

$$\Pr [\text{no cascade} \mid V = 1] = \frac{(p - 2)(p - 1) [1 - (p - p^2)^{n/2}]}{2(1 - p + p^2)} \quad (2.17)$$

From equation 2.15, the probability of being in the correct cascade is increasing in p and n . From equation 2.16 even if p is above 0.5, so signals are very informative, the probability of the wrong cascade being initiated is high.

Analysing the results so far, a number of points stand out. Firstly, there is a stark contrast with a *normal learning model* (where signals rather than actions are observed) in which we might presume that as the information in the sequence is aggregated later individuals would converge to the right action. This does not occur because it is actions rather than private signals which are revealed over time, and once a cascade is initiated, actions no longer convey any useful information. Later individuals' potentially useful

information is lost and with it the prospect of convergence to the right action. Secondly, the probability of a cascade seems to be very high, particularly for a high n and p . The probability of initiating the wrong cascade is also high, especially where p is far from 0.5. This produces an interesting paradox: *as signals become more informative the chance of a possibly incorrect cascade being initiated increases*. Thirdly, in this model informational cascades are irreversible, and so once initiated last forever, and is the direct consequence of the observability of actions rather than signals. If the signals of predecessors were observable then a long enough series of opposing signals would eventually cause people to switch actions. Finally, the failure to converge to the correct outcome might be phrased in terms of a discernible negative externality.

Definition 2. *Herd Externality (Banerjee, 1992).* The herd externality is the loss of the information contained in later agents' private signals that comes about when agents in a sequence ignore their own private information and join a herd.

2.2.2 Generalizing the Nature of Signals

Now we follow Bikhchandani *et al* in generalizing the binary nature of signals, again considering a sequence of individuals, $i = 1, 2, \dots, n, \dots$, each deciding whether to adopt some behaviour or technology. As before, agents can observe the actions of their predecessors in the sequence, but not their signals. Again we limit the ordering to be exogenous and common knowledge. The gain function, V , has a finite set of possible values, $v_1, v_2, \dots, v_l, \dots, v_s$, where v_1 and v_s bound the function, and $v_1 < v_2 < \dots < v_l < \dots < v_s$. There is again a non-trivial cost of adoption, C , i.e. $v_1 < C < v_s$. Denote the prior probability $\mu_l \equiv \Pr[V = v_l]$. Bikhchandani *et al* then call upon the concept of *perfect Bayesian equilibrium*. They argue that when an individual decides upon his action he need not consider the actions of later individuals in the sequence, and so there is no incentive to make an out-of-equilibrium move to try to influence later behaviour. Having noted this we can assume without loss of generality that any observed deviation out-of-equilibrium will not alter the beliefs of later individuals who will have the same beliefs as if the deviating individual had chosen his equilibrium action.

As in section 2.2.1 individuals act in sequence, but have two distinguishing characteristics: their place in the exogenously determined sequence; and the private signals they observe. An individual i observes one of a conditionally independent and identically distributed (henceforth IID) sequence of signals, X_i , with possible values $x_1 < x_2 < \dots < x_q < \dots < x_r$. Let p_{ql} be the probability that an individual observes signal value x_q given a true value of adoption v_l . Furthermore, it is assumed that $p_{ql} \equiv \Pr [x_q | v_l] > 0$ for all q and l . Let P_{ql} be the cumulative distribution function of X_i given $V = v_l$, that is:

$$P_{ql} \equiv \Pr [X_i \leq x_q | V = v_l] = \sum_{j=1}^q p_{j l}$$

Let J_i be the set of signal realizations that lead individual i to adopt the behaviour or technology considered. The action of this individual reveals to those after him in the sequence whether or not he observed a signal in the set J_i or its complement, J_i^c . An individual's action conveys no information about his realization if $J_i = \{x_1, x_2, \dots, x_r\}$ or if $J_i = \emptyset$, i.e. if J_i is complete or empty. At this point definition 1 can be shown to imply that cascades have infinite duration.

Proposition 1. *A cascade once started will last forever.*

Proof. There exist many proofs of this in the literature. The following is therefore short and intuitive. If an agent i is in a cascade, then by definition, regardless of his signal he will follow his predecessor's action. Therefore his action conveys no information and agent $i + 1$ can only draw the same inference from all previous actions as agent i . Agent $i + 1$'s information set is therefore made up of exactly the same public information plus one private observation. Since agent $i + 1$'s signal is drawn from the same distribution as agent i and since agent i 's action was not dependent on his signal; neither will the action of agent $i + 1$ be dependent on his own private signal. By induction, as all agents after agent $i + 1$ also have draws from the same distribution they can only draw the same

inference from all previous actions as agent i . Therefore all agents after i will simply repeat the action of agent $i - 1$. \square

Let $a_i \in \{\text{adopt}, \text{reject}\}$ be individual i 's action, and let $A_i = (a_1, a_2, \dots, a_i)$ represent the history of actions taken by individuals $1, 2, \dots, i$. Given history A_{i-1} , let $J_i(A_{i-1}, a_i)$ be the set of signal realizations that lead individual i to choose action a_i . The individual $n + 1$'s conditional expectation of V given his own signal realization, x_q , and the history, A_n , is:

$$V_{n+1}(x_q; A_n) \equiv E[V \mid X_{n+1} = x_q, X_i \in J_i(A_{i-1}, a_i), \text{ for all } i \leq n]$$

At this stage Bikhchandani *et al* introduce the tie-breaking assumption that indifferent individuals choose to adopt, rather than randomize as in section 2.2.1, which they justify with reference to the fact that they are no longer restricting themselves to a symmetric example, i.e. individual $n + 1$ adopts if $V_{n+1}(x_q; A_n) \geq C$. Therefore, the inference drawn from individual $n + 1$'s action a_{n+1} is that:

$$X_{n+1} \in J_{n+1}(A_n, a_{n+1})$$

$$\text{where } J_{n+1}(A_n, \text{adopt}) = \{x_q \text{ such that } V_{n+1}(x_q; A_n) \geq C\}$$

$$\text{and } J_{n+1}(A_n, \text{reject}) = \{x_q \text{ such that } V_{n+1}(x_q; A_n) < C\}$$

Bikhchandani *et al* impose two regularity conditions upon the model. The first, assumption 2, ensures that if an individual observes a higher signal realization, he infers

that the value of adoption is higher.

Assumption 2. Monotone Likelihood Ratio Ordering. Conditional distributions $\Pr[X_i | V = v_l]$ are ordered by the monotone likelihood ratio property, i.e. for all $l < S$: $\frac{p_{q,l-1}}{p_{q+1,l}} \geq \frac{p_{q,l}}{p_{q+1,l+1}}$ for all $q < R$ with strict inequality for at least one q .

Assumption 3. No Ties. There are no long-run ties, i.e. $v_l \neq C$ for all l .

Assumption 2 ensures that the conditional expectation of each individual increases in his signal realization. Thus, if individual i is not in a cascade and he adopts, later individuals conclude that $X_i \geq x_q$ for some q . If individual i does not adopt then the conclusion is that $X_i < x_q$.³ Assumption 3 ensures that if individuals learn enough about value by observing those before them in the sequence, they will not be indifferent between adoption and rejection.

Proposition 2. (Bikchandani, Hirshleifer, and Welch, 1992). If assumptions 2 and 3 hold then, as the number of individuals increase, the probability that a cascade eventually starts approaches one, i.e. $\lim_{n \rightarrow \infty} \Pr[\text{cascade begins}] = 1$.

Proof. Suppose that individual n , late in the sequence, is still making a decision based upon his own information, i.e. we are not yet in an informational cascade. The decisions of the earlier individuals, A_{n-1} , convey some information about their signals. Now if individual n is far enough down the line, then, by the *strong law of large numbers*, with probability close to one, he can infer the true value of adoption with almost perfect certainty. But then his own signal, x_n , contributes arbitrarily little to his information set, J_n , and he acts according to the information conveyed by the actions of previous individuals, A_{n-1} . Therefore he ignores his own private information, x_n , and initiates an informational cascade. \square

As in the binary signals model in section 2.2.1 informational cascades can often be wrong in the sense that they result in herding into the lower quality action. A very

³For more on the general importance of assumption 2 see Milgrom (1981).

noisy signal can induce a probability of an incorrect informational cascade of close to 0.5. Welch (1992) has shown that with binary signals and a uniform prior on the true value, cascades will start and can often be wrong. Banerjee assuming a continuous uniform prior distribution on the correct action, also finds that a herd will start and this could easily be on the incorrect action; however, Bikhchandani *et al* note that incorrect cascades in Banerjee's model derive from a degenerate payoff function.

2.3 THE CONTINUITY OF ACTION SPACES

In this section the main focus will be on the assumption of a continuous action space, first examined by Lee (1993). Later in the chapter we will go on to examine the inclusion of a stochastic environment by Moscarini, Ottaviani, and Smith (1996). Other major points have been made by Vives (1991) and Vives (1996) on the use of prices to overcome the informational inefficiency in information cascade models, and on the speed of learning, and by Smith and Sorensen (1996), Smith and Sorensen (1997a) and Smith and Sorensen (1997b) on generalizing the model to incorporate more general belief processes, and considering the situation when only part of past history can be observed. The continuity of the action space and stochastic environments will be examined in this and the following sections since it is these issues which are most directly relevant for the following chapters of this thesis.

Lee (1993) bases his model on Bikhchandani, Hirshleifer, and Welch (1992), using the binary signal assumption as in section 2.2.1. The state of nature is drawn randomly at the beginning from a finite number of feasible states, $s = 1, \dots, S$, and does not change. Then agents are allowed to take their actions in an exogenously determined sequence. As before there is a public body of information based upon an initial prior plus the history of previous agents' actions; this is supplemented by the private information of the agent which is a random variable, conditionally IID given the state. As before we are concerned only with rational learning, so we require that our agent update his prior distribution after observing his private information and the history of others' actions, via

Bayes' rule, and having done this choose optimally with respect to the updated posterior distribution.⁴

2.3.1 The Model

Signals are binary, $x \in X = \{1, 0\}$, and states are distinguished only by the probability of signal x , so $p_{1s} = \Pr[x = 1 | s] = p_s$ and $p_{0s} = \Pr[x = 0 | s] = 1 - p_s$. The signal carries information about the state, s , because p_s differs for different states. The probability of the signal $x = 1$ given state s also denotes the mean of the signal x given state s since $E[x | s] = 1p_s + 0(1 - p_s) = p_s$. Agent n takes an action $a \in A \subset \mathbb{R}$. The feasible action set A is the same for all agents and can be finite, countably infinite or uncountably infinite, but is assumed to be compact to guarantee the existence of an optimal action. All agents minimize the same loss function, $l_s(a)$ which depends on the state and the action taken. The information set of agent n includes the history of actions, $h^n = (a^1, a^2, \dots, a^{n-1})$ where $h^1 = \emptyset$, and the private signal, x^n , where superscript n denotes agent n . Write the prior distribution given history h^n before the signal as $\mu_s^n \equiv \mu(s | h^n)$ and the posterior distribution given history h^n and the private signal x^n as $\pi_s^n \equiv \pi(s | h^n, x^n)$ where subscript s denotes state s . Write μ^n and π^n to denote $(\mu_1^n, \mu_2^n, \dots, \mu_s^n)$ and $(\pi_1^n, \pi_2^n, \dots, \pi_s^n)$ respectively. The structure of the model including the probability of the signal values given each state and the initial prior distribution over states, μ^1 , is common knowledge. Agent n solves the problem:

$$\min_{a \in A} E[l_s(a) | h^n, x^n] \text{ or equivalently } \min_{a \in A} \sum_{s=1}^S \pi_s^n l_s(a) \quad (2.18)$$

At this point Lee makes a number of assumptions.

Assumption 4. *The loss function is written as $l_s(a) = \{a - E[x | s]\}^2 = (a - p_s)^2$*

⁴Or alternatively we are looking for a *Bayesian Nash equilibrium* choosing optimally after updating via Bayes' rule.

Assumption 5. *The probability of the signal, $x = 1$, is strictly between 0 and 1, and increases with the state s , i.e. $0 < p_1 < p_2 < \dots < p_s < 1$*

Assumption 6. *The initial prior distribution is non-degenerate, i.e. $0 \leq \mu_1^1, \dots, \mu_s^1 < 1$*

Assumption 7. *The action set A contains the means of signal for all states, i.e. $\{p_1, p_2, \dots, p_s\} \subset A$*

Assumption 4 is merely a simplification and applies without loss of generality if the model is extended for well behaved concave utility functions, for example assuming that agents maximize a monotone transform of the loss function given in assumption 4. The first part of assumption 5 and assumption 6 rule out degenerate cases. The second part of assumption 5 implies that the signal, $x = 1$, is observed more often from a higher state than from a lower state.⁵

Assumption 4 and assumption 6 together imply that if an agent knows the true state his action will equal the true state's mean value. Lee re-defines an *informational cascade*, and another concept, a *fully revealing informational cascade*, in terms which better fit his model.

Definition 3. *Informational Cascade (Lee, 1993). An informational cascade arises if:*

$$\lim_{n \rightarrow \infty} \arg \min_a E [l_s(a) | h^n, x^n] = \bar{a} \in A \quad (2.19)$$

Definition 4. *(Lee, 1993). A fully revealing informational cascade arises if:*

$$\lim_{n \rightarrow \infty} \arg \min_a E [l_s(a) | h^n, x^n] = \hat{a}_{\tilde{s}} = \arg \min_a l_{\tilde{s}}(a) \quad (2.20)$$

where \tilde{s} is the true state.

⁵This ensures that the signal is *informative* which can be seen as the binary signal variant of assumption 2.

A non-fully revealing informational cascade arises if definition 4 is not satisfied. In Lee's definition an informational cascade comes about when agents have posterior distributions which are close together and this leads to similar actions being taken. A fully revealing informational cascade occurs when there is no asymmetry of information: the true state is revealed by the action choices. The Bikhchandani *et al* definition depends upon whether the action choice reveals the private signal. Lee, in the proposition below shows that definition 1 requires a uniform action for an informational cascade to arise. Lee's definition does not require this, and therefore encompasses the Bikhchandani *et al* definition. If an informational cascade in the sense of Bikhchandani *et al* arises this will imply that one in the sense of Lee has also arisen. Definition 3 is therefore a more general definition.

2.3.2 Results

Lee sets up and proves a number of propositions and lemmas which culminate in establishing two major theorems.

Proposition 3. (Lee, 1993). *If there exists an N and h^N such that for all $x^N \in X$, $\hat{a}^N = \arg \min_a E [l_s(a) | h^N, x^N] = \bar{a} \in A$, then $\hat{a}^n = \arg \min_a E [l_s(a) | h^n, x^n] = \bar{a}$ for all $n \geq N$*

Proof. The proof is given in Lee (1993). □

An informational cascade in the sense of definition 1 has a positive probability of being non-fully revealing, as established by the theorem below. Comparing the two sets of definitions, the claim in Bikhchandani, Hirshleifer, and Welch (1992) that, for an action set with two elements, an informational cascade in the sense of definition 1 arises with probability 1, is equivalent to the claim that for such an action set there is a positive probability of the occurrence of a non-fully revealing informational cascade in the sense of definition 4. It is shown below that any discrete action set allows a positive probability of the occurrence of a non-fully revealing informational cascade and this result generalizes to settings with more than two actions without change.

Theorem 1. (Lee, 1993). *It is the case that:*

$$\Pr \left\{ \left| \lim_{n \rightarrow \infty} \arg \min_a E [l_s(a) | h^n, x^n] - \hat{a}_s \right| > 0 \right\} > 0$$

if and only if there exists an N and h^N such that for all $x^N \in X$,

$$\hat{a}^N = \arg \min_a E [l_s(a) | h^N, x^N] = \bar{a} \in A$$

Proof. The proof is given in Lee (1993). □

Intuitively, theorem 1 states that whenever learning from observed action choices stops in finite steps there is a positive probability that all the agents from then on are taking an action which is not optimal under the true state and as a result the true state is never fully revealed. Lee then goes on to consider the necessary and sufficient condition to guarantee a fully revealing information cascade for any initial prior probability, μ . Firstly, lemma 1 established below implies that the optimal action minimizes the distance to the conditional expectation.

Lemma 1. *Optimal Action Choice (Lee, 1993). Given a posterior distribution $\pi = (\pi_1, \pi_2, \dots, \pi_s)$ the optimal action is $\hat{a} = \arg \min_{a \in A} |a - E_\pi p|$*

Proof. The proof is given in Lee (1993). □

The lemma below establishes that the conditional expectation given a higher signal is higher than the conditional expectation when a low signal is observed.

Lemma 2. (Lee, 1993). *$E[p | \mu, x = 1] \geq E[p | \mu, x = 0]$ for all prior μ*

Proof. The proof is given in Lee (1993). □

The theorem below provides the necessary and sufficient conditions for the occurrence of a fully revealing informational cascade with probability 1 given any priors. The theorem is stated as an inequality condition to be satisfied by each point in the action set.

Each point v in the action set has gaps to its right, radius ε_v^r , and its left, radius ε_v^l . An interior point would have $\varepsilon_v^r = \varepsilon_v^l = 0$.

Theorem 2. (Lee, 1993). *A fully revealing informational cascade arises with probability 1 for all prior μ , if and only if for all v in the action set A such that $v \in [p_{\tilde{s}}, p_{\tilde{s}+1}]$ for some \tilde{s} , either $\varepsilon_v^r = \varepsilon_v^l = 0$, or at least one of the two gaps of v is strictly positive for which*

$$\frac{p_{\tilde{s}} (p_{\tilde{s}} - v - \varepsilon_v^r)}{(1 - p_{\tilde{s}}) (p_{\tilde{s}} - v + \varepsilon_v^l)} \leq \frac{p_{\tilde{s}+1} (p_{\tilde{s}+1} - v - \varepsilon_v^r)}{(1 - p_{\tilde{s}+1}) (p_{\tilde{s}+1} - v + \varepsilon_v^l)}$$

Proof. The proof is given in Lee (1993). □

Lee gives an example of the use of this theorem. Firstly, define the criterion function to be:

$$f(p) = \frac{p(p - v - \varepsilon_v^r)}{(1 - p)(p - v + \varepsilon_v^l)} \quad (2.21)$$

Then consider the parameter values $v = 0.5$, $\varepsilon_v^r = \varepsilon_v^l = 0.01$, $p_1 = 0.2$ and $p_2 = 0.8$. The condition of theorem 2 is satisfied since $f(p_1) \leq f(p_2)$. Therefore, when the true states are $p = 0.2$ or $p = 0.8$, the action set which has an isolated point at $v = 0.5$ with gaps of radius 0.01 to the left and right does not allow a non-fully revealing information cascade. There are two corollaries which follow from theorem 2.

Corollary 1. (Lee, 1993). *A fully revealing informational cascade arises with probability 1 for all initial priors μ , if the action set A contains the interval $[p_1, p_s] : [p, p_s] \subset A$*

Corollary 2. (Lee, 1993). *If the action set does not contain an open interval between p_1 and p_s such that for some s , p_s is one of its end points, there are initial priors μ which generate a non-fully revealing informational cascade.*

Corollary 1 is true since it relates to the case when $\varepsilon_v^r = \varepsilon_v^l = 0$ for all v in the action set. Corollary 2 implies that a discrete action set always allows a positive probability of the occurrence of a non-fully revealing informational cascade. However, an action set with gaps is only a necessary condition for the occurrence of a non-fully revealing informational cascade, not a sufficient condition.⁶

In conclusion, Lee (1993) characterizes the action set which guarantees a fully revealing informational cascade almost everywhere. In particular he shows that a discrete action set may allow non-fully revealing informational cascades, but this is not certain for all priors, and furthermore, there are many action sets which would ensure that a fully revealing informational cascade arises with probability 1. Therefore, the results of Banerjee (1992) and Bikchandani, Hirshleifer, and Welch (1992) are perhaps misleading in that they both use make use of action sets which produce non-fully revealing informational cascades with probability 1. Lee (1993) challenges the generality of these results and suggests that care be taken when choosing the form of the action set allowed since this will greatly affect the evolution of decision-making in a sequential model.

2.4 STOCHASTIC ENVIRONMENTS

The working paper by Moscarini, Ottaviani, and Smith (1996) demonstrates that if a stochastic environment is added to a sequential decision-making model of the type developed by Banerjee (1992) and Bikchandani, Hirshleifer, and Welch (1992) then only temporary informational cascades can arise. Furthermore, if the environment changes in a sufficiently unpredictable way then no informational cascade will arise, because past information will depreciate so quickly that beliefs can never be too extreme.

⁶It is important to note that corollary 1 is not taken to imply that in all cases a continuous action set rules out informational cascades in the sense of definition 1. Huck and Oechssler (1997) have shown that, while it is often claimed that with continuous action spaces informational cascades are impossible, stylized models can be designed in which the action space need not be coarser than the signal space, but where agents ignore their signals and enter informational cascades in the sense of definition 1. Banerjee (1992) did allow a continuous action space and still obtained herding behaviour. However, he did so using degenerate payoffs (where payoffs are zero unless the optimal action is chosen exactly) while Huck and Oechssler allow for non-degenerate payoffs.

2.4.1 The Model

Moscarini, Ottaviani, and Smith (1996) consider a countable number of agents acting sequentially who take one of two possible actions, $a \in \{a_0, a_1\}$. Payoffs to actions are contingent on an unknown state of the world, $\omega \in \{\omega_0, \omega_1\}$. The common prior belief that the state is initially ω_1 is given as q^1 . Action a_1 is more rewarding than action a_0 in state ω_1 , while the opposite is true in state ω_0 . The payoffs, π , are $\pi [a_i | \omega = \omega_i] = 1$ and $\pi [a_i | \omega = \omega_j] = 0$ where $i \neq j$ and $i, j \in \{0, 1\}$. After an agent decides upon an action, the state of the world changes with shift probability ε , assumed to be Markovian and independent of the current state. Therefore, for $i \neq j$ and $i, j \in \{0, 1\}$:

$$\Pr [\omega^n = \omega_i | \omega^{n-1} = \omega_j] = \Pr [\omega^n = \omega_j | \omega^{n-1} = \omega_i] = \varepsilon \quad (2.22)$$

Before choosing an action, individual n observes a private signal $\sigma^n \in \{\sigma_0, \sigma_1\}$ and the public history of action decisions (not signals) of all the preceding individuals $1, 2, \dots, n-1$. Private signals are drawn from a state-dependent Bernoulli distribution, and are independent conditional on the current state. For $i, j \in \{0, 1\}$:

$$\Pr [\sigma = \sigma_i | \omega = \omega_j] = \begin{cases} \alpha > 0.5 & \text{if } i = j \\ 1 - \alpha < 0.5 & \text{if } i \neq j \end{cases} \quad (2.23)$$

The quality of the private signal is assumed bounded, i.e. $\alpha < 1$. For $n \geq 2$ let $H^n \equiv \{a_0, a_1\}^{n-1}$ be the space of all possible period n histories of actions chosen by the $n-1$ predecessors of agent n . Let h^n denote an element of H^n . Let $q^n \equiv \Pr [\omega = \omega_1 | h^n]$ be the public probability belief that the state is ω_1 in period n conditional on the publicly observed history of actions chosen by the predecessors of agent n . Similarly let $r_i^n \equiv \Pr [\omega = \omega_1 | h^n, \sigma_i]$ be the posterior belief that the state is ω_1 conditional on both the public action history h^n and the realization σ_i of the private signal observed by agent n . An application of Bayes' rule yields:

$$r_i^n = \frac{\Pr[\omega_1 \cap \sigma_i | h^n]}{\Pr[\sigma_i | h^n]} = \frac{\Pr[\sigma_i | h^n, \omega_1] \Pr[\omega_1 | h^n]}{\Pr[\sigma_i | h^n]} \quad (2.24)$$

So that:

$$r_0^n = \frac{(1 - \alpha) q^n}{\alpha(1 - q^n) + (1 - \alpha) q^n} \quad (2.25)$$

$$r_1^n = \frac{\alpha q^n}{\alpha q^n + (1 - \alpha)(1 - q^n)} \quad (2.26)$$

Agent n wishes to choose the action a^n which yields the highest expected payoff. If agent n receives the private signal $\sigma^n = \sigma_1$, then it is optimal to take action $a^n = a_1$ if and only if $r_1^n \geq 0.5$. Using equation 2.26 the requirement becomes $q^n \geq 1 - \alpha$. Moscarini *et al* assume that when indifferent the agent will act so as to avoid herding. The decision rule can be summarized as:

$$\text{If } a^n = 0 \text{ then } a^n = a_0 \Leftrightarrow q^n \leq \alpha \text{ and } a^n = a_1 \Leftrightarrow q^n > \alpha \quad (2.27)$$

$$\text{If } a^n = 1 \text{ then } a^n = a_0 \Leftrightarrow q^n < 1 - \alpha \text{ and } a^n = a_1 \Leftrightarrow q^n \geq 1 - \alpha \quad (2.28)$$

2.4.2 Results

If we set $\varepsilon = 0$ then we return to a model much like that in section 2.2.1. There will be an informational cascade on action a_1 (respectively a_0) as soon as $q^k > \alpha$ (respectively $q^k < 1 - \alpha$), since at this point the action a_1 (respectively a_0) will be taken by all the

proceeding agents regardless of their private signals. With $\varepsilon > 0$ the dynamics change. The cascade region is unaffected by the state switching, since that event occurs only after the decision has been made. However, when the possibility that the state of the world has changed in the meantime is accounted for, the public prior belief of agent $n + 1$, coming after agent n who chose $a^n = a_i$ according to the signal $\sigma^n = \sigma_i$, satisfies $q^{n+1} = (1 - \varepsilon) r_i^n + \varepsilon (1 - r_i^n)$, which can be rewritten using equations 2.25 and 2.26 as:

$$q^{n+1} = \begin{cases} f_0(q^n) \equiv \frac{(1-\varepsilon)(1-\alpha)q^n + \varepsilon\alpha(1-q^n)}{(1-\alpha)q^n + \alpha(1-q^n)} & \text{if } a^n = a_0 \\ f_1(q^n) \equiv \frac{(1-\varepsilon)\alpha q^n + \varepsilon(1-\alpha)(1-q^n)}{\alpha q^n + (1-\alpha)(1-q^n)} & \text{if } a^n = a_1 \end{cases} \quad (2.29)$$

Considering the case when $q^k > \alpha$ (the other case can be treated symmetrically) the action chosen will be $a^k = a_1$, regardless of signal σ^k . The next agent $k + 1$ knows that $a^k = a_1$ is uninformative, and computes the public prior belief $q^{k+1} = (1 - \varepsilon) q^k + \varepsilon (1 - q^k)$. In general, the following agent $n + 1$, as long as $q^n > \alpha$ or $q^n < 1 - \alpha$, will update his prior belief during the cascade in the same fashion, according to the (uninformative) cascade dynamics:

$$q^{n+1} = \varphi(q^n) \equiv (1 - \varepsilon) q^n + \varepsilon (1 - q^n) \quad (2.30)$$

The public belief dynamics are stochastic and determined by expression 2.29 as long as $1 - \alpha \leq q^n \leq \alpha$ (when not in a cascade) and are deterministic and follow expression 2.30 when either $q^n > \alpha$ or $q^n < 1 - \alpha$ (during the cascade). Moscarini, Ottaviani, and Smith (1996), having established a belief dynamic, give the conditions for the occurrence of an informational cascade and show that such a cascade must eventually end.

Proposition 4. *(Moscarini, Ottaviani and Smith, 1996). For any $\varepsilon \in (0, 1)$, if a cascade exists, then for some $k = k(\varepsilon) < \infty$, the cascade must end in $k(\varepsilon)$ periods.*

Proof. The proof is given in Moscarini, Ottaviani, and Smith (1996). □

Proposition 4 shows that an informational cascade must eventually stop. It is possible to compute the maximum length of an informational cascade. The longest possible informational cascade on a_1 starts with a belief $f_1(\alpha)$, and after h periods in an informational cascade the belief is:

$$\varphi^h(f_1(\alpha)) = \varepsilon \left[\sum_{i=0}^{h-1} (1 - 2\varepsilon)^i \right] + (1 - 2\varepsilon)^h f_1(\alpha) \quad (2.31)$$

The informational cascade will terminate when:

$$\varphi^h(f_1(\alpha)) \leq \alpha \text{ or equivalently } (1 - 2\varepsilon)^{h+1} \leq [1 - 2\alpha(1 - \alpha)]$$

We can find the tight upper bound to the informational cascade's length as:

$$K(\alpha, \varepsilon) \equiv \{\log(1 - 2\varepsilon)\}^{-1} \log[1 - 2\alpha(1 - \alpha)] \quad (2.32)$$

Furthermore, the higher the quality of private information and the more predictable the evolution in the state of the world, the longer an informational cascade can possibly last, since:

$$\frac{\partial K(\alpha, \varepsilon)}{\partial \alpha} > 0 \text{ and } \text{sign} \left\{ \frac{\partial K(\alpha, \varepsilon)}{\partial \alpha} \right\} = \text{sign}(1 - 2\varepsilon) \text{ for } \varepsilon \neq 0.5$$

Proposition 5. (Moscarini, Ottaviani and Smith, 1996). For any $q^1 \in (1 - \alpha, \alpha)$, with probability 1, an informational cascade on some action arises in finite time if and only if $\varepsilon < \underline{\varepsilon}(\alpha) \equiv \alpha(1 - \alpha)$.

Proof. The proof is given in Moscarini, Ottaviani, and Smith (1996). □

Proposition 5 provides the condition that only if the state of the world is sufficiently

persistent will an informational cascade on one action arise.

Corollary 3. (*Moscarini, Ottaviani and Smith, 1996*). For $q^1 \in (1 - \alpha, \alpha)$, with probability 1, informational cascades on alternating actions arise in finite time if and only if $\varepsilon > \bar{\varepsilon}(\alpha) \equiv 1 - \alpha(1 - \alpha)$.

Proof. The proof is given in Moscarini, Ottaviani, and Smith (1996). \square

Corollary 3 describes the interesting case which can occur when the state of the world changes rapidly enough and agents alternate between the two actions. The system may enter an *alternating cascade* in which agents alternate between the two action regardless of private signals.

Corollary 4. (*Moscarini, Ottaviani and Smith, 1996*). No cascade ever arises for $\varepsilon \in [0.25, 0.75]$.

Proof. This follows immediately from proposition 5, corollary 3 and the fact that $\alpha > 0.5$. \square

Corollary 4 provides the final major result from Moscarini, Ottaviani, and Smith (1996), that an environment changing in a sufficiently unpredictable way will stop the development of an informational cascade.

2.5 CONCLUSIONS

This chapter focused on the seminal works by Banerjee (1992) and Bikchandani, Hirshleifer, and Welch (1992). The main result is perhaps the welfare cost attached to the perfectly rational use of the information that comes through the observation of others. This welfare cost is an externality as it harms those who come later in the sequence at no cost to the initiator who begins the herd towards one particular action. The first two parts of this thesis will make heavy use of the results from this chapter, and develop them further. In particular we will look at how the herd externality can be manipulated

through the provision of additional information in chapters 3 and 5. Part II will change the framework in a fairly dramatic way, but the herd externality will remain a central focus. Lee (1993) also received a good deal of space in this chapter as it represented the first major attempt to generalize herding. In particular it pointed out the basic requirements for herding to take place, and has therefore provided later researchers interested in this field with a place to start when building models designed to capture the herd externality. Moscarini, Ottaviani, and Smith (1996) show that results regarding informational cascades in a sequential model are not automatically robust to the addition of a stochastic environment. Intuitively, information depreciation increases the immediate value of new information but diminishes its long-term value. In a social learning context, since agents are myopic, only the former effect is present, and depreciation unambiguously discourages herding. Moscarini, Ottaviani, and Smith (1996) was examined not simply because the results were considered of interest, but mainly because, from the small pool of available work on sequential herding, it best represents something of a link between parts I and II of this thesis.

In part II the environment will become stochastic, although in a way not envisaged by Moscarini, Ottaviani, and Smith (1996). Agents will be able to delay their decision-making, and this will produce a stream of new information about the action (or inaction) of others. Part II is a more complex situation as together with the decision of *what* to do, agents will have to concern themselves with *when* to make a decision. The timing of decision becomes endogenized, and the results of Moscarini, Ottaviani, and Smith (1996) will not be directly applicable.

CHAPTER 3

OPTIMIZING INFORMATION IN THE HERD

3.1 INTRODUCTION

Should a firm with a new product release it to the entire market on the same day or pre-release to a select set of customers first?⁷ Should movie premieres be made available to members of the press and high profile celebrities or should films simply be released world-wide on the same day? Should a firm's product be released on the same day in several markets or released sequentially in different markets? Is it optimal for every country in the European Community to conduct separate drug trials or should they pool and have one drug trial? Is it sensible for the government to slowly release new drug treatments to doctors or should there be a core release to a select group of doctors first?

All of these questions raised in chapter 1 relate to whether a slow sequential release of information is better than a discrete simultaneous release in the first instance followed by a slow sequential release thereafter. The term *better* may relate to consumer welfare (in the case of government planners) or profitability (in the case of firms). In both cases the approach taken in this chapter is to model the learning process as a potential herding phenomenon. This allows a quick identification of the crucial trade-off involved between the gains for those late in the sequence who have access to more information if an initial group of "guinea pigs" is used, and the costs for those within the group of guinea pigs.

⁷This chapter is based on SgROI (2000a), forthcoming in *Games and Economic Behavior*, and was presented at the First World Congress of the Game Theory Society, Bilbao, July 2000 and the Gorman Workshop, Department of Economics, University of Oxford, March 2000.

More generally, forcing a sub-set of agents to make decisions quickly is important in any herd context when agents learn by observing each others' actions. Later decision-makers in such a model can gain a great deal from early movers, but they are also liable to become trapped in an information-damaging herd in which it is impossible to derive information from agents' actions, as they are simply copying their predecessors regardless of their own private information. Forced early movement overcomes this problem by providing later movers with genuinely informative actions and easily inferable private information. However, this is clearly damaging for those forced to decide early since they can no longer gain from observational learning themselves. We therefore have a clear trade-off.

This chapter begins in section 2 by developing a herding or *informational cascade* model based on the work of Bikchandani, Hirshleifer, and Welch (1992) which has been usefully categorized by Jackson and Kalai (1997) as a *recurring game*. The chapter then moves on in section 3 by developing individual payoff functions, then a total consumer welfare function. Section 4 then examines the central trade-off from a social planner's perspective, finding that the optimal size of this sub-set of guinea pigs depends on the total number of agents in the sequence and the probability that a signal points in the right direction. This part of the chapter shows that decision-making by a sequence of pairs is superior to sequential individual decision-making, and might therefore be an interesting informational justification for marriage! Section 5 considers the problem from a firm's perspective. The firm is naturally not interested in maximizing consumer welfare, but is instead interested in ensuring as many successful sales as possible. The firm can still use guinea pigs through some form of promotional campaign to raise its sales, and the chapter produces various comparative statics to show how knowledge of the number of consumers and the probability that their private information is correct is joined with the firms' knowledge of the true state to provide a clear optimal policy. A firm with a good product will make more use of guinea pigs and intervene aggressively, doing so all the more as the number of consumers in the market increases and the quality of consumers' private information falls. There are clear implications for welfare as it is in the firm's interest to assist consumers to make the right choice. With a bad product the firm is

indirectly interested in minimizing consumer welfare by convincing the maximum number to purchase when doing so is not optimal.

The findings in this chapter provide at least one way to improve on the consumer welfare damaging herds that can start in many circumstances. Using the results in this chapter it is possible to look for any potential herd situation in which a herd might move in the wrong direction and calculate an optimal number of guinea pigs, plus the percentage improvement over *laissez-faire*, that an attempt to force early movement could achieve. For example, if we were dealing with a group of 100 individuals, with private information which is correct $\frac{2}{3}$ of the time, it would be optimal to force 22 of these to move early, perhaps through the use of schemes to provide extra incentives for early movement (such as price reductions, or free gifts), limiting the number in this scheme to 22 (through randomization, or via prior selection). This would maximize overall consumer welfare and minimize the chance of an incorrect herd. For a firm we can now find some number of early and perhaps high profile consumers which would provide maximum sales. The firm could seek out members of the press or celebrities and provide them with its product at some reduced rate, and the extra sales certainly justify reducing prices to these early movers. In this way we can justify various observed practices, from a card-holder day at a department store in which a select number of customers are urged to try out new stock, to the use of movie premieres in which celebrities and members of the press gather to pass judgement on a new film before the general release of the film. Although there are undoubtedly many reasons for such promotional activity by firms, such practices increase the likelihood of products achieving success in the marketplace, and work against the herd externality.

3.2 THE MODEL

This section sets up a version of the model first used in the seminal herd paper by Bikchandani, Hirshleifer, and Welch (1992) more generally characterized by Jackson and Kalai (1997) as a recurring game. Most early herd papers were primarily concerned

with herd probabilities and proving that herds were likely or even certain to occur. The main focus of this chapter is rather with payoff functions and the development of a total consumer welfare function. However, in order to derive such functions we cannot avoid looking at herd probabilities. We can motivate this in two ways. Firstly, we will consider the role of a social planner, government or regulatory agency. The agents might simply be consumers of a product and the social planner might simply wish to maximize consumer welfare. Later we will consider a firm which has a different aim. The firm wishes to maximize profit. We proxy this aim by maximizing sales since price is not a variable we consider, so the firm will attempt to induce as many agents to purchase as possible.

3.2.1 Preliminaries

Consider a sequence of $N \in \mathbb{N}^{++}$ agents, the ordering of which is exogenous and common knowledge, each deciding whether to adopt/purchase (Y) or not (N) some product or technology. Each agent observes the actions (Y or N) of his predecessors. The cost of adoption is $C = \frac{1}{2}$, and results in the gain of V which has prior probability $\frac{1}{2}$ of returning 0 or 1. The agents each receive a conditionally independent signal about V defined as $X_i \in \{H, L\}$ for agent i . The signals are informative in the sense that:

$$\Pr[X_i = H \mid V = 1] = \Pr[X_i = L \mid V = 0] = p \in (0.5, 1)$$

$$\Pr[X_i = H \mid V = 0] = \Pr[X_i = L \mid V = 1] = 1 - p \in (0, 0.5)$$

We assume that signals are identically distributed and note that the restriction on p suffices to produce informative but non-fully revealing signals. Define the history up to agent n as the set of actions of agents 1 to $n - 1$ so $H_{n-1} \equiv \{A_1, A_2, \dots, A_{n-1}\}$ where $A_i \in \{Y, N\}$. Now define the information set of agent i as $I_i \equiv \{H_{i-1}, X_i\}$. It will be the case that in certain circumstances X_i will be inferable from A_i but this will not always be true. Now define N^{odd} as the set of agents from N indexed by only odd numbers from \mathbb{N}^{++} , and equivalently define N^{even} . Define also \mathbb{N}^{odd} as the set of odd numbers in \mathbb{N}^{++} ,

and equivalently define N^{even} . Define $E[\pi_i]$ to be agent i 's *ex ante* expected payoff (i.e. his expected payoff before his signal draw). Finally define $\#X_i$ as the number of signals or actions of type X_i drawn or taken up to and including agent i .

Now $X_1 = H \Leftrightarrow A_1 = Y$ and $X_1 = L \Leftrightarrow A_1 = N$. Agent 2 can infer agent 1's signal, X_1 , from his action, A_1 , and so has an information set $I_2 = \{X_1, X_2\}$. If $X_2 = H$ and $A_1 = Y \Rightarrow X_1 = H$ then agent 2 adopts so $A_2 = Y$. If $X_2 = H$ and $A_1 = N \Rightarrow X_1 = L$ or if $X_2 = L$ and $A_1 = Y \Rightarrow X_1 = H$ agent 2 will have two conflicting signals so we require a tie-breaking rule. We use a simple coin-flipping rule which is known to all agents:

Condition 1. (*Tie-breaking rule*) If I_i includes an equal weighting of H and L signals then $\Pr[A_i = Y] = \Pr[A_i = N] = \frac{1}{2}$. This rule is common knowledge.

3.2.2 Cascades

Consider a possible chain of events. The *first agent* will purchase if $X_1 = H$ and reject if $X_1 = L$. The *second agent* can infer the signal of the first agent from his action. He will then purchase if $X_2 = H$ having observed purchase by the first agent. If he observed rejection but received the signal $X_2 = H$ then he will flip a coin following the tie-breaking rule. If he receives $X_2 = L$ and $A_1 = N$ then he too will choose $A_2 = N$. If the first agent purchased then he would be indifferent and so flip a coin. The *third agent* is the first to face the possibility of a herd. If he observed two purchases, so $H_2 = \{Y, Y\}$ then $A_3 = Y$ for all X_3 since he knows that $X_1 = H$ and the second agent's signal is also more likely to be H than L , so the weight of evidence is in favour of purchase *regardless of* X_3 . This initiates a *Y cascade*: the *forth agent* will also adopt as will the fifth, etc. Similarly if the third agent observes that both previous choices were rejections then he too will reject, initiating a *N cascade*. This process can be described in terms of a stylized probability tree, and is done so for the case when $V = 1$, in figure 1 of chapter 2. Definition 1 in chapter 2 defines an informational cascade, but to provide a more model-specific definition:

Definition 5. Informational Cascades. A *Y cascade* is said to occur if $A_{i-1} = Y \Rightarrow A_i = Y$ for all X_i . A *N cascade* is said to occur if $A_{i-1} = N \Rightarrow A_i = N$ for all X_i .

For the sake of clarity define the *initiator* of a herd or cascade as the agent whose decision to go Y or N makes the following agent's signal irrelevant. The cascade *traps* the agent who first faces a deterministic optimal choice regardless of his signal value, and all subsequent agents. So in the case of $H_2 = \{Y, Y\}$ a Y cascade is initiated by agent 2 and agent 3 finds himself trapped in the Y cascade. Note that if $H_2 = \{Y, N\}$ or $H_2 = \{N, Y\}$ then agent 3 will be in the same position, pre-signal draw, as agent 1. Note also that if agent 3 finds himself trapped in a cascade so to will agents 4, 5, 6, ..., N .

Proposition 6. *A cascade once started will last forever.*

Proof. This is proposition 1 of chapter 2, and is proved there. \square

Consequently, a cascade once started will last forever, even if it is based on an action which would not be chosen if all the agents' signals were common knowledge. Finally, we might consider the possibility of convergence to the incorrect outcome, in terms of a discernible negative externality as in definition 2 of chapter 2. A social planner will wish to minimize the impact of this negative externality on consumer welfare. In some cases it will also be in the interests of a firm to work against this externality. As we see later, in some cases a firm will actually use this externality to its own advantage, when it wishes to sell a low quality product.

3.2.3 Calculating Herd Probabilities

From the model specifications we can derive the unconditional *ex ante* probabilities of a Y cascade, N cascade, or no cascade after n agents. Define $Y(n)$ to be a Y cascade initiated by agent n and similarly define $N(n)$ for a N cascade and $No(n)$ for no cascade by agent n . For example $\Pr[Y(2)]$ is simply the probability that the first two agents both choose Y . So $1 - \Pr[Y(n)] - \Pr[N(n)] = \Pr[No(n)]$ for all n . Starting with 2 agents we have $\Pr[Y(2) | V = 1] = p^2 + \frac{(1-p)p}{2}$ & $\Pr[Y(2) | V = 0] = (1-p)^2 + \frac{p(1-p)}{2}$. Therefore $\Pr[Y(2)] = \frac{1-p+p^2}{2}$. Similarly, we have $\Pr[N(2)] = \frac{1}{2}(1-p+p^2)$. No cascade by agent 2 will occur with probability $1 - \Pr[Y(n)] - \Pr[N(n)]$, therefore $\Pr[No(2)] = p - p^2$. Note of course that this can be alternatively calculated as the occurrence of HL or

LH and a coin flip by agent 2, so $\Pr[No(n)] = \frac{1}{2}p(1-p) + \frac{1}{2}(1-p)p$. Further note that $\Pr[Y(2)]$ and $\Pr[N(2)]$ are not conditional on V since they are fully symmetric so $\Pr[N(n)] = \frac{1}{2}(1 - \Pr[No(n)])$. Now note that $\Pr[Y(4)] = \Pr[Y(2)] + \Pr[No(2)]\Pr[Y(2)]$ and similarly for $\Pr[N(4)]$. Further $\Pr[No(4)] = (\Pr[No(2)])^2$. Using this we can easily deduce the general probabilities after an even number of n agents to be $\Pr[No(n)] = (\Pr[No(2)])^n = (p - p^2)^{\frac{n}{2}}$ for no cascade, and $\Pr[Y(n)] = \Pr[N(n)] = \frac{1}{2}\{1 - \Pr[No(n)]\} = \frac{1}{2}\left[1 - (p - p^2)^{\frac{n}{2}}\right]$ for a Y or N cascade. Now note that as $p \rightarrow 1$ cascades tend to start sooner, so more precise signals raise the probability of histories that lead to the correct cascades where we define correct cascades as a Y cascade if $V = 1$ or a N cascade if $V = 0$. As Bikhchandani *et al* note, the probability of not being in a cascade falls exponentially with the number of agents, for example for a very noisy signal, $p = \frac{1}{2} + \varepsilon$ with $\varepsilon \rightarrow 0$ we have $\Pr[No(10)] < 0.1$. Now we conclude this part of the analysis by considering the probability of the correct or incorrect cascade occurring:

$$\Pr[Y(2) | V = 1] = p^2 + \frac{1}{2}p(1-p) = \frac{1}{2}p(p+1)$$

$$\Pr[No(2) | V = 1] = \frac{1}{2}p(1-p) + \frac{1}{2}p(1-p) = p(1-p)$$

$$\Pr[N(2) | V = 1] = (1-p)^2 + \frac{1}{2}p(1-p) = \frac{1}{2}(p-2)(p-1)$$

After an even number of n agents we have:

$$\Pr[No(n) | V = 1] = (\Pr[No(2) | V = 1])^{\frac{n}{2}} = (p - p^2)^{\frac{n}{2}}$$

$$\Pr[Y(n) | V = 1] = \Pr[Y(2) | V = 1] + \Pr[Y(2) | V = 1]\Pr[No(2) | V = 1]$$

$$\begin{aligned}
& + \Pr [Y (2) | V = 1] \Pr [No (4) | V = 1] + \dots + \Pr [Y (2) | V = 1] \Pr [No (\frac{n}{2}) | V = 1] \\
& = \Pr [Y (2) | V = 1] \left[1 + (p - p^2) + \dots + (p - p^2)^{\frac{n}{2}} \right]
\end{aligned}$$

Now using the sum of a geometric series we have:

$$\Pr [Y (n) | V = 1] = \frac{p(p+1)}{2} \frac{1 - (p-p^2)^{\frac{n}{2}}}{1 - (p-p^2)} \quad (3.1)$$

Similarly we can calculate for $\Pr [N (n) | V = 1]$:

$$\Pr [N (n) | V = 1] = \frac{(p-2)(p-1)}{2} \frac{1 - (p-p^2)^{\frac{n}{2}}}{1 - (p-p^2)} \quad (3.2)$$

Note that from equation 3.1 $\Pr [Y (n) | V = 1]$ is increasing in p and n but from equation 3.2 we have that $\Pr [N (n) | V = 1]$ is high even for p much higher than $\frac{1}{2}$. Therefore, even when a great majority of the signals are of type H , a product still faces the prospect of a possible herd against its purchase. This is worrying for both a social planner and for a firm with a high quality product. The symmetric case where $V = 0$ would apply when the product is of low quality, and the results provide some hope for the manufacturer of such a product, since there is always the chance of a Y cascade. As we will see, firms and planners can manipulate the probability of a Y cascade, so they need not remain passive in the face of a potential herd.

3.3 PAYOFFS AND CONSUMER WELFARE

In this section we move away from standard herding concerns and instead focus on the calculation of individual payoffs for agents along the sequence. These calculations are then used to derive an expression for total consumer welfare. Evaluating the payoffs to the potential consumers and the total consumer welfare function, represents the first

step in finding a consumer welfare-improving policy for a social planner. We will look at a more active role for the social planner, in which it can improve on the *laissez-faire* outcome in the next section. As we see later, consumer welfare also indirectly plays an important role in finding the optimal policy for a firm.

3.3.1 Individual Payoffs

We will begin by looking at the *ex ante* expected payoff of the first agent. Note the two conditions prior probabilities are $\Pr[V = 1] = \Pr[V = 0] = \frac{1}{2}$ and signal probabilities are $\Pr[X_i = H | V = 1] = \Pr[X_i = L | V = 0] = p$. So we have:

$$E[\pi_1] = \frac{E[\pi_1|V=0]}{2} + \frac{E[\pi_1|V=1]}{2} = \frac{2p-1}{4} > 0 \quad (3.3)$$

Note also that $A_i = N \Rightarrow \pi_i = 0$ so we need only consider $A_i = Y$ when calculating payoffs. Since $p > \frac{1}{2}$ (signals are informative) we have that $E[\pi_1] > 0$. Now consider the second agent:

$$E[\pi_2] = \frac{\left[(1-p)^2 + \frac{p(1-p)}{2} + \frac{(1-p)p}{2} \right] \left(-\frac{1}{2} \right) + \frac{[p^2 + p(1-p)] \left(\frac{1}{2} \right)}{2} = \frac{2p-1}{4} = E[\pi_1] \quad (3.4)$$

So we have the interesting result, that $E[\pi_2] = E[\pi_1]$. In fact in general we can say that:

Proposition 7. *Agent k will have the same ex ante expected payoff as agent $k+1$ where $k \in N^{odd}$.*

Proof. Consider a sequence of N agents and let $k < N$ be an odd number. Now consider an arbitrary agent k and agent $k+1$. Agent k will either be in a cascade or not. We will consider these possibilities in turn. 1. If agent k is in a cascade then agent $k+1$ will simply duplicate agent k 's action and so will obtain the same expected payoff. 2. If agent k is not in a cascade then all the revealed information from agents 1 to $k-1$ must be neutral, i.e. can be ignored. This is only possible if k is odd, as otherwise the set of information arising from agents 1 to $k-1$ would always be biased towards one choice. Agent k will then decide based upon his own signal, choosing $A_K = Y$ if $X_k = H$ and

$A_K = N$ if $X_k = L$. The signal of agent k is therefore perfectly inferable to agent $k + 1$. Now we examine the return to agent $k + 1$. If agent $k + 1$ receives the same signal as agent k , so $X_{k+1} = X_k$ then he will choose as agent k did and therefore receive the same expected payoff. If agent k receives a different signal, so $X_{k+1} \neq X_k$, then he will have received one signal suggesting Y and one suggesting N and will therefore be perfectly indifferent between the two. Indifference implies the expected return from choosing either option is the same. Therefore one optimal policy for agent $k + 1$ is always to duplicate the choice of agent k regardless of his own signal. Since this is always optimal the expected payoffs of the two agents must be the same. For completeness we should note that as shown earlier payoffs are the same for agents 1 and 2, and since it has been shown true for arbitrary agents k and $k + 1$, where k is odd, it will be generally true by induction for $k < N$ where k is an odd number. \square

It is interesting to note the similarity between this proposition and the role of information played in Meyer (1991). Meyer (1991) shows that in certain circumstances an extra draw from an informative distribution will not increase the probability of a decision-maker choosing correctly between two alternatives. Meyer (1991) goes on to suggest biasing the second draw in such a way as to improve the probability of success and applies this to show the optimality of biased promotion tournaments.

We have the result that $E[\pi_n] = E[\pi_{n+1}]$ for $n \in N^{odd}$. This allows us to concentrate on agent $n \in N^{odd}$. Now consider what happens by agent $n - 1$. There are three possibilities:

- (1) We have a Y cascade, so agent n will go Y .
- (2) We have a N cascade, so agent n will go N .
- (3) We have no cascade, so agent n will go Y if $X_n = Y$ and will opt for N if $X_n = L$.

This allows the calculation of agent n 's payoff since we know that $\pi_n = \frac{1}{2}$ if $A_n = Y$ and $V = 1$, and $\pi_n = -\frac{1}{2}$ if $A_n = Y$ and $V = 0$, otherwise there will be no payoff. Therefore:

$$E [\pi_n \mid n \in N^{odd}] = \frac{\{\Pr[Y(n-1)|V=1]+p\Pr[No(n-1)|V=1]\}}{4} - \frac{\{\Pr[Y(n-1)|V=0]+(1-p)\Pr[No(n-1)|V=0]\}}{4}$$

This simply evaluates the probability that agent n goes for Y in both states of the world then weights this by the payoffs of $\frac{1}{2}$ and $-\frac{1}{2}$. Note that if there is no cascade we have to weight the chance of agent n going Y by the chance of his observing a H signal in each state of the world, that is $(1-p)$ if $V=0$ and p if $V=1$. This can be simplified to:

$$E [\pi_n \mid n \in N^{odd}] = \frac{2p-1}{4} \frac{1-(p-p^2)^{\frac{n+1}{2}}}{1-p+p^2} \quad (3.5)$$

It is also interesting to note that:

$$E [\pi_n \mid Y(n)] = \frac{1-(p-p^2)^{\frac{n-1}{2}}}{1-(p-p^2)} E [\pi_1]$$

So all payoffs are related to the payoff of the first agent. Similarly calculations for agent $n \in N^{even}$ yield:

$$E [\pi_n \mid n \in N^{even}] = \frac{2p-1}{4} \frac{1-(p-p^2)^{\frac{n}{2}}}{1-p+p^2} \quad (3.6)$$

3.3.2 Aggregate Consumer welfare

Define aggregate consumer welfare Ω as equal to *ex ante* expected payoffs summed over all N agents, so $\Omega = \sum_{n \in N} \pi_n$. Proposition 7 immediately tells us that since $E [\pi_n] = E [\pi_{n+1}]$ for $n \in N^{odd}$ we can consider aggregate payoffs to be simply:

$$\sum_{n \in N} \pi_n = \begin{cases} 2 \sum_{n \in N^{odd}} E\pi(n) & \text{for } N \in \mathbb{N}^{even} \\ 2 \left[\sum_{n \in N^{odd}} E\pi(n) \right] - \pi(N) & \text{for } N \in \mathbb{N}^{odd} \end{cases} \quad (3.7)$$

Based on equation 3.7 we can now calculate total *ex ante* expected consumer welfare since:

$$\sum_{n \in N^{odd}} E[\pi_n] = \frac{(2p-1) \left[N - \sum_{n \in N^{odd}} (p-p^2)^{\frac{n+1}{2}} \right]}{4(1-p+p^2)} \quad (3.8)$$

And so we have:

$$\Omega = \begin{cases} \frac{2p-1}{2(1-p+p^2)} \left[N - \sum_{n \in N^{odd}} (p-p^2)^{\frac{n+1}{2}} \right] & \text{for } N \in \mathbb{N}^{even} \\ \frac{2p-1}{2(1-p+p^2)} \left[N - \sum_{n \in N^{odd}} (p-p^2)^{\frac{n+1}{2}} \right] - \frac{1}{4} (2p-1) \frac{1 - (p-p^2)^{\frac{N+1}{2}}}{1-p+p^2} & \text{for } N \in \mathbb{N}^{odd} \end{cases} \quad (3.9)$$

Or alternatively:

$$\Omega = \left[N - \sum_{n \in N^{odd}} (p-p^2)^{\frac{n+1}{2}} - \sum_{n \in N^{even}} (p-p^2)^{\frac{n}{2}} \right] \quad (3.10)$$

This then provides the benchmark for an interventionist social planner. Any intervention that aims to strictly improve aggregate consumer welfare will have to raise overall payoffs above the non-intervention level of Ω .

3.4 THE SOCIAL PLANNER

Now we consider the role of a consumer welfare-maximizing social planner. We allow our social planner to force an additional $M \subset N$ agents to move in the first period. These agents will join with the first agent to give us a set of $M + 1$ “guinea pigs”. This has two major effects on consumer welfare:

(1) Those $M + 1$ who move first will not get access to later information and so will have less to use to help them make the optimal decision.

(2) Those $N - (M + 1)$ who move later will have more information at their disposal and so should have a better chance of making the correct decision.

We can therefore use the consumer welfare equation from the previous section and make two adjustments. We can use the sum for $N - (M + 1)$ agents and then add a further $ME[\pi_1]$ to the total, but first we must adjust the probability of being caught in a cascade to reflect the extra information made available to later movers. For simplicity we will assume that $N \in \mathbb{N}^{even}$, though the results in the section above allow us to look at $N \in \mathbb{N}^{odd}$ in an equivalent way. Now define $\Omega = \Omega(M, N, p)$ as the general level of consumer welfare. Since $N \in \mathbb{N}^{even}$ we can use the simple form for $M = 0$ (no intervention) given as:

$$\Omega(0, N, p) = \frac{2p-1}{2(1-p+p^2)} \left[N - \sum_{n \in N^{odd}} (p - p^2)^{\frac{n+1}{2}} \right] \quad (3.11)$$

3.4.1 Polar Examples

Consider what happens when $M = 1$ so we force two agents to act without social learning, rather than just a single agent as in the model in section 3. We then get 2 initial moves followed by a sequence of $N - 2$ agents. The two start-agents will share the same *ex ante* expected payoff as the first agent in the standard sequence, that is $\frac{1}{4}(2p - 1)$. However, the new sequence from the third agent onwards will have access to more noiseless information; 3 signals instead of 2. Therefore the agent deciding after the start-agents

may well face the prospect of a cascade before drawing his signal.

Consider the payoff of the third agent. He faces three possibilities, a Y cascade, N cascade or no cascade as before, but the probabilities are slightly different. Consider $V = 0$:

$$E[\pi_3 \mid V = 0, M = 1] = (1 - p)^2 \left(-\frac{1}{2}\right) + 2p(1 - p)^2 \left(-\frac{1}{2}\right)$$

Now consider $V = 1$:

$$E[\pi_3 \mid V = 1, M = 1] = p^2 \left(\frac{1}{2}\right) + 2p^2(1 - p) \left(\frac{1}{2}\right)$$

Therefore his unconditional payoff is:

$$E[\pi_3 \mid M = 1] = \frac{1}{4}(2p - 1)(2p - 2p^2 + 1) \quad (3.12)$$

Now note that in the non-intervention case we have:

$$E[\pi_3 \mid M = 0] = \frac{1}{4}(2p - 1)(p - p^2 + 1) \quad (3.13)$$

Now consider which has the higher value. Let us assume that equation 3.12 has the higher value than equation 3.13 and see if this is true: $\frac{1}{4}(2p - 1)(2p - 2p^2 + 1) > \frac{1}{4}(2p - 1)(p - p^2 + 1) \Rightarrow p - p^2 > 0$ which is clearly true for $p \in (0, 0.5)$. So the third agent gains slightly from the extra noiseless information, while the first two agents each receive what they would have anyway under non-intervention, expected payoffs of $\frac{1}{4}(2p - 1)$. Clearly agent 4 cannot be any worse off, so we have found an unambiguously better situation, hence $\Omega(1, N, p) > \Omega(0, N, p)$.

Now consider $M = N - 1$. In this case all the agents that usually move after the first agent are instructed to move immediately without scope for learning. The result here requires no analysis, all will move and gain the same *ex ante* expected payoff of $\frac{1}{4}(2p - 1)$, so $\Omega(N - 1, N, p) = \frac{1}{4}N(2p - 1)$. Now since $\Omega(0, N, p) > \frac{1}{4}N(2p - 1)$ this is clearly a worse situation for total consumer welfare than non-intervention. So we see by example that $\Omega(0, N, p)$ is not the highest consumer welfare achievable, but is also not the lowest. We need to find another candidate for optimal consumer welfare and therefore the optimal structure for our problem. This also constitutes a proof that:

Proposition 8. *The optimal level of M (number of guinea pigs) will lie strictly between 0 and $N - 1$.*

Proof. As shown above $\Omega(1, N, p) > \Omega(0, N, p)$ and $\Omega(0, N, p) > \Omega(N - 1, N, p)$ which immediately proves the proposition. \square

3.4.2 The Welfare-Maximizing Structure

So far we have $\Omega(1, N, p) > \Omega(0, N, p)$ and $\Omega(0, N, p) > \Omega(N - 1, N, p)$. The agent moving immediately after the $M + 1$ agents make the initial move, agent $M + 2$, will see various possibilities. Divide the set of size $M + 1$ into subsets M_Y and M_N which are made from those who choose Y and N respectively. Agent $M + 2$ faces three mutually exclusive possibilities:

- (1) $M_Y - M_N \geq 2$ which will create a Y cascade.
- (2) $M_N - M_Y \geq 2$ which will create a N cascade.
- (3) $M_Y - M_N \in (-2, 2)$ which produces no cascade.

Part 1 of the appendix derives the function:

$$\Omega(M, N, p) = \frac{(2p-1)(M+1)}{4} + \frac{\frac{1}{2}(2p-1)(M+1)!p^{\frac{M+2}{2}}(1-p)^{\frac{M+2}{2}}}{\left[\left(\frac{M+1}{2}\right)!\right]^2} \sum_{z=1}^{\frac{N-M}{2}} \sum_{s=1}^{\frac{2z-2}{2}} \left(\frac{p^2-p+1}{2}\right)^s$$

$$+ \frac{N-M-1}{4} \left\{ \frac{(2p-1)(M+1)! p^{\frac{M+2}{2}} (1-p)^{\frac{M+2}{2}}}{\left[\left(\frac{M+1}{2}\right)!\right]^2} + \sum_{x=0}^{\frac{M-1}{2}} \frac{(M+1)! [p^{M+1-x} (1-p)^x - (1-p)^{M+1-x} p^x]}{(M+1-x)! x!} \right\}$$

We then need to optimize this with respect to M and further calculations are made in the appendix. Despite the complexity of the function, a number of features stand out:

- (1) There is a unique maximum for all p and N assured by the concavity of $\Omega(M, N, p)$.
- (2) The maximum is in the interior of the range of M for all p and non-trivial N .

To give some examples, figures 2 to 4 show how the total consumer welfare value $\Omega(M, N, p)$ evolves as $M+1 \in \mathbb{N}^{even}$ increases, with N and p held fixed. Figure 2 fixes $N = 10$, figure 3 fixes $N = 50$ and figure 4 fixes $N = 100$. Figure 3 provides different optimal values for M for different values of p . However, as in figure 2 the optimal value of M is never 0 or N , instead it is given at some interior value. Figure 4 provides a convex shape and a set of interior optima for M . For $N = 10$ the optimal value of M is 3 for the five values of p examined. The existence of a trade-off between the value of guinea pigs and the loss to the guinea pigs' consumer welfare because of their failure to learn from others is captured by the convex shape of the function, which is a general feature, and is clearly visible in figures 3 and 4. Figure 5 summarizes figures 2, 3 and 4 by providing optimal M values for the given values of p and N . It is important to note the restriction $M+1 \in \mathbb{N}^{even}$.

Consider a particular example. For $N = 100$ and $p = \frac{2}{3}$, we have an optimal value of $M = 21$. This states that total consumer welfare will be maximized if we have a structure in which 21 of 100 agents join agent 1 and move immediately. Then the remaining 78 agents move in sequence. This gives agent 23 access to 23 signals unpolluted by possible herding, and guarantees that all agents will get full access to at least 23% of all signals. This provides significantly more consumer welfare than a standard herd when all act in strict sequence (twice as much) or when there is no social learning so all act independently (over twice as much).

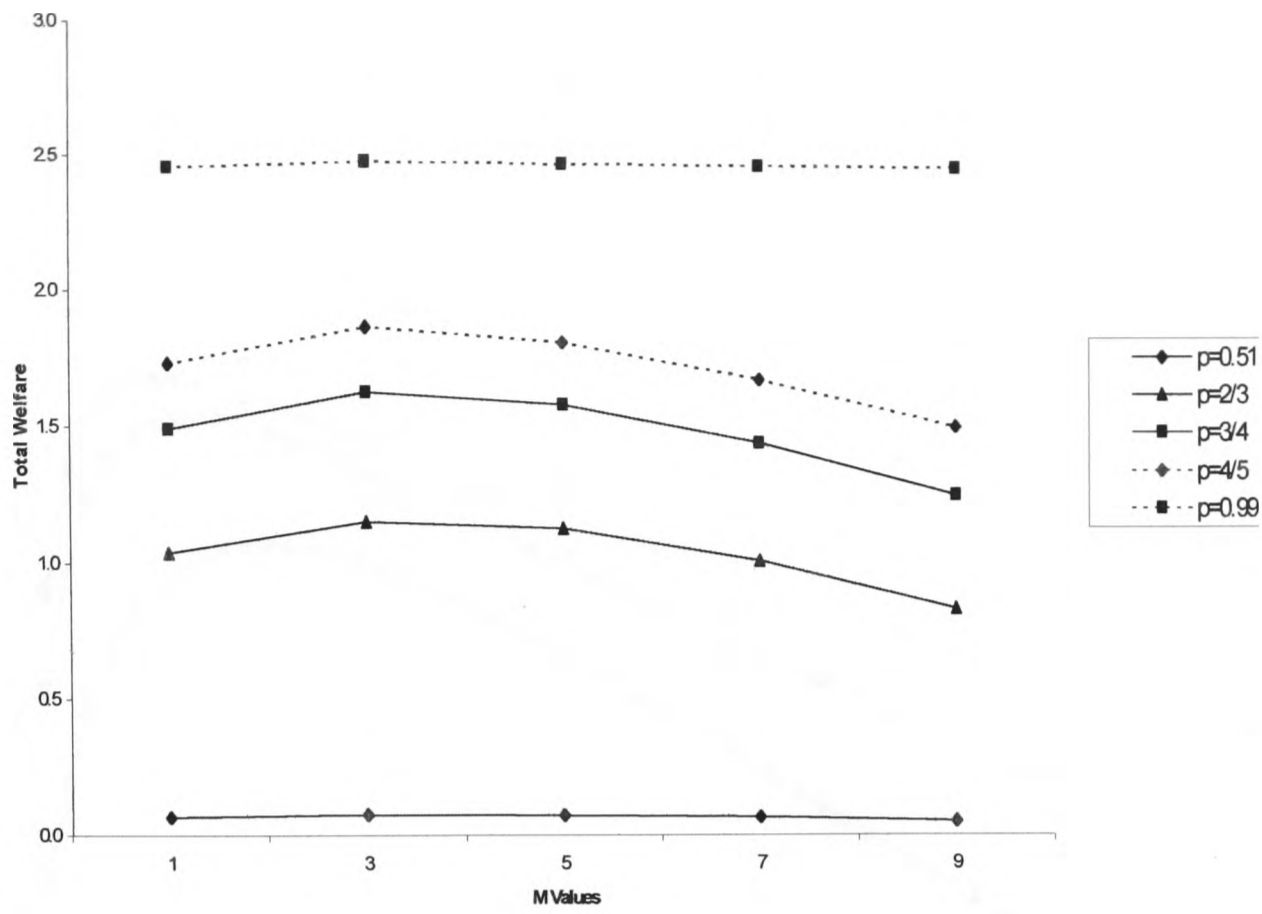


FIGURE 2: Optimal M Values, N=10

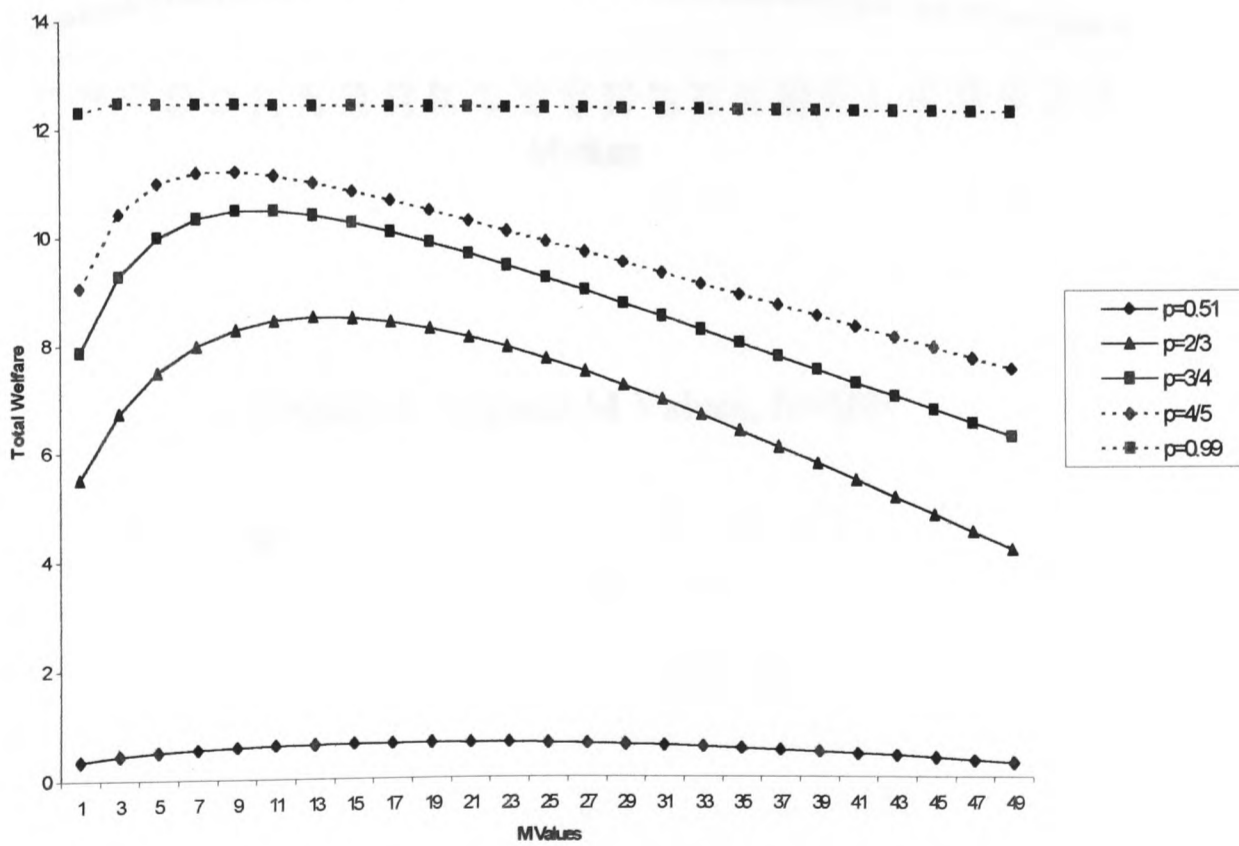


FIGURE 3: Optimal M Values, N=50

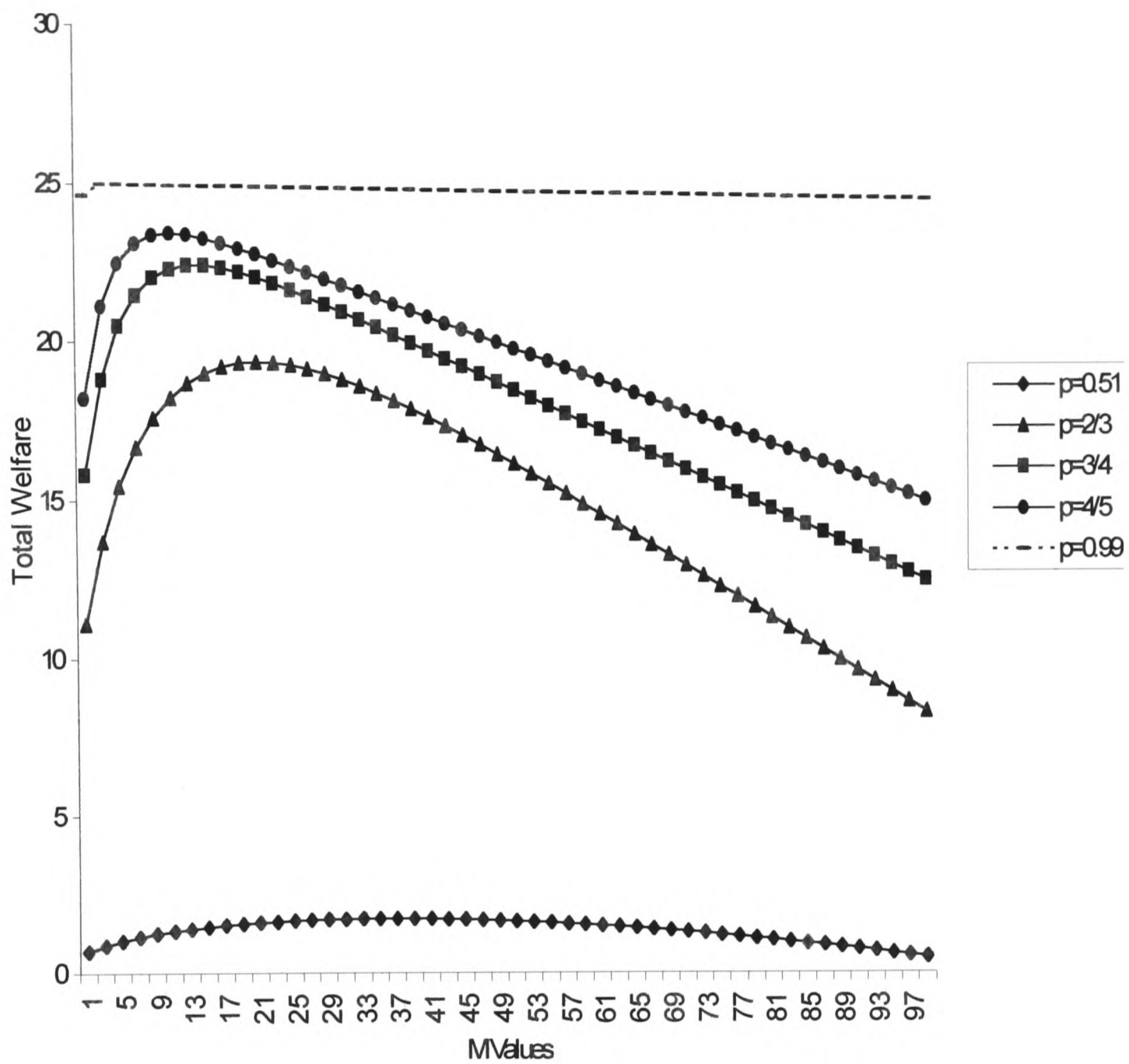


FIGURE 4: Optimal M Values, N=100

To give some comparative statics note that as p rises the optimal number of guinea pigs falls. This is the case since a rising p increases the probability of a good decision and a correct herd which is good for consumer welfare. The gains from increased guinea pigs are therefore not so great, while the disadvantage of reducing the information available to them is still present. As N rises the number of guinea pigs needs to rise, though the proportion falls.

p		0.51	2/3	3/4	4/5	0.99
N	50	19	13	11	9	3
	100	37	21	13	11	5

FIGURE 5: Optimal M Values for Consumer Welfare Given Values of p and N

3.4.3 A Note about Marriage

Before we move on to consider the firm, we will examine one alternative structure implied by proposition 7. The proposition reveals that for $k \in N^{odd}$ payoffs will always be identical for agents k and $k + 1$, the reasoning revolved around the valueless nature of agent k 's information to agent $k + 1$. This immediately tells us that it is at least weakly better for welfare to have all decisions "made in pairs".

Remark 1. *Welfare is weakly improved by grouping by going from a system of sequential decision-making with no guinea pigs to a system where all decision-making is done by a sequence of pairs, where each pair-member must decide without knowledge of their partner's decision.*

Consider a structure in which agents 1 and 2 decide simultaneously, then agents 3 and 4 observe agents 1 and 2, and also decide simultaneously, etc. We have eliminated only the useless observation of the direct predecessor when that predecessor can be indexed by an odd number. This will increase the information available to each successive pair in the same way as increasing the number of guinea pigs, but without the associated cost to the pair. The net result suggests that simultaneous decision-making by pairs is

at least as good as a strict sequence of decision-making. Taken literally this can provide an interesting argument for the informational gain of splitting the population into pairs via formal marriage or a similar informal link, but where each individual decides without reference to the decision made by their partners!

Remark 2. *In a sequential decision-making system with no guinea pigs, deference to one's partner is optimal.*

This remark is simply a corollary to proposition 7. Since agent $k + 1 \in N^{even}$ will be behaving optimally by copying the decision made by agent k , we have a theory of deference to one's partner! On informational grounds one possible optimal policy in a sequential decision-making world with no guinea pigs would be for all husbands to defer to their wives' decisions, or *vice versa*.

3.5 THE FIRM

Now we move away from the objective of consumer welfare maximization and instead examine the aims of the firm. Consider a single firm with a product it wishes to sell. Abstracting from profit-maximization, in the context of the current model we will consider the firm's aim to be simply to sell as many units as possible. We will first detail the optimal strategy of a firm which has a *good* product, i.e. for which $V = 1$ and then examine a firm which has a *bad* product, i.e. for which $V = 0$. Then we consider a firm that is uninformed about the value of V . We then compare the optimal actions of a firm with the optimal actions of a social planner.

3.5.1 Promoting a Good Product

Assume that $V = 1$ so the decision to purchase is the right one. Unfortunately the firm cannot convince all consumers that this is the case. However, the firm can manipulate the structure of the herd in much the same way as the social planner through the use of guinea pigs. In this case we might imagine the firm approaching a sub-set of all consumers and offering some incentive to make a quick decision. This can come in a variety of forms. The

firm might send time-limited money-off coupons to certain potential consumers, or offer free products to high profile consumers or members of the press who agree to advertise their experience through writing a review. Perhaps the best example is that of a movie premiere full of high profile celebrities and members of the press whose opinion will be sought.

Define the number of units sold as $Q_N(M+1) \equiv \#Y_N(M+1)$ which is a function of $M+1$ for a population of agents of size N and simply reads the number of Y decisions made by a population of N agents when there are an additional M guinea pigs chosen to decide with the first agent. In terms of the model the firm's objective is clearly to maximize the number of units sold. In order to do this the firm faces an important trade-off.

(1) It wishes to maximize the probability of a Y cascade by choice of M , since this will raise the number of purchases by those outside the initial decision group. For any given choice of M there will only be a remainder population outside the group of guinea pigs of size $N - M - 1$, so the population which learns is of size $N - M - 1$. Therefore the firm is interested in ensuring that this remainder population opts for a Y cascade, so intuitively it is interested in maximizing $(N - M - 1) \Pr [Y(M+1) | V = 1]$. A Y cascade will be initiated by the group of $M+1$ guinea pigs if $Q_{M+1} \geq \frac{M+1}{2} + 1$, a N cascade will be initiated if $Q_{M+1} \leq \frac{M+1}{2} - 1$ or alternatively there will be no net public information and no cascade will occur if $Q_{M+1} = \frac{M+1}{2}$. Having noted this it is easy to see that the probability of a Y cascade being initiated by given number of $M+1$ guinea pigs will be $\Pr [Q_{M+1} \geq \frac{M+1}{2} + 1 | V = 1]$.

(2) It also wishes to sell its product to as many of the guinea pigs as possible. The sales to the first $M+1$ is very simply defined as $p(M+1)$ since there will be no learning within this group.

Furthermore, the firm also knows that even if a Y cascade is not initiated by the initial group of guinea pigs later agents may still initiate a Y cascade.

The second part of the appendix reduces the firm's problem to:

$$\max_M \left\{ p(M+1) + (N-M-1)(M+1)! \sum_{x=0}^{\frac{M-1}{2}} \frac{p^{M+1-x}(1-p)^x}{x!(M+1-x)!} + \right. \quad (3.14)$$

$$\left. \left(1 - (M+1)! \sum_{x=0}^{\frac{M-1}{2}} \frac{p^{M+1-x}(1-p)^x + p^x(1-p)^{M+1-x}}{x!(M+1-x)!} \right) \frac{p(p+1)}{2} \sum_{n=2}^{N-M-1} \frac{\binom{N-M-\frac{n}{2}}{n} \left(1 - (p-p^2)^{\frac{n-M-1}{2}} \right)}{1-(p-p^2)} \right\}$$

Differentiating this requires the use of the digamma and hypergeometric distributions and produces a fairly complex result. However, some comparative statics should provide some intuition for the result. Figure 6 gives the optimal choice of M for various values of p and N , and provides an interesting comparison with figure 5.

p		0.6	2/3	3/4	4/5
N	50	23	19	13	11
	100	37	27	17	15

FIGURE 6: Optimal M Values for the Firm when $V = 1$ for Given Values of p and N

Figure 7 gives the expected number of units sold for various different choices of M by the firm for a market of size $N = 100$ for different values of p .

M		9	29	49	69	89
	0.51	41	48	51	51	51
p	2/3	77	87	83	77	70
	4/5	95	94	90	86	82

FIGURE 7: Expected Units Sold for Different Values of M , $N = 100$

Figure 8 holds p constant at $2/3$ and varies the size of the market, again looking at the impact on the expected number of units sold (with percentage of market size in brackets) of a change in M .

M	9	29	49	69	89	
	100	77 (77)	87 (87)	83 (83)	77 (77)	70 (70)
N	150	117 (78)	135 (90)	132 (88)	126 (84)	120 (80)
	250	195 (78)	230 (92)	231 (92)	226 (90)	220 (88)

FIGURE 8: Expected Units Sold for Different Values of M , $p = \frac{2}{3}$

Finally, figure 9 considers the percentage of the market which purchases the product when $p = \frac{2}{3}$ and we vary N and the ration of M/N .

M/N	9%	25%	49%	75%	91%	
	100	77%	87%	83%	75%	69%
N	150	83%	89%	83%	75%	69%
	250	91%	91%	83%	75%	70%

FIGURE 9: Success Rate for Different Percentages of the Market Forced to Decide

Early, $p = \frac{2}{3}$

Analysing figures 6 to 9 reveals a number of interesting comparative statics. Firstly the impact of raising M on total number of units purchased is non-monotonic. So we do not expect corner-solutions. Secondly, the impact of M is very dependent on the value of N and p . Thirdly, optimal M is rising in N but falling in p . Finally, switching to percentages reduces the importance of N but does not eliminate it, so the solution cannot be expressed as a fixed percentage of the market for a given p . Some casual observations would add that a figure of around 25% of the market for $p = \frac{2}{3}$ whilst not optimal seems reasonable for an N between 100 and 250, though it is a little high for N approaching 250. So the trade-off gives us a value of M which is nicely in the interior, and not too

high a level for a reasonable value of p . As for the impact of N and p we can reason as follows. As p rises the chance of a Y cascade without resort to guinea pigs rises and this seems sufficient to outweigh the similarly beneficial fall in the number of guinea pigs who do not purchase from the firm. Therefore, a rising p value indicates that the number of guinea pigs should be reduced, holding N constant. A rising N value indicates that the number of guinea pigs should rise, though not as a percentage of N . So the firm should raise the absolute number but reduce the percentage of the market acting as guinea pigs. This seems sensible given that market size is decreasingly important for learning in a herding model, since once a herd has started it will not stop, regardless of the number of agents remaining in the sequence.

3.5.2 Promoting a Bad Product

Now we consider the case when $V = 0$. Part 3 of the appendix shows that the firm's problem has now changed to become:

$$\max_M \left\{ (1-p)(M+1) + (N-M-1)(M+1)! \sum_{x=0}^{\frac{M-1}{2}} \frac{(1-p)^{M+1-x} p^x}{x!(M+1-x)!} + \right. \quad (3.15)$$

$$\left. \left(1 - (M+1)! \sum_{x=0}^{\frac{M-1}{2}} \frac{p^{M+1-x} (1-p)^x + p^x (1-p)^{M+1-x}}{x!(M+1-x)!} \right) \frac{(p-2)(p-1)}{2} \sum_{n=2}^{N-M-1} \frac{(N-M-\frac{n}{2}) \left(1 - (p-p^2)^{\frac{n-M-1}{2}} \right)}{1-(p-p^2)} \right\}$$

Now we carry out some of the comparative statics from the previous subsection with the only difference being the move from the $V = 1$ state to the $V = 0$ state. Figure 10 repeats the findings of figure 6 for the new state. Note the collapse in the value of M that would be selected by the firm as we move from state $V = 1$ to $V = 0$. The firm has to carefully balance the desire to initiate a Y cascade by manipulating the number of

guinea pigs, by the need to avoid too much information being revealed and a N cascade being initiated.

p		0.6	2/3	3/4	4/5
N	50	17	13	9	7
	100	27	19	11	9

FIGURE 10: Optimal M Values for the Firm when $V = 0$ for Given Values of p and N

Figure 11, much like figure 7, gives the expected number of units sold for various different choices of M by the firm for a market of size $N = 100$ for different values of p . The figures for optimal M when $p = 0.51$ are not surprisingly very similar, but moving to a higher figure for p yields very different results with far fewer units being sold especially for higher values of M supporting the findings in figure 10.

M		9	29	49	69	89
	0.51	41	48	50	50	49
p	2/3	74	77	66	53	40
	4/5	89	76	60	44	28

FIGURE 11: Expected Units Sold for Different Values of M , $N = 100$

Figure 12 carries out the same process as figure 8 but for $V = 0$, holding p constant at $2/3$ and varying the size of the market, looking at the impact on the expected number of units sold (with percentage of market size in brackets) of a change in M .

M		9	29	49	69	89
	100	74 (74)	77 (77)	66 (66)	53 (53)	40 (40)
N	150	113 (75)	125 (83)	116 (77)	103 (69)	90 (60)
	250	192 (77)	220 (88)	215 (86)	203 (86)	190 (76)

FIGURE 12: Expected Units Sold for Different Values of M , $p = \frac{2}{3}$

Finally, figure 13 mirrors figure 8 by evaluating the percentage of the market which purchases the product when $p = \frac{2}{3}$ for various values of N and M/N .

M/N		9%	25%	49%	75%	91%
	100	74%	77%	66%	53%	40%
N	150	80%	81%	67%	49%	39%
	250	88%	83%	67%	50%	39%

FIGURE 13: Success Rate for Different Percentages of the Market Forced to Decide
Early, $p = \frac{2}{3}$

3.5.3 Summarizing the Effect on Welfare

The firm's desire to maximize the number of units sold when the state is $V = 1$ is intuitively consumer welfare maximizing, since consumers would wish to purchase given $V = 1$. Therefore the firm's problem is effectively identical to the problem $\max_M \Omega(M, N, p | V = 1)$. However, if we consider $V = 0$ it should be immediately obvious that the firm will not maximize consumer welfare. In fact the firm is interested in maximizing the probability of an incorrect herd by choice of M and so is interested in *minimizing* consumer welfare. To summarize:

Agent	Social Planner	Firm, $V = 1$	Firm, $V = 0$
Optimal Policy	Maximize Ω	Maximize Ω	Minimize Ω

FIGURE 14: Consumer Welfare

3.5.4 A Note about Revelation

If a social planner has a reasonable knowledge of p and of N it might be possible to deduce whether the firm is promoting a good product ($V = 1$) or a bad product ($V = 0$) based on the firm's choice of M . In fact a consumer with a very vague knowledge of p may be able to isolate the firm's own perceptions of its product by its choice of M . This would have a reinforcing effect on the policy of a firm when $V = 1$ helping to ensure success for the product, and equivalently reduce the success of a product when $V = 0$. For example,

if a consumer feels very uncertain about the quality of his signal, considering it to range between 0.4 and 0.6 and knows the market is of size 100, and he observes an impressive promotion campaign targeting about 40% of the market he might deduce that the firm was selecting the optimal policy given $V = 1$. Of course, this might convince a firm in the state $V = 0$ to increase its efforts and raise M to provide a misleading signal. In fact since it is generally true that a firm will wish to opt for a much higher M when $V = 1$, a high value of M is suggestive of $V = 1$. If we follow this avenue of thought we would need to consider ways in which a firm in state $V = 1$ can differentiate itself from a firm in the state $V = 0$ providing some cost function for raising M that is lower if $V = 1$ and solving the firm's problem given a set of incentive compatibility and individual rationality constraints. This returns us to more traditional signalling models which might work in combination with herding to provide an interesting extension to this chapter. One way of completely removing this concern is to assume that *only the firm knows the value of N* . This makes any backward inference from M to V impossible, and is in fact quite a reasonable assumption in most markets.

3.6 CONCLUSIONS

The herd externality is potentially a force which can damage welfare and profits, and we would therefore expect a social planner or firm to act against the potential loss of information when their interests are at risk. This chapter suggested that by altering the structure of the sequence to include guinea pigs, both planner and firm can partly overcome the herd externality. Allowing a new variable to enter into the maximization of welfare or profits greatly complicates the problem. However, the solution will be that a positive number of consumers will be forced to move early, and their loss will benefit later decision-makers.

We might consider rejecting the sequence completely, and allowing consumers to make decisions whenever they wish. This endogenizes the timing of the decision-making process, and the next part of this thesis is devoted to just such an alternative structure.

APPENDIX FOR CHAPTER 3

1. The Optimal Choice of M by a Social Planner

Note that when $V = 1$, we have a probability of a Y cascade being introduced by our $M + 1$ agents of:

$$p^{M+1} + \frac{(M+1)!}{M!} p^M (1-p) + \frac{(M+1)!}{(M-1)!2!} p^{M-1} (1-p)^2 + \dots + \frac{(M+1)!}{(M+1-x)!x!} p^{M+1-x} (1-p)^x$$

Where x is the highest whole number less than $\frac{M-1}{2}$. We can simplify this to:

$$\Pr [M_Y - M_N \geq 2 \mid V = 1] = \sum_{x=0}^{\frac{M-1}{2}} \frac{(M+1)!}{(M+1-x)!x!} p^{M+1-x} (1-p)^x$$

For $V = 0$ by a similar calculation we have:

$$\Pr [M_Y - M_N \geq 2 \mid V = 0] = \sum_{x=0}^{\frac{M-1}{2}} \frac{(M+1)!}{(M+1-x)!x!} (1-p)^{M+1-x} p^x$$

Which yields the unconditional probability:

$$\Pr [M_Y - M_N \geq 2] = \frac{1}{2} \sum_{x=0}^{\frac{M-1}{2}} \frac{(M+1)!}{(M+1-x)!x!} \left[p^{M+1-x} (1-p)^x + (1-p)^{M+1-x} p^x \right] \quad (3.16)$$

Now for $M_Y = M_N \in (-2, 2)$ which results in no cascade we have $M_Y = M_N$ in the case when $M + 1 \in \mathbb{N}^{even}$. We will from now on concentrate on this case for simplicity.

So we have:

$$\Pr [M_Y = M_N] = \frac{(M+1)!}{\left[\left(\frac{M+1}{2}\right)!\right]^2} p^{\frac{M+2}{2}} (1-p)^{\frac{M+2}{2}} \quad (3.17)$$

We would need $X_{M+2} = Y$ to induce our $M + 2$ nd agent to choose Y after observing $M_Y = M_N$, which will happen with probability p if $V = 1$ and probability $(1 - p)$ if $V = 0$ and yield a payoff of $\frac{1}{2}$ or $-\frac{1}{2}$ respectively. Therefore we have a payoff under the no cascade assumption of:

$$E [\pi_{M+2} | M_Y = M_N] = \frac{2p-1}{4} \left\{ \frac{(M+1)!}{\left[\left(\frac{M+1}{2}\right)!\right]^2} p^{\frac{M+2}{2}} (1-p)^{\frac{M+2}{2}} \right\} \quad (3.18)$$

Now we add in the probability of our agent being caught in a herd on Y , to yield an unconditional expected payoff:

$$E [\pi_{M+2}] = \frac{1}{4} \left\{ \frac{(2p-1)(M+1)! p^{\frac{M+2}{2}} (1-p)^{\frac{M+2}{2}}}{\left[\left(\frac{M+1}{2}\right)!\right]^2} + \sum_{x=0}^{M-1} \frac{(M+1)! [p^{M+1-x} (1-p)^x - (1-p)^{M+1-x} p^x]}{(M+1-x)! x!} \right\} \quad (3.19)$$

For example, in the $M = 1$ case, equation 3.19 gives us, just as in equation 3.12:

$$E [\pi_3 | M = 1] = \frac{1}{4} (2p - 1) (1 + 2p - 2p^2)$$

The total consumer welfare expression is:

$$\Omega (M, N, p) = \frac{1}{4} (2p - 1) (M + 1) + \{E [\pi_{M+2}] + E [\pi_{M+3}] + \dots + E [\pi_N]\}$$

Which we simply need to express fully in terms of M and then optimize with respect to M . We need to note the relationship between $E[\pi_{M+2}]$ and $E[\pi_{M+i}]$ for $i \in \{3, 4, \dots, N\}$. First note that from proposition 7. Therefore we have:

$$\Omega(M, N, p) = \frac{1}{4}(2p - 1)(M + 1) + 2 \sum_{i=1}^{\frac{N-M}{2}} E[\pi_{M+2i}]$$

Comparing agent $M + 4$ to agent $M + 2$ we see that agent $M + 4$ basically gets the same payoff but faces a slightly higher chance of being in a Y cascade which is good when $V = 1$ and bad when $V = 0$. This extra probability is basically just the chance that no cascade was initiated before agent $M + 2$, but that agent $M + 2$ and agent $M + 3$ both went for action Y . This occurs with probability

$$\Pr[No \mid M + 2] \left\{ \frac{1}{2} [p^2 + \frac{1}{2}p(1-p)] + \frac{1}{2} [(1-p)^2 + \frac{1}{2}p(1-p)] \right\}$$

and the net gain is simply $\frac{1}{4}(2p - 1)$. So we get:

$$E[\pi_{M+4}] = E[\pi_{M+2}] + \frac{p^2 - p + 1}{2} \frac{\frac{1}{4}(2p-1)(M+1)! p^{\frac{M+2}{2}} (1-p)^{\frac{M+2}{2}}}{\left[\left(\frac{M+1}{2}\right)!\right]^2}$$

For agent $M + 6$ we have a similar calculation except that now we need no cascade before agent $M + 4$:

$$E[\pi_{M+6}] = E[\pi_{M+4}] + \left(\frac{p^2 - p + 1}{2}\right)^2 \frac{\frac{1}{4}(2p-1)(M+1)! p^{\frac{M+2}{2}} (1-p)^{\frac{M+2}{2}}}{\left[\left(\frac{M+1}{2}\right)!\right]^2}$$

And so on, so in general for $z > 1$:

$$E[\pi_{M+2z}] = E[\pi_{M+2z-2}] + \left(\frac{p^2-p+1}{2}\right)^{\frac{2z-2}{2}} \frac{\frac{1}{4}(2p-1)(M+1)!p^{\frac{M+2}{2}}(1-p)^{\frac{M+2}{2}}}{\left[\left(\frac{M+1}{2}\right)!\right]^2}$$

Clearly we have a nested structure and can therefore resolve this as follows:

$$E[\pi_{M+6}] = E[\pi_{M+2}] + \left[\frac{p^2-p+1}{2} + \left(\frac{p^2-p+1}{2}\right)^2\right] \frac{\frac{1}{4}(2p-1)(M+1)!p^{\frac{M+2}{2}}(1-p)^{\frac{M+2}{2}}}{\left[\left(\frac{M+1}{2}\right)!\right]^2}$$

And in general:

$$E[\pi_{M+2z}] = E[\pi_{M+2}] + \frac{\frac{1}{4}(2p-1)(M+1)!p^{\frac{M+2}{2}}(1-p)^{\frac{M+2}{2}}}{\left[\left(\frac{M+1}{2}\right)!\right]^2} \sum_{s=1}^{\frac{2z-2}{2}} \left(\frac{p^2-p+1}{2}\right)^s \quad (3.20)$$

So for total consumer welfare we have:

$$\Omega(M, N, p) = \frac{(2p-1)(M+1)}{4} + 2 \sum_{z=1}^{\frac{N-M}{2}} E[\pi_{M+2z}] \quad (3.21)$$

$$= E[\pi_{M+2}](N-M-1) + \frac{(2p-1)(M+1)}{4} + \frac{\frac{1}{4}(2p-1)(M+1)!p^{\frac{M+2}{2}}(1-p)^{\frac{M+2}{2}}}{\left[\left(\frac{M+1}{2}\right)!\right]^2} \sum_{z=1}^{\frac{N-M}{2}} \sum_{s=1}^{\frac{2z-2}{2}} \left(\frac{p^2-p+1}{2}\right)^s$$

We can now insert the expression for $E[\pi_{M+2}]$ to yield:

$$\Omega(M, N, p) = \frac{(2p-1)(M+1)}{4} + \frac{\frac{1}{2}(2p-1)(M+1)!p^{\frac{M+2}{2}}(1-p)^{\frac{M+2}{2}}}{\left[\left(\frac{M+1}{2}\right)!\right]^2} \sum_{z=1}^{\frac{N-M}{2}} \sum_{s=1}^{\frac{2z-2}{2}} \left(\frac{p^2-p+1}{2}\right)^s$$

$$+ \frac{N-M-1}{4} \left\{ \frac{(2p-1)(M+1)!p^{\frac{M+2}{2}}(1-p)^{\frac{M+2}{2}}}{\left[\left(\frac{M+1}{2}\right)!\right]^2} + \sum_{x=0}^{\frac{M-1}{2}} \frac{(M+1)! \left[p^{M+1-x}(1-p)^x - (1-p)^{M+1-x} p^x \right]}{(M+1-x)!x!} \right\}$$

Now we differentiate with respect to M using a convenient set of abbreviations:

$$A = \frac{1}{2} (N - M + 2)$$

$$B = \left(\frac{M+1}{2}\right)!$$

$$C = \frac{1}{2} (p^2 + p + 1)$$

$$D = -\frac{A(p^4 - 2p^3 + p^2 - 1) - 4C^A - p^2 + 2p - 3 + p^4 - 2p^3}{(p^2 - p - 1)^2}$$

$$E = (2p - 1)(M + 1)! p^{\frac{M+2}{2}} (1 - p)^{\frac{M+2}{2}}$$

$$\Omega'(M, N, p) = \frac{2p-1}{4} + \frac{E}{B^2} \left(\frac{2D+N-M-1}{2} \frac{2\Psi(M+2)+\ln(p)+\ln(1-p)-2\Psi\left(\frac{M+3}{2}\right)}{4} - \frac{2C^A \ln C - \frac{1}{2}p^4 + p^3 - \frac{1}{2}p^2 + \frac{1}{2}}{2(p^2-p-1^2)} - 1 \right)$$

$$- \frac{1}{4} \sum_{x=0}^{\frac{M-1}{2}} \frac{(M+1)! (p^{M+1-x}(1-p)^x - (1-p)^{M+1-x} p^x)}{(M+1-x)! x!} + \frac{N-M-1}{4} \frac{\partial}{\partial M} \sum_{x=0}^{\frac{M-1}{2}} \frac{(M+1)! (p^{M+1-x}(1-p)^x - (1-p)^{M+1-x} p^x)}{(M+1-x)! x!}$$

Where $\Psi(\cdot)$ is the digamma function. This expression can then be set equal to zero to yield an implicit function for the optimal value of M given p and N . Furthermore $\Omega_{MM}(M, N, p) < 0$ across the whole range of p and N implying that $\Omega(M, N, p)$ is concave, which is sufficient to provide a unique maximum.

2. The Firm's Problem when $V=1$

Expression 3.14 is composed of three terms. The first term is simply $p(M+1)$, the number of units purchased within the group of guinea pigs, and is simply the probability of a high signal given $V=1$ multiplied by the size of the initial group. The second term is more complex $(N-M-1)(M+1)! \sum_{x=0}^{\frac{M-1}{2}} \frac{p^{M+1-x}(1-p)^x}{x!(M+1-x)!}$. This is the size of the remaining population of agents, $N-M-1$, multiplied by the probability of a Y cascade being induced by the initial group, which is:

$$\Pr [Q_{M+1} \geq \frac{M+1}{2} + 1 \mid V=1] = (M+1)! \sum_{x=0}^{\frac{M-1}{2}} \frac{p^{M+1-x}(1-p)^x}{x!(M+1-x)!}$$

The derivation of the final term in expression 3.14 incorporates the possibility that the initial group failed to initiate a Y cascade. Despite this there is still a good chance of a Y cascade being initiated by later agents. Start with a signal which is on aggregate

neutral, being revealed by the guinea pigs, which occurs with probability:

$$\Pr \left[Q_{M+1} = \frac{M+1}{2} \mid V = 1 \right] = 1 - (M+1)! \sum_{x=0}^{\frac{M-1}{2}} \frac{p^{M+1-x}(1-p)^x + p^x(1-p)^{M+1-x}}{x!(M+1-x)!}$$

Which is simply one minus the combined probability of a Y cascade and a N cascade. Now consider the actions of the decisions made by the post-guinea pig agents. If n agents make decisions and initiate a Y cascade this must involve $\frac{1}{2}(n-2)$ agents choosing Y and $\frac{1}{2}(n-2)$ agents choosing N with a crucial 2 agents tipping the balance in favour of a Y cascade. So of the n we have $\frac{1}{2}(n-2)+2$ agents deciding to purchase. Now we have only a population of size $N-M-1-n$ remaining. So we have a total of $N-M-\frac{1}{2}n$ who purchase in the event of a Y cascade being initiated by agent n . For example, if $n=2$ then the Y cascade failed to be initiated by the initial group of guinea pigs, but the $M+2$ nd agent and the $M+3$ rd agent both decide to purchase initiating a Y cascade which still results in the entire $N-M-1$ purchasing. If $n=4$, then from the first 4 after the initial group of $M+1$, 3 will decide to purchase and 1 will decide otherwise, resulting in $N-M-2$ units being purchased. This all has to be multiplied by the probability of no cascade being initiated by the group of guinea pigs and the probability of a Y cascade being initiated by the group of n immediately following the guinea pig, which is therefore:

$$\sum_{n=2}^{N-M-1} \left[(N-M-\frac{n}{2}) \frac{p(p+1)}{2} \frac{1-(p-p^2)^{\frac{n-M-1}{2}}}{1-(p-p^2)} \left(1 - (M+1)! \sum_{x=0}^{\frac{M-1}{2}} \frac{p^{M+1-x}(1-p)^x + p^x(1-p)^{M+1-x}}{x!(M+1-x)!} \right) \right]$$

Combining these three parts yields a function giving the total number of units sold by the firm as a function of p , N and M . Of these we allow the firm to vary only M making the final problem to maximize expression 3.14 by choice of M .

3. The Firm's Problem when $V=0$

Expression 3.15 is also composed of three terms. The first term is now $(1-p)(M+1)$, since the probability of a high signal given $V=0$ has changed to be $1-p$. The second term has also slightly changed to now be: $(N-M-1)(M+1)! \sum_{x=0}^{\frac{M-1}{2}} \frac{(1-p)^{M+1-x} p^x}{x!(M+1-x)!}$. This is the size of the remaining population of agents, $N-M-1$, multiplied by the new probability of a Y cascade being induced, now that $V=0$, which is:

$$\Pr [Q_{M+1} \geq \frac{M+1}{2} + 1 \mid V=0] = (M+1)! \sum_{x=0}^{\frac{M-1}{2}} \frac{(1-p)^{M+1-x} p^x}{x!(M+1-x)!}$$

The derivation of the final term in the expression is much as in the case when $V=1$ except we now use the probability that a Y cascade occurs given $V=0$. Note that the aggregate neutral signal being revealed by the guinea pigs occurs with the same probability as before, so:

$$\Pr [Q_{M+1} = \frac{M+1}{2} \mid V=0] = 1 - (M+1)! \sum_{x=0}^{\frac{M-1}{2}} \frac{p^{M+1-x}(1-p)^x + p^x(1-p)^{M+1-x}}{x!(M+1-x)!}$$

The population size is also the same, at $N-M-\frac{n}{2}$. However, the new Y cascade probability changes the final term to:

$$\sum_{n=2}^{N-M-1} \left[(N-M-\frac{n}{2}) \frac{(p-2)(p-1)}{2} \frac{1-(p-p^2)^{\frac{n-M-1}{2}}}{1-(p-p^2)} \left(1 - (M+1)! \sum_{x=0}^{\frac{M-1}{2}} \frac{p^{M+1-x}(1-p)^x + p^x(1-p)^{M+1-x}}{x!(M+1-x)!} \right) \right]$$

Combining these three parts once again yields the function which the firm will maximize by choice of M .

PART II
HERDING IN ENDOGENOUS TIME

CHAPTER 4

HERDING IN TIME AND CHOICE

4.1 INTRODUCTION

This part of the thesis moves into the realm of endogenous timing. Here we consider agents to have the ability to choose both *what* action to take, and *when* to take it. Therefore, the ordering of actions becomes no longer sequential, but is determined endogenously within the model. Since the ordering is most commonly indexed by time, this class of models can usefully be called *endogenous timing models*. Perhaps the earliest work in this direction occurred in Bikchandani, Hirshleifer, and Welch (1992), with a simple argument based on the precision of signals. However, two other papers have more or less defined the research agenda into endogenous timing. The paper by Chamley and Gale (1994) models decision-making in a discrete action space, using an options methodology, and finds considerable inefficiency. The model developed by Gul and Lundholm (1995) uses continuous action spaces, agents choosing from an interval, and finds no such inefficiency. Both types of model are examined in depth in this chapter before the next two chapters develop and test some of the ideas first seen within these two papers.

4.2 THE PRECISION OF SIGNALS

As a precursor to the development of a full endogenous-timing model in which agents can drop out of sequence if they wish, we will examine the final major element of the

Bikhchandani, Hirshleifer, and Welch (1992) model, concerning different signal precisions among agents. They introduce a very crude notion of endogenous timing as part of their examination of a scenario in which agents have different signal precisions. Returning to the binary signal case in section 2.2.1, assume as before that $\Pr[V = 1] = \Pr[V = 0] = 0.5$.

Proposition 9. *(Bikhchandani, Hirshleifer, and Welch, 1992). Suppose the model is as described in section 2.2.1 of chapter 2. If $C = 0.5$ and if the agent with the highest precision signal decides first, then the first agent's decision is copied by all later agents.*

Proof. The second agent will infer the first agent's signal, and since agent 1's signal is superior to agent 2's signal, agent 2 will ignore his own signal and follow agent 1, initiating a cascade. \square

Proposition 10. *(Bikhchandani, Hirshleifer, and Welch, 1992). Again, suppose that the model is as described by section 2.2.1 of chapter 2. Assume that all agents $n > 1$ observe signals of identical precision. Then all agents $n > 2$ are better off if the first agent's signal precision is slightly lower rather than slightly higher than theirs.*

Proof. If agent 1's signal is slightly higher in precision the second agent will defer to the first. If agent 1's signal is of slightly lower precision, agent 2 will follow his own signal, thus revealing more information for later agents. \square

Bikhchandani *et al* assume that the first agent has the highest precision signal as a simple form of endogeneity in the ordering. If agents can choose whether to decide or delay, with some cost of delay, then they have an incentive to wait and hope to free-ride on the first to decide. There are two competing influences, a desire to wait and free-ride and a growing cost to holding back.⁸

Bikhchandani, Hirshleifer, and Welch (1992) argues that all things being equal the cost of deciding early is lowest for the agent with the highest precision. By allowing signal

⁸These two forces will recur again and again, and in particular are examined in the next chapter as the competing forces of accuracy and delay.

precision to vary a new possibility emerges: that a later agent with high precision could break a cascade because he is more inclined to use his own high precision information than to follow the herd. This possibility, which Bikhchandani *et al* call *cascade reversal*, is socially beneficial since it leads to more information being revealed and thus better decisions being taken later on in the sequence.

4.3 MODELS OF DISCRETE CHOICE

The Chamley and Gale (1994) and Gale (1996) papers concentrate on a world in which only discrete choice is allowed. Chamley and Gale consider N agents choosing whether to invest or delay, n of whom have an investment option, exercisable at only one discrete and countable date chosen by the investor: $t \in \mathbb{N}^{++}$. Discounting is at the common factor $\delta \in (0, 1)$. Therefore, for one of the n agents with an investment option, payoffs are $\delta^{t-1}V(n)$ from investing and 0 from never investing. Denote the agent i 's action at time t as $x_{it} = 1$ if agent i invests and $x_{it} = 0$ if not. Outcomes are given by a vector of decisions at time t , $x_t = (x_{1t}, x_{2t}, \dots, x_{Nt})$. History is given by the sequence of outcomes at t , $h_t = (x_1, x_2, \dots, x_{t-1}) \in H_t$. Let $H_1 = \{\emptyset\}$ be the initial history and $H = \bigcup_{t=1}^{\infty} H_t$ be the set of all histories.

Chamley and Gale (1994) look for a symmetric equilibrium in which agents actions depend only on their type and publicly observed information. Those $N - n$ with no options are passive and so we only need to describe strategies and beliefs for the n agents with options. For any history h , let $\lambda(h)$ denote the probability that an agent who has not yet exercised his option does so after observing history h . A behavioural strategy is a function $\lambda : H \mapsto [0, 1]$. Since a player knows only his own type he is facing an extensive form game of incomplete information and must make a probability assessment $\mu : H \times N \mapsto [0, 1]$ with the interpretation that $\mu(n | h)$ is the probability that n agents have options given the history h . A *perfect Bayesian equilibrium* (henceforth PBE) in this game consists of a strategy λ and a probability assessment μ such that: (i) each agent's strategy is a best response at every information set; (ii) the probability assessments are

consistent with Bayes' rule at every information set reached with positive probability. An *equilibrium path* E is the set of histories that occur with positive probability in equilibrium.

Initially, Chamley and Gale characterize E for a fixed but arbitrary symmetric PBE (λ, μ) . Let $V(h)$ denote the payoff from immediate investment at the information set h , and let $W(g, h)$ denote the undiscounted payoff from waiting one period at the information set h and then making an irrevocable decision, when others invest with probability g . Let (h, k) denote the information set reached if k players invest after history h . Write $V(\emptyset)$ for V , $V(\emptyset, k)$ for $V(k)$, $W(\lambda, \emptyset)$ for $W(\lambda)$, and so on. Finally, let $W^*(h)$ be the equilibrium payoff from waiting at the information set h . Suppose a player waits at h and makes a once-for-all decision either to invest at (h, k) or never to invest. His payoff at (h, k) will be $\sum_k p(k | h) \max\{V(h, k), 0\}$ where $p(k | h)$ is the distinguished agent's probability assessment that k others will invest at h . At this point Chamley and Gale use the following concept:

Definition 6. *The One-Step Property.* (Chamley and Gale, 1994). *The one-step property (henceforth OSP) is satisfied at h , if:*

$$W^*(h) = W(\lambda(h), h) \equiv \sum_k p(k | h) \max\{V(h, k), 0\}$$

If the OSP is satisfied it is *optimal* for an agent who waits at h to make an irrevocable decision next period. The proposition below shows that the OSP is satisfied at every information set that is reached with positive probability. The key observation in the proof is that $V(h, k) > 0 \Rightarrow \lambda(h, k) > 0$. Either it is strictly optimal to invest at (h, k) or else $V(h, k) \leq 0$ and no one chooses to invest. In the latter case, because no information is revealed, this becomes an absorbing state and no one ever invests.

Proposition 11. (Chamley and Gale, 1994). *For any fixed but arbitrary symmetric PBE (λ, μ) the OSP holds at any information set $h \in E$.*

Proof. The proof is given in Chamley and Gale (1994). □

The OSP allows the characterization of the equilibrium path.⁹ Using the OSP Chamley and Gale show that:

$$\begin{aligned}
 W(\xi, h) &= \sum_{k=0}^{N-1} p(k | h) \max \{V(h, k), 0\} \\
 &= \sum_{k=0}^{N-1} \max \left\{ \sum_{n=0}^N b(k; n - K(h) - 1, \xi) \mu(n | h) v(n), 0 \right\} \quad (4.1)
 \end{aligned}$$

Where $p(k, n | h, \xi)$ is the distinguished agent's probability assessment of k investments and n options, given the history h , and $K(h)$ is the number of players who have invested already. Chamley and Gale show that $W(\xi, h)$ is continuous and increasing in ξ whenever $W(\xi, h) > V(h) > 0$ using the properties of the OSP and characterize the equilibrium path in the following proposition:

Proposition 12. (Chamley and Gale, 1994). *Let (λ, μ) be a fixed but arbitrary symmetric PBE After any history $h \in E$, one of the following mutually exclusive situations occurs:*

- (a) $V(h) < 0$ and $\lambda(h) = 0$;
- (b) $V(h) \geq \delta \sum_n \mu(n | h) \max \{V(h), 0\} > 0$ and $\lambda(h) = 1$;
- (c) Neither (a) nor (b) apply, in which case $0 < \lambda(h) < 1$ is the unique value such that $V(h) = \delta W(\lambda(h), h) = \delta W^*(h)$.

Proof. The proof is given in Chamley and Gale (1994). □

Proposition 12 characterizes the equilibrium path by dividing it into three cases. In case (a), sufficiently optimistic beliefs result in all agents immediately investing and the game ending. In case (b) beliefs are sufficiently pessimistic that no one is willing to invest,

⁹The OSP does not hold for asymmetric equilibria and as such, they are much more difficult to study. For this reason Chamley and Gale concentrate almost entirely on the symmetric case.

and with the revelation of no new information the game will also end immediately. In case (c) there are intermediate beliefs resulting in indifference between investing and waiting and so randomization occurs. The proposition can also be used to define the equilibrium path recursively. Given any history $h \in E$, define the probability assessment $\mu(h)$ using Bayes' theorem and then calculate the value of $V(h)$. We can then use the conditions of proposition 12 to define $\lambda(h)$ uniquely which leads us to the following theorem:

Theorem 3. (Chamley and Gale, 1994). *There exists a unique symmetric PBE for the game described above.*

Examining the nature of the equilibrium we say that a symmetric PBE exhibits delay if agents who have a positive payoff from investment choose not to invest, i.e. if $V(h) > 0$ and $\lambda(h) < 1$, for some history h that is reached with positive probability. Define δ_1 implicitly by:

$$V(\emptyset) = \delta_1 W(1, \emptyset) \equiv \delta_1 \sum_N g(n) \max\{v(n), 0\} \quad (4.2)$$

Proposition 13. (Chamley and Gale, 1994). *Any symmetric PBE must exhibit delay if $\delta > \delta_1$. No equilibrium will exhibit delay if $\delta < \delta_1$.*

Proof. The proof is immediate from proposition 12 and the observation that $V(\emptyset) < \delta W(1, \emptyset)$ if and only if $\delta > \delta_1$. \square

The length of a time period measures the reaction lag between the release of information at one date and its incorporation in decisions at the next. For example, this lag may be substantial if it takes *time to build* and decision-makers are responding to physical evidence of investment decisions. Letting γ denote the length of time period and ρ the rate of time preference, the discount factor δ is equal to $e^{-\rho\gamma}$. Let γ_1 be the period length corresponding to δ_1 . Proposition 13 then implies that there will be delay if $0 < \gamma < \gamma_1$ but not if $\gamma > \gamma_1$, so lengthening the period can reduce delay. However, as the period

length γ becomes very short, the delay also becomes very short, as the proposition below shows.

Proposition 14. (*Chamley and Gale, 1994*). *In any symmetric PBE all investment ends after at most N periods. Thus, as $\gamma \rightarrow 0$, the length of the game converges to zero.*

Proof. This is proved in Chamley and Gale (1994). □

Proposition 14 identifies conditions under which the game ends quickly.¹⁰ Agents make their decision either to invest or else never to invest within an arbitrarily short time after the start of the game.¹¹

Despite the quick end to the game, the outcome is not efficient. In fact, the information revealed must be imperfect or the first movers would be unwilling to invest. The imperfection takes the form of a collapse of investment: with positive probability agents stop investing, even though the true return to investment is positive. Thus, when the period length is vanishingly small, the model produces something like the herd behaviour of Banerjee (1992) or the informational cascades of Bikchandani, Hirshleifer, and Welch (1992) analysed in chapter 2. As the period increases towards γ_1 , the probability of herd behaviour and informational cascades decreases towards zero. In the Chamley and Gale model there are two major conclusions: firstly, there is likely to be inefficiency in the form of a positive probability of a failure to invest when there is a positive true value to investment; secondly, endogenous timing plays a role in determining the size of this positive probability.

4.4 CHOOSING FROM A CONTINUUM

Gul and Lundholm (1995) present a model which seeks to analyse similar forces to those at work in the Chamley and Gale (1994) model. However, they present an alternative

¹⁰This proposition only holds with finitely many agents. With an infinite number of agents Chamley and Gale (1994) show that a modified proposition and proof can still be found.

¹¹Proposition 14 does not give conditions for *no delay* since a decision never to invest might be seen as enforcing infinite delay. It is the process of investment and information revelation which ends almost instantaneously when the period length becomes vanishingly short.

approach with different conclusions. Each agent is interested in predicting the future value of a project, denoted by the realization of a random variable W and, with the accuracy of the prediction held constant, would prefer to make their prediction sooner rather than later. Each agent has information about the realization of W ; in particular, $W = \tilde{s}_1 + \tilde{s}_2$, and agent $i \in \{1, 2\}$ observes the realization s_i . For simplicity assume that the \tilde{s}_i 's are independent and have a uniform distribution on the interval $I = [0, 1]$. Denote agent i 's prediction by z_i and the time of the prediction by t_i . Each agent makes only one prediction and the second agent observes the first agent's prediction. Agent i 's utility is given by:

$$u(w, z, t) = -(w - z_i)^2 - \alpha w t_i \quad (4.3)$$

The utility function trades off the cost of an error in the agent's prediction (the first term) against the cost of delaying the prediction (the second term). The constant $\alpha > 0$ measures the relative sensitivity of utility to accuracy versus delay. The only reason an agent has to wait is in order to observe the other agent's prediction. The delay cost is increasing in the realized w , capturing the idea that there is more urgency in forecasting more valuable projects. There is no relative performance term in the utility function, therefore both could reveal their information and achieve the highest utility. Gul and Lundholm do not allow pre-play communication and use the utility form in equation 4.3 to keep the analysis as simple as possible. The variable z_i can be thought of as an actual prediction or as a more tangible action choice, such as the size of an initial investment in a new project with w as the optimal level of investment based on all available information. In general, all that is required of z_i is that it is a one-to-one function of the agent's expectation of W .

Denote the strategy profile of the two agents as $\sigma = (\sigma_1, \sigma_2)$. The optimal prediction for agent i will minimize the mean squared error of the forecast conditional on the agent's signal s_i . Once an agent makes a prediction the other agent will predict immediately

afterward. A strategy for agent i is fully described by a function $t_i : I \mapsto \mathbb{R}^+$, where $t_i(s_i)$ specifies the latest possible time at which agent i with signal s_i will make a prediction. As a terminological note, Gul and Lundholm refer to the agent predicting first as the *first* agent, regardless of whether that agent has signal s_1 or s_2 .

Much like Chamley and Gale (1994), Gul and Lundholm concentrate on the symmetric equilibrium in which $t_1(s) = t_2(s) = t(s)$ for all $s \in I$. In this equilibrium $t'(s) < 0$ for all s and $t(I) = 0$. The $t(\cdot)$ function cannot be increasing since it is more costly for agents with higher signal realizations to wait than it is for agents with lower signals, and the gain to waiting does not depend on signals. Furthermore, there can be no region in which $t(\cdot)$ is constant. If there were, an agent could wait an arbitrarily small amount of time and gain a strictly positive amount of additional information. Since no agent would be willing to wait the strictly positive amount of time represented by the discontinuity gain an infinitesimal amount of additional information, $t(\cdot)$ must be continuous. The second agent can infer the first agent's signal from the time made by the first agent's forecast because $t(\cdot)$ is invertible. The inverse of $t(\cdot)$ is denoted $s(\cdot)$. Since $t(\cdot)$ is downward sloping, if the game proceeds to time $\bar{\tau}$ without a forecast, each agent knows that the other agents's signal is not in the region $[s(\bar{\tau}), 1]$. Thus, if the first agent chooses to forecast at time $\bar{\tau}$, then his forecast is $s_i + \frac{1}{2}s(\bar{\tau})$. The equilibrium is detailed in the proposition below.¹²

Proposition 15. *(Gul and Lundholm, 1995). There exists a unique Nash equilibrium outcome for the game described above. In this equilibrium, agent i predicts $(3/2)s_i$ at time $t(s_i) = (1 - s_i)/6\alpha$ if his opponent has not made a prediction; otherwise he predicts $s_i + \frac{2}{3}z_j = s_1 + s_2$ immediately after his opponent's announcement (at time $t(\frac{2}{3}z_j)$). If agent i observes that $\tau_j \neq \frac{2}{3}z_j$, then agent i forms an arbitrary conjecture about the distribution of s_j and forecasts s_i plus the mean of s_j given his new conjecture.*

Proof. The proof is given in Gul and Lundholm (1995). □

¹²The arbitrary conjecture referred to in the proposition is admissible since the second agent's out-of-equilibrium beliefs do not affect the first agent's expected utility calculations and so there are a multiplicity of out-of-equilibrium conjectures associated with the unique symmetric equilibrium outcome.

The remarkable result in Gul and Lundholm (1995) concerns the fact that there are no informational cascades in the sense of definition 1, but agents' decisions are still clustered together. An agent's decision variable is chosen from a continuum, so the second agent always uses his own information to improve his decision. The economy in their model is very efficient in its use of information. Not only do both agents use their own information, but the second agent can recover the first agent's information by observing his forecast.

4.5 A CRITICAL APPRAISAL: ADDRESSING SOME CONCERNS

In Gul and Lundholm (1995) agents delay even though they make decisions in continuous time. However, the analogue in Chamley and Gale (1994) would be to let $\gamma \rightarrow 0$ and this eliminates delay in their model. Chamley and Gale stress that the Gul and Lundholm model involves the choice of an action from an interval, and consider this to be the explanation for the essential differences between the two models.¹³ In the Chamley and Gale (1994) model a positive option value is necessary for delay. The option value is only positive if an agent's decision depends in a non-trivial way on the information revealed. In any period investment must stop for some value of k , the number of players. The amount of information revealed in each period must be large in order to change an agent's decision from 0 to 1. However, when an agent chooses a continuous decision variable, he can respond to a small amount of new information by making a small change in the variable. Even small amounts of new information have a positive option value, so it is possible to keep the process going with only a small amount of information being released each period. The process can continue like this for a long time, producing more and more delay.

The Chamley and Gale (1994) model can be used to analyse a fairly general class of problems and by letting $\gamma \rightarrow 0$ we can analyse a close approximation to continuous time. A comparison with the Gul and Lundholm (1995) model suggests that it is the implicit assumption of indivisibilities in the action set that produces the difference between the

¹³This comment is based on Gul and Lundholm (1993), but is applicable to Gul and Lundholm (1995).

two sets of results. In games in continuous time there are certain technical difficulties with terms like “immediately after”, ensuring strategies imply well-defined outcomes. Therefore, their continuous-time game should be considered to be the limit of a series of discrete-time games as the length of periods goes to zero. Gul and Lundholm (1993) show that there is a unique symmetric equilibrium outcome to the discrete time game and that the equilibrium outcome converges to the outcome given for the continuous-time game.

The Gul and Lundholm (1995) model shows that indivisibilities are crucial in generating informational cascades, and that the empirical regularity of some form of clustering cannot be taken to imply that information is being unused and inefficiency is present. Other empirical results may be needed to detect some form of inefficiency and Gul and Lundholm suggest that one sign of an informational cascade might be that the *ex post* accuracy of decisions does not improve with time, whereas in an efficient clustering decisions would become more accurate.

Given the nature of investment decisions, it seems reasonable to consider indivisibilities as an essential part of the decision-making process. With this point in mind, much of the work in chapters 5 and 6 concentrates on models which are closer to Chamley and Gale (1994) in spirit, where investment is a discrete binary variable, so a decision-maker either invests or delays.¹⁴ In this sense there is an assumption of some form of inefficiency before the analysis even begins; however it is still possible to examine the welfare-effects of other alterations to the class of models examined in this chapter.

One area we might wish to explore is building a rationale into the model for the irreversibility of decision-making which is so essential to Chamley and Gale (1994). The next chapter does this through the use of an alternative utility function, close in spirit to that used in Gul and Lundholm (1995). Chapter 5 will show how incorporating a mean square error term into the utility function can be seen as directly capturing the costs of reversing any decision.

¹⁴However, in chapter 5, a form of utility similar to that specified in Gul and Lundholm (1995) is used to capture the idea that there is an optimal level of sunk cost when deciding to invest.

A further area we might wish to explore (much as in chapter 3) is the scope for improving upon the *status quo* through intervention. Simply revealing the true state or forcing all agents to work together, and therefore reveal all of their useful information, will undoubtedly raise welfare. However, a more interesting question concerns the scope to improve welfare through more realistic forms of intervention. Both Chamley and Gale (1994) and Gul and Lundholm (1995) assume that co-operation is not an option because to allow co-operation removes all interest from the model. They justify this by suggesting that there are strategic reasons outside the model why such co-operation is not possible. Chapter 5 is therefore all about what a third party might be able to do (and that third party might represent some attempt by the players in the game to merge their information).

Finally, we might be interested in the role of a stochastic environment. The role of the discount factor is such that when sufficient information is available players will jump into a simultaneous herd, rather than sequential process, in which all players instantly take action. This is the very feature of endogenous-time models which makes them different from the models examined in the previous two chapters, and in a sense they automatically include a stochastic element; hence the role of the option value in finding an equilibrium. The next chapter demonstrates this by incorporating a simple Markov process into the model and shows the immediate effect of this inclusion in a Chamley and Gale-style model. A new term appears which adds a further cost to delay, and unlike in the myopic agent world of Moscarini, Ottaviani, and Smith (1996) and sequential herding models more generally, this will be appreciated by the players in an endogenous-time world, as it directly affects their utility.

This chapter and the next chapter are both vulnerable to the standard criticism that the decision-making assumed requires very high levels of computational ability. Chapter 6 addresses these concerns through the use of an experiment designed to show that for small monetary incentives most players will get very close to behaving optimally, and will certainly follow the central decision-making rule asserted in part I of this thesis: that rational agents will take into account the observation of others' relevant actions.

CHAPTER 5

MODELS OF ENDOGENOUS-TIME HERDING

5.1 INTRODUCTION

As was stressed in the first part of this thesis, when taking an economic decision an agent might well benefit from observing the choices made by others.¹⁵ Furthermore, as seen in the last chapter, an agent may be willing to delay his decision in order to gain more information inferable from the actions, or inaction, of others. The classic example comes in the theory of irreversible investment under uncertainty, where an investor may delay his decision in order to try to increase his understanding of the payoffs of different projects by observing the actions of other potential investors. As we shall see in this chapter this problem is as much about the herd externality as is the sequential herding problem of chapters 2 and 3.

In the broader context of asymmetric information there is also an interesting and recurring theme: it may be the case that the prospect of additional information becoming available can actually damage welfare. It is the main aim of this chapter to develop a model which can be used to see whether this is true in a social learning context, or at least to discover how valuable additional public information can be. In order to do this agents must be allowed to make decisions whenever they wish. The approach of this chapter is therefore to concentrate on a model in which agents can choose when to act, so

¹⁵This chapter is based in part on SgROI (1998), and was presented at the Fifth Young Economists' Conference, Amsterdam, April 1999.

the timing of decisions becomes endogenous. Finally, much as in chapter 3, this chapter details an attempt to come to terms with the herd externality rather than simply accept the welfare-damaging implications.

The analysis in this chapter suggests that decisions will usually be reached quickly, with some delay to account for social learning. Introducing the prospect of some future revelation of information will produce additional delay as agents wait to capitalize on new information. This may reduce the decision-makers' expected payoffs. If the gathering of public information involves even a small positive cost then such information can seem prohibitively expensive. So we see that the dangers of herding and the problem of delay cannot be easily mitigated by providing further information.

5.1.1 Overview

In section 5.2 a simple model will be constructed in which two agents must decide whether, and if so when, to invest in a project about which they have some private information. It will be shown that decisions will usually be reached very quickly but that the ability to wait in order to capitalize on new information will create some delay.

Section 5.3 develops the full model. While the full model loses some of the simplicity and clarity of simple model in the previous section, the various new additions are essential for the examination of several of the point raised in the last chapter. The first aim is to directly capture the problems inherent in a model of irreversible sunk cost investment via a simple reduced form utility function, which increases the importance of having access to accurate information. The second aim is to assess the scope for mitigating the herd externality through the provision of additional information, which is examined in the next section.

Section 5.4 starts with an examination of a specific problem: it is possible that in a common value multi-agent model of investment under uncertainty, worthwhile investments will not be undertaken because of information asymmetries between agents. *A priori* it seems plausible that welfare will unambiguously improve if a third party perfectly reveals the value of the investment at some point in the future. However, it will

be shown that, while this will eliminate the danger of worthwhile investments *not* being made, it will generate further delay which will partly offset this benefit. A necessary condition for the use of some form of public revelation is given which shows that if gathering extra information involves even a small positive cost, *ex ante* welfare may actually decrease if a policy of complete revelation is undertaken.

Section 5.5 offers some conclusions and suggests further applications for the class of models examined in the chapter.

5.2 INTRODUCING A SIMPLE MODEL

The model developed within this section builds on the methodology for introducing endogenously-timed decision making outlined in Gale (1996) which provides perhaps the simplest way to extend the herding problem into endogenous time.

5.2.1 Preliminaries

In general we will consider N agents, but initially we will restrict ourselves to $N = 2$. These agents have a decision problem which operates in two dimensions: whether to invest in a project, and if so when to invest. The return to this project is the state of the world, w , which is initially assumed fixed at the beginning of time. Time is indexed by $t \in \mathbb{N}^{++}$.¹⁶ Agents do not directly observe w , instead receiving a signal, μ , at $t = 1$. We use superscript to index agents and subscript to index time, so μ_t^i is the signal of agent $i \in \{1, 2\}$ at time t . We will use i and j to denote our two agents; usually i is the agent whose decision problem we are considering and j will be the other agent. The signals μ^i and μ^j are independent and identically drawn from the uniform distribution with support $[-1, 1]$, so $\mu^i \sim U[-1, 1]$ for $i \in \{1, 2\}$. These signals do not change over time, and the state of the world w is set equal to the sum of all signals, $w = \mu^i + \mu^j$.¹⁷

¹⁶We are therefore restricting ourselves to discrete time and following Chamley and Gale (1994) rather than Gul and Lundholm (1995).

¹⁷ $w = f(\mu^i, \mu^j)$ would be more general, but a simple additive function will keep the results as simple as possible.

Actions are defined as: $x^i = 1 \Leftrightarrow$ “invest”; and $x^i = 0 \Leftrightarrow$ “do not invest”. An agent can observe his own signal, but not the signal of the other agent. In each period actions are made simultaneously, so the two agents cannot observe each others’ actions. However, in period 2, the agent will know the action that the other agent performed in period 1, and through the observed choice of action some information about the nature of the other agent’s signal may be revealed. While there is no explicit reason why the agents could not meet and reveal their signals, pre-play communication will not be allowed. In reality there may be issues of relative performance or a zero-sum element to the payoffs; by excluding pre-play communication we can simplify the utility function and the analysis without explicitly considering a relative performance measure. Finally we have payoffs, π_t^i , where $t \in \mathbb{N}^{++}$ and $i \in \{1, 2\}$, discounted strictly by $\delta \in (0, 1)$:

$$\pi_t^i = \begin{cases} \delta^{t-1}w & \text{if } x^i = 1 \\ 0 & \text{if } x^i = 0 \end{cases}$$

5.2.2 Solving the Decision Problem

Consider the problem faced by agent i : whether, and if so when, to invest. Myopically we could consider the following simple rules: (a) invest (i.e. $x_t^i = 1$) if and only if $E[\pi_t^i] > 0$; (b) if an investment is to be made, then make it at $t = 1$ if and only if $E[\pi_1^i] > E[\pi_2^i]$, if not then wait. In these rules the profit function explicitly includes discounting. This might seem a sensible rule to adopt, but although it captures a notion of the cost of delay since we have an implicit $\delta < 1$ in the second period payoff, it fails to capture the benefit of delay, namely the *option value* of waiting. This option value comes about because of the possibility that for some reason agent i may have invested at time 1 when doing so was foolish given the information available to him at time 2. We will consider the cost and benefit of delay in turn, but first we will define a symmetric signal value $\bar{\mu}$ such that $\mu^i > \bar{\mu} > 0 \Leftrightarrow x^i = 1$. We have not yet said anything about what to do at $t = 2$, but we have defined an alternative to decision rule (b) for $t = 1$: (b*) invest at $t = 1$ (i.e. set $x_1^i = 1$) if and only if $\mu^i > \bar{\mu} > 0$.

Proposition 16. *There is some symmetric $\bar{\mu}$ such that it is optimal for agent i to invest at time $t = 1$ if and only if $\mu^i > \bar{\mu} > 0$. Furthermore this value can be roughly approximated by the linear function $\bar{\mu} = \frac{1}{3}\delta$ over the relevant range of values of δ .*

Proof. See appendix. □

Proposition 17. *(i) The game will end by $t = 2$, i.e. if agent i did not invest at time $t = 1$ he will either invest when $t = 2$ or never invest. (ii) Agent i will only invest at $t = 2$ if agent j invested at $t = 1$.*

Proof. See appendix. □

5.2.3 Results

We have now specified a simple example of an endogenous-timing model. The next section presents the full model, which adds a more complex form of utility, incorporating concerns related to the irreversibility of decision-making, and a stochastic state. However, as we shall see the simple model captures a great deal of the intuition that arises when the timing of decisions becomes endogenous.

We will from now on refer to the value of $\bar{\mu}$ found in this section as $\bar{\mu}(\delta)$ since it is a function of δ only. This will enable us to distinguish the signal value found here in the simplest case to the more complex function derived later. There are numerous features of this model which are very much in keeping with the herding literature: information is not fully revealed; there is no direct mapping from signal to action which can be inverted to reveal agents' signals; errors are made and private information may be ignored, in particular even if $\mu^i > 0$ for $i = 1, 2$ neither will invest unless $\mu^i > \bar{\mu}(\delta)$ for at least some i ; and the errors which lead to incorrect decisions in turn lead to welfare losses, even though there is minimal delay in this model. It has also been shown that the game will effectively end at $t = 2$, as beyond this point agents have either invested or will never do so. The addition of further agents would allow the game to continue beyond two periods of interest, but we need at least one agent to invest in a period or investment will stop,

as in the two agent case. This is formally shown to be true in the statement and proof of proposition 18 which extends proposition 17 to the multi-agent case.

Proposition 18. *A single period of no investment will end the prospect of any further investment in a model with $N \in \mathbb{N}^{++}$ agents.*

Proof. See appendix. □

Gale (1996) provides an intuition for results of this kind, pointing out that in a model of this type there must be a possibility of investment collapse as a necessary condition of equilibrium. This comes about because in order to have any delay there must be a positive option value, and this in turn implies a positive probability that agents will never invest.

5.3 THE FULL MODEL: INCORPORATING IRREVERSIBILITY

Introducing accuracy and delay in this section allows the generalization of the form of utility examined in section 5.2 to highlight two concerns: the desire to predict accurately the value of the project, w , and the desire to minimize delay. We will do this by altering the utility function of agent i so that it is similar in form to the utility function used by Gul and Lundholm (1995). It is shown that making this change increases the value of delay, though most of the results of the previous section still apply with only minor modifications. In particular the various propositions require a weakly higher threshold signal value, but qualitatively remain similar. Consider the following scenario:

(1) An agent $i \in \{1, 2\}$ is faced with a decision about whether to make an investment, $x_i^i = \{0, 1\}$, and if so, when to make it, $t \in \mathbb{N}^{++}$.

(2) Agents each receive a signal about the true state, μ^i , drawn from an independent and identically distributed uniform $[-1, 1]$ distribution. We assume initially that $w = \sum_i \mu^i = \mu^1 + \mu^2$. In general the valuation w is a function of time but we will assume otherwise in this section as we did in section 5.2, though w is allowed to change in time in section 5.4. The value of the project is the same for both agents. Neither know their rivals' expected valuations, or equivalently neither know the value of w .

(3) Having invested the return is equal to the value of the project to the agent, about which there is uncertainty. Suitably discounted, $\pi_t^i = \delta^{t-1} E(w_t)$.

(4) An agent pays a sunk-cost to invest which effectively makes the decision to invest possible only once: *the investment is irreversible*. This sunk-cost is not explicitly modelled but we are modelling it implicitly. The optimal value of the sunk-cost is such that it is lower if the true value of the world is estimated closely, so it may represent the building of a factory or setting up of machines, or the hiring of staff. A good guess about the scale of the project and its value will enable a best choice of sunk-cost. Too high a valuation might lead to an excessive sunk-cost (too big a factory or too many people hired) and too low a valuation might lead to an insubstantial sunk-cost (the factory may need to be extended or new machines bought later after a period of capacity under-utilization). This sunk-cost problem is modelled in reduced form by including an additional component in the payoff function: a desire to minimize the mean squared error (henceforth MSE) in prediction. $E[w_t - J_t^i E(w_t | J_t^i)]$ is the information about the state of the world at time t available to agent i at time t (just before the decision is made). In another sense we can implicitly consider w to include the optimal fixed cost - and we model the difference from optimum as the MSE term. The constant $\alpha \in [0, 1]$ indexes the importance of this effect.¹⁸ The MSE term will also be discounted as part of the net payoff.

The scenario provides a justification for the irreversible nature of investment and the choice of utility function given in expression 5.1 below. Consider a utility function in which a desire to estimate correctly the value of the investment opportunity, w , appears directly in the utility function via a MSE term. The MSE term also captures indirectly the idea that individuals are concerned not only with the expected value of a project, but also the variance. If the outcome is easily predicted it is in some sense better for the potential investors than where they are more likely to make an error, even if on average the expected outcome is the same. This form of utility might also be seen as a

¹⁸If α were too high then the certain payoff of zero from no investment would dominate any uncertain payment involving a potentially very high accuracy penalty. Bounding the α variable to be in the interval $[0, 1]$ removes this possibility and makes the decision a non-trivial one. The bound also seems reasonable given the rationale behind the form of utility.

reduced-form attempt to capture risk aversion.

$$\pi_t^i = \begin{cases} \delta^{t-1} \left\{ w - \alpha [w - E(w | J_t^i)]^2 \right\} & \text{if } x_t^i = 1 \\ \pi_t^i = 0 & \text{if } x_t^i = 0 \end{cases} \quad (5.1)$$

Where $\delta \in (0, 1)$, $\alpha \in [0, 1]$, $t \in \mathbb{N}^+$, and $i \in \{1, 2\}$. The expectation of w is conditioned on the information set of agent i at time t denoted J_t^i . This information set will certainly contain the agent's own signal, so for agent i that is μ^i , but it will also contain the observed history of agent j 's actions, so $J_\tau^i = \{\mu^i, h_\tau^j\}$ where $h_1^j = \emptyset$, but for $t = 2$ and onwards h_t^j is the observed history of agent j 's actions from the beginning of the game until time $t = \tau - 1$.

To complete the model, consider the case when the state $\{w_t, t \in \mathbb{N}^{++}\}$ follows a general discrete time Markov process defined by:

$$w_t = \sum_{n=1}^N \mu^n + \varepsilon_t = \mu^i + \mu^j + \varepsilon_t \quad (5.2)$$

Where $E[\varepsilon_t | J_t] = E[\varepsilon_t] = 0$, $\text{Var}[\varepsilon_t] = \text{Var}[\varepsilon] = \eta$ for all $t \in \mathbb{N}^{++}$, and $w_1 = \mu^i + \mu^j$ (or equivalently $\varepsilon_1 = 0$). The error term in equation 5.2 is assumed to be independent of the signal drawings, i.e. $E[\mu^i \varepsilon_t] = E[\mu^j \varepsilon_t] = 0$ for all $t \in \mathbb{N}^+$. It is further assumed that μ^i is independent of μ^j , and that the error term is not correlated with past values of itself, so $E[\varepsilon_t | \varepsilon_\tau] = E[\varepsilon_t] = 0$ for all $t \neq \tau \in \mathbb{N}^+$. The drawings of the μ 's will continue to be from a uniform $[-1, 1]$ distribution.

5.3.1 Solving the Model

It is immediate from equation 5.2, the linearity of expectation and the definition of a Markov process that:

$$E[w_t | J_t] = E\left[\sum_{n=1}^N \mu^n + \varepsilon_t | J_t\right] = \mu^i + E[\mu^j | J_t] \quad \text{for all } t \in \mathbb{N}^{++} \quad (5.3)$$

Consider the decision problem of agent i where $\{w_t, t \in \mathbb{N}^{++}\}$ follows the stochastic process described in equation 5.2. As in the last sections we know that by time $t = 2$ the agent will have made a final decision about whether to invest at time $t = 2$ or whether never to invest. Now we can once again search for the value of the threshold signal $\bar{\mu}$ for this case. The expected profit at time $t = 1$ is given by $E[\pi_1^i] = \mu^i - \frac{1}{3}\alpha$. With probability $\frac{1}{2}(1 - \bar{\mu})$ it is the case that $\mu^j > \bar{\mu}$. When this is true (including discounting) we can derive $E[\pi_2^i | \mu^j > \bar{\mu}]$, noting the appearance of a $-\alpha\delta\eta$ term, the cross terms involving ε_2 disappearing because of the properties of ε_t :

$$E[\pi_2^i | \mu^j > \bar{\mu}] = \delta\mu^i + \frac{1}{2}\delta(\bar{\mu} + 1) - \frac{1}{4}\alpha\delta(\bar{\mu} + 1)^2 - \alpha\delta\left[\frac{\frac{1}{6}\bar{\mu}^3 + \frac{1}{2}\bar{\mu} - \frac{1}{2}\bar{\mu}^2 - \frac{1}{6}}{1 - \bar{\mu}}\right] - \alpha\delta\eta$$

With probability $\frac{1}{2}(1 + \bar{\mu})$ it is the case that $\mu^j < \bar{\mu}$.

$$E[\pi_2^i | \mu^j < \bar{\mu}] = \delta\mu^i + \frac{1}{2}\delta(\bar{\mu} - 1) - \frac{1}{4}\alpha\delta(\bar{\mu} - 1)^2 - \alpha\delta\left[\frac{-\frac{1}{6}\bar{\mu}^3 + \frac{1}{2}\bar{\mu}^2 + \frac{1}{2}\bar{\mu} - \frac{1}{6}}{1 + \bar{\mu}}\right] - \alpha\delta\eta$$

Calculating the unconditional probability:

$$E[\pi_2^i] = \frac{1 - \bar{\mu}}{2}E[\pi_2^i | \mu^j > \bar{\mu}] + \frac{1 + \bar{\mu}}{2}E[\pi_2^i | \mu^j < \bar{\mu}] = \delta\left[\mu^i - \alpha\left(\frac{1}{12} + \frac{1}{2}\bar{\mu} - \frac{1}{4}\bar{\mu}^2 + \eta\right)\right]$$

We can now calculate the cost of delay:

$$\begin{aligned}
E[\pi_1^i] - E[\pi_2^i] &= (\mu^i - \frac{\alpha}{3}) - \delta \left[\mu^i - \alpha \left(\frac{1}{12} + \frac{\bar{\mu}}{2} - \frac{\bar{\mu}^2}{4} + \eta \right) \right] \\
&= (1 - \delta) \mu^i - \alpha \left[\frac{1}{3} - \delta \left(\frac{1}{12} + \frac{\bar{\mu}}{2} - \frac{\bar{\mu}^2}{4} + \eta \right) \right]
\end{aligned}$$

The option value of delay is $-\delta \Pr[\mu^j < \bar{\mu}] \{\mu^i + E[\mu^j \mid \mu^j < \bar{\mu}]\}$. That is the expected loss avoided by agent i by not investing at $t = 1$ in the event that agent j does not invest at $t = 1$. Setting the cost of delay as equal to the benefit (option value) of delay and noting that at indifference $\mu^i = \bar{\mu}$ yields the value matching condition which is a quadratic in $\bar{\mu}$:

$$\begin{aligned}
(1 - \delta) \bar{\mu} - \alpha \left[\frac{1}{3} - \delta \left(\frac{1}{12} + \frac{\bar{\mu}}{2} - \frac{\bar{\mu}^2}{4} + \eta \right) \right] &= -\frac{\delta(1+\bar{\mu})(3\bar{\mu}-1)}{4} \\
\Rightarrow \frac{\delta(3-\alpha)\bar{\mu}^2}{4} + \frac{(2-\delta+\alpha\delta)\bar{\mu}}{2} - \frac{4\alpha+3\delta-\alpha\delta-12\alpha\delta\eta}{12} &= 0
\end{aligned}$$

We will denote this value $\bar{\mu}(\alpha, \delta, \eta)$. Ruling out the negative root since we require $\bar{\mu} \in [-1, 1]$ and taking values of $\alpha \in [0, 1]$:

$$\bar{\mu}(\alpha, \delta, \eta) = \frac{1}{\delta(3-\alpha)} \left\{ \delta - \alpha\delta - 2 + \left[(2 - \delta + \alpha\delta)^2 + \frac{\delta(3-\alpha)(4\alpha+3\delta-\alpha\delta-12\alpha\delta\eta)}{3} \right]^{\frac{1}{2}} \right\} \quad (5.4)$$

Setting $\eta = \alpha = 0$ returns the value of $\bar{\mu}(\delta) = \bar{\mu}(0, \delta, 0)$ in section 5.2 with no MSE term or stochastic state. In this sense the utility function for the full model generalizes that used in section 5.2. Examining on the weight on accuracy, α , as we might have expected $\frac{\partial \bar{\mu}(\alpha, \delta, \eta)}{\partial \alpha} > 0$ and $\frac{\partial \bar{\mu}(\alpha, \delta, \eta)}{\partial \delta} > 0$, therefore a higher weight on accuracy (or

equivalently, greater patience) makes delay more likely. The threshold value $\bar{\mu}(\alpha, \delta, \eta)$ is decreasing in η which is intuitive: a more uncertain future (a higher variance error term) would imply that waiting to observe the other agent's signal is less revealing and therefore less valuable, which will increase the likelihood of investment at time $t = 1$.

5.3.2 Results

Start by setting $\eta = 0$. Now since all that has been done is to add a term to the utility function, propositions 16, 17 and 18 apply unchanged except that the threshold value is now $\bar{\mu}(\alpha, \delta, 0)$, where for $\alpha \in [0, 1]$ we have that $\bar{\mu}(\alpha, \delta, 0) > \bar{\mu}(\delta) > 0$ since delay is more likely with a positive MSE term. Furthermore the only real effect of the stochastic state is to add a new term, η , which produces more concern about the future and therefore if positive will change the threshold value. Therefore the proofs of propositions 19, 20 and 21 are analogous to the proofs of propositions 16, 17 and 18 respectively, with $\bar{\mu}(\delta)$ replaced by $\bar{\mu}(\alpha, \delta, \eta)$ as given in equation 5.4.

Proposition 19. *For values of $\alpha \in [0, 1]$ there exists some $\bar{\mu}(\alpha, \delta, \eta) \in [-1, 1]$ such that it is optimal for agent i to invest at time $t = 1$ if and only if $\mu^i > \bar{\mu}(\alpha, \delta, \eta) > 0$.*

We have $\bar{\mu}(\alpha, \delta, \eta) > 0$ since there is a positive option value to delay *and* a benefit in terms of increased information about the other agent's signal. The unique value of $\bar{\mu}(\alpha, \delta, \eta)$ is now a function of $\alpha \in [0, 1]$ and the variance term η as well as $\delta \in (0, 1)$ and is given in equation 5.4.

Proposition 20. *(i) The game will end at $t = 2$, i.e. if agent i did not invest at time $t = 1$ he will either invest when $t = 2$ or never invest. (ii) Agent i will only invest at $t = 2$ if agent j invested at $t = 1$.*

Proposition 21. *A single period of no investment will end the prospect of any further investment in a model with $N \in \mathbb{N}^{++}$ agents.*

5.4 WILL INFORMATION REVELATION HELP?

In this section the model developed in the previous section is used to assess the impact of the herd externality as defined in chapter 2, definition 2, in a framework incorporating endogenous timing. We see that the herd externality does indeed feature, as once again the early decision-maker will reveal only his action. As seen in chapter 4, in a discrete time world the time of action cannot be used to provide a one-to-one mapping between signal and action where signals are continuous, and we would expect inefficiencies to occur.

The main result in this section can also be viewed as a different way of focusing on a classic problem in multi-agent investment under uncertainty with common values. It is the case that if an agent believes a project is probably worthwhile, but is not strong in his beliefs, he is likely to wait in order to try to accumulate more information from the observed actions of others. If the other agents are in a similar position then there may be no investment taking place while all wait to gain new information. The agents are then caught in a trap: the only new information they have is to lower their expectation of the state of the world conditional upon not observing any others deciding to invest. The net result is that they become even more unlikely to invest and investment breakdown occurs, although the sum of all signals (and even all individual signals) may suggest that the value of the project is strictly positive. *This problem is the exact analogue of the herd externality in a sequential model.*

Much like chapter 3, this section will examine the scope for the provision of further information to aid the decision-makers. The information as modelled here may be a joint effort by agents to survey conditions, or may be released by a regulator or government body. If a failure to invest has occurred when the project is actually worthwhile, then information revelation will increase joint profits *ex post*. However, using the model developed in this section it is shown that *ex ante* the agents may be better off in terms of joint payoffs if no revelation occurs. The rationale behind this result stems from the extra delay which common knowledge of future revelation induces, through a rise in the

threshold signal value above which investment takes place, and the failure of the extra information to be of sufficient use to offset even a small cost of information gathering. This effect is only possible in a multi-agent model in which strategic effects come into play, and would not be a feature of a single agent model in which additional information is always weakly beneficial (if it is not misleading) since it can always be ignored. A necessary condition for additional information to be useful is given and it is shown to be difficult to satisfy when there exists a cost of information gathering.

5.4.1 Complete Revelation of the True State

We now examine the role of information revelation in the decision process. In particular, consider the possibility of *complete revelation* of the true state.

Definition 7. *Complete revelation of the true state of the world $\{w_t, t \in \mathbb{N}^{++}\}$ at some pre-determined point in the future $t = \tau^*$, where $\tau^* \in \mathbb{N}^{++}$ is common knowledge to all agents, is said to occur when the true value of the state of the world w_{τ^*} at time $t = \tau^*$ is revealed to all agents.*

Assume that there exists a benevolent third party who can observe the true state of the world at certain points in time. He can reveal this information to the two agents, which will effectively eliminate the MSE component in their payoff functions.¹⁹ Continue to assume that $\{w_t, t \in \mathbb{N}^{++}\}$ follows the stochastic process as outlined in expression 5.2. Now consider a three period model: the first two periods are as in the previous sections. However, in the third period we allow the agents to know the true state of the world. Therefore the MSE term drops out and for $\delta \in (0, 1)$ and $i \in \{1, 2\}$ the agents are left with:

$$\pi_3^i = \begin{cases} \delta^2 w_3 & \text{if } x_3^i = 1 \\ 0 & \text{if } x_3^i = 0 \end{cases}$$

¹⁹The definition of complete revelation implies that $\{w_{\tau^*} - E[w_{\tau^*}]\}^2 = 0$.

Proposition 22. *If $\tau^* > 2$ then the game will end at time $t = \tau^*$ with a decision to invest or never invest, where τ^* is the time of complete revelation. In particular, if $\tau^* = 3$, then the game will end in period 3 with a decision to invest or never invest.*

Proof. See appendix. □

Agent i has three potential periods in which it might be optimal to invest. At time $t = 1$ when the agent first considers this problem his expected payoff at time $t = 3$ will be:

$$E [\pi_3^i | J_3^i] = \delta^2 E [w_3 | J_1^i] = \delta^2 \mu^i \quad (5.5)$$

The natural way to examine the decision problem is via a backward induction or dynamic programming approach. We will consider what the agent would do in period 3, assuming he is at period 3, then examine decisions in period 2 in the light of actions in period 3. Finally, we will look at period 1 having considered the optimal decision in period 2. This is made feasible by the simple observation that having reached period 3 the agent's best decision is to invest if and only if $\pi_3^i = \delta^2 w_3 > 0 \Rightarrow w_3 > 0$. Therefore we can disregard period 4 and onwards. Agent i then knows that his worst possible payoff is $\max \{\delta^2 w_3, 0\}$ and the expectation at time $t = 1$ of this is given by equation 5.5 to be simply $E [\pi_3^i] = \max \{\delta^2 \mu^i, 0\}$.²⁰ We will define the threshold signal value used here as $\hat{\mu}$ to differentiate it from the previous three signal values. $\hat{\mu}$ is a function of α, δ and η . It is also generally a function of the time of revelation τ^* , but here we have set $\tau^* = 3$. So generally we have $\hat{\mu} = \hat{\mu}(\alpha, \delta, \eta, \tau^*)$. In period 2 the agent will invest if and only if:

$$E [\pi_2^i | J_2^i] - \max \{\delta^2 \mu^i, 0\} > -\delta \frac{1}{4} (1 + \hat{\mu}) (3\hat{\mu} - 1)$$

²⁰Here and throughout this chapter payoffs will always be discounted back to time $t = 1$, so the $t = 3$ payoff is expressed as the discounted value at $t = 1$. When comparing time $t = 2$ and $t = 3$ the payoff at time $t = 3$ should only be discounted once by δ to return a payoff discounted back to $t = 2$ but by discounting all payoffs to $t = 1$ there is greater consistency throughout.

In period one agent i knows that if he delays beyond $t = 1$:

$$E[\pi_2^i] = \max \{ E[\pi_2^i | J_2^i], \max \{ \delta^2 \mu^i, 0 \} \}$$

Therefore:

$$x_1^i = 1 \Leftrightarrow E[\pi_1^i | J_1^i] - \max \{ E[\pi_2^i | J_2^i], \max \{ \delta^2 \mu^i, 0 \} \} > -\frac{1}{4} \delta (1 + \hat{\mu}) (3\hat{\mu} - 1)$$

We can determine $E[\pi_2^i | J_2^i]$ as in the previous section by considering the two possibilities of $x_1^j = 1$ or $x_1^j = 0$ and the probabilities of each occurring, the rule becomes:

$$x_1^i = 1 \Leftrightarrow E[\pi_1^i | \mu^i] - \max \left\{ \frac{(1-\hat{\mu})E[\pi_2^i | \mu^j > \hat{\mu}]}{2} + \frac{(1+\hat{\mu})E[\pi_2^i | \mu^j < \hat{\mu}]}{2}, \max \{ \delta^2 \mu^i, 0 \} \right\} \\ > -\frac{\delta(1+\hat{\mu})(3\hat{\mu}-1)}{4}$$

$$\Leftrightarrow \mu^i - \frac{1}{3}\alpha - \max \left\{ \delta \mu^i - \frac{\alpha\delta}{12} - \frac{\alpha\delta\mu^i}{2} + \frac{\alpha\delta\mu^{i2}}{4} - \alpha\delta\eta, \max \{ \delta^2 \mu^i, 0 \} \right\} > -\frac{\delta(1+\hat{\mu})(3\hat{\mu}-1)}{4}$$

We can once again find the threshold value by setting $\mu^i = \hat{\mu}$ and equating the cost of delay to the option value of delay.

$$\max \left\{ \frac{\delta(3-\alpha)\hat{\mu}^2}{4} + \frac{(2-\delta+\alpha\delta)\hat{\mu}}{2} - \frac{4\alpha+3\delta-\alpha\delta-12\alpha\delta\eta}{12}, \frac{3\delta\hat{\mu}^2}{4} + \frac{(2+\delta-2\delta^2)\hat{\mu}}{2} - \frac{4\alpha+3\delta}{12} \right\} = 0$$

To differentiate the two cases we will use a subscript 1 for the first case where:

$$\frac{\delta(3-\alpha)\hat{\mu}^2}{4} + \frac{(2-\delta+\alpha\delta)\hat{\mu}}{2} - \frac{4\alpha+3\delta-\alpha\delta-12\alpha\delta\eta}{12} > \frac{3\delta\hat{\mu}^2}{4} + \frac{(2+\delta-2\delta^2)\hat{\mu}}{2} - \frac{4\alpha+3\delta}{12}$$

$$\Leftrightarrow \left(\frac{\alpha}{2} + \delta\right) \hat{\mu} + \left(\frac{1}{12} + \eta\right) \alpha > \frac{\alpha \hat{\mu}^2}{4} + \hat{\mu} \quad (5.6)$$

The subscript 2 will be used to denote the second case where the reverse inequality holds. Consider the two cases in turn:

$$\frac{1}{4} \delta (3 - \alpha) \hat{\mu}_1^2 + \frac{1}{2} (2 - \delta + \alpha \delta) \hat{\mu}_1 - \frac{1}{12} (4\alpha + 3\delta - \alpha \delta - 12\alpha \delta \eta) = 0 \quad (5.7)$$

The quadratic in $\hat{\mu}$ can be solved in the usual way (ruling out the unfeasible root) to give:

$$\hat{\mu}_1 = \frac{1}{\delta(3-\alpha)} \left\{ \delta - \alpha \delta - 2 + \left[(2 - \delta + \alpha \delta)^2 + \frac{\delta(3-\alpha)(4\alpha+3\delta-\alpha\delta-12\alpha\delta\eta)}{3} \right]^{\frac{1}{2}} \right\} \quad (5.8)$$

Equation 5.8 is in fact the expression already calculated in equation 5.4, therefore if this condition holds, complete revelation makes no effective difference to the decision-making of the two agents. The alternative case is:

$$\frac{3}{4} \delta \hat{\mu}_2^2 + \frac{1}{2} (2 + \delta - 2\delta^2) \hat{\mu}_2 - \frac{1}{12} (4\alpha + 3\delta) = 0$$

Solving this and ruling out the negative root leaves:

$$\hat{\mu}_2 = \frac{2}{3} \delta - \frac{1}{3} + \frac{2}{3} \delta^{-1} \left\{ [1 + \delta - \delta^2 - \delta^3 + \delta^4 - \alpha \delta]^{\frac{1}{2}} - 1 \right\}$$

The addition of the possibility of complete revelation of the true state at time $t = 3$ may have absolutely no effect if expression 5.6 holds, or it may result in a new threshold value if the condition does not hold. If expression 5.6 holds the possibility of complete revelation is effectively ignored. It is only when the condition holds and the threshold value is $\hat{\mu}_2$, that complete revelation affects profits.

5.4.2 Measuring Welfare and Comparing Outcomes

The first major question to address is whether the prospect of complete revelation at some known point in the future will unambiguously increase expected welfare. There is no obvious best way to model welfare in this model since consumers are not explicitly considered. However, perhaps the most straightforward way to examine the value of complete revelation is to find the effect on joint profits in the case where complete revelation does have an effect, i.e. when the threshold signal value is $\hat{\mu}_2$ and to compare this with the reference case where there is no complete revelation, i.e. where the threshold signal value is $\hat{\mu}_1 = \bar{\mu}(\alpha, \delta, \eta)$.

Remark 3. *Complete revelation need not unambiguously increase expected joint payoffs, where the expectation is calculated before the social learning process begins, but after agents' observe their own signals.*

By looking at the time after agents have discovered their own signals but before social learning begins we can highlight the scope for extra information to be damaging for certain signal types, which could easily be missed by either averaging over all signal types or equivalently looking at true *ex ante* welfare (before signals are drawn). Nevertheless, *ex ante* welfare (pre-signal draws) is considered later in the chapter (in propositions 19 and 21 below). Remark 3 can be justified with reference to two effects produced when complete revelation is important: one with a positive impact on profits; the other with a negative impact. First, complete revelation removes the prospect of worthwhile investment not occurring, where worthwhile is simply taken to mean that the true value of the state is positive, so it precludes *full investment breakdown*.

Definition 8. *Full investment breakdown is said to occur when a project has positive value but it is not carried out by any agent because of problems of asymmetric information and uncertainty about the true value of the project.*

Close examination of this definition reveals that full investment breakdown is a form of informational cascade, capturing an endogenous-time analogue to the herd externality

as defined in definition 2. Assume that $\varepsilon_t = 0$, for all $t \in \mathbb{N}^{++}$. Consider a situation in which both agents have signal values below the threshold $\bar{\mu}$ and where $w = \mu^i + \mu^j > 0$. Neither agent would invest at $t = 1$, then having observed a period of no investment, they would never invest. The agents are effectively trapped in an informational cascade on the action “do not invest”, producing an investment breakdown. Complete revelation effectively bounds profits. When there is complete revelation expected profits for agent i must be weakly greater than $\delta^2 \mu^i$ which is a higher bound than in the case without complete revelation, where the only bound is that expected profits are weakly positive (since investment need not take place). We could call this the *profit bounding effect*.

However, the second effect weakly lowers expected joint profits, where the expectation is calculated after the agents’ signals have been revealed but before the social learning process has begun. In the case without complete revelation we may see investment taking place earlier than if the decision is delayed to $t = 3$, and if this proves worthwhile all will be better off. We could call this an *investment delay effect*, which renders the net effect ambiguous. Whether joint profits improve therefore depends on the signal values observed and it is possible that agents may be better off in a world without complete revelation.

Comparing $\hat{\mu}_2$ and $\hat{\mu}_1$ it is quickly apparent that while it is likely that $\hat{\mu}_2 > \hat{\mu}_1$ for the feasible range of signals we are interested in, when the discount rate is low the difference between $\hat{\mu}_1$ and $\hat{\mu}_2$ is not great. Intuitively this is the case since information about the future is not as important when the future is not highly valued. As the discount rate rises the threshold signal value when the information matters diverges strongly from the signal value when the information does not matter. If complete revelation is important and it makes delay more likely then actual profits may be lower via discounting.²¹

Remark 3 can be shown to be true with the help of two scenarios given in Figure 15 overleaf. In both cases the symmetric signal case is taken for simplicity, but example sce-

²¹Consider a situation in which an agent would have invested if there was no prospect of revelation, but decides against doing so now that revelation is due to occur at time $t = 3$. If the agent’s original desire to invest would have lead to positive profits, the change of heart will reduce profits through discounting. See example 1 on page 106.

narios can easily be constructed in which signal values are different and new alternatives such as non-Pareto improving joint profit increases can occur.

In scenario 1 the signal value to agent i is 0.2. Regardless of his rival's signal he will delay to $t = 2$. This is the case since his signal value is below the threshold value; he will wait because the option value is more than enough to outweigh the payoff difference between investment at $t = 1$ and $t = 2$. This all changes with complete revelation at $t = 3$. When complete revelation occurs agent i delays to $t = 3$ and faces a higher expected payoff. All of this occurs before social learning has begun: the expectations are taken before $t = 1$, but after agent i receives his own signal, i.e. during phase 0 of the timing schedule given in Figure 15. Clearly once he observes the other agent's actions social learning will begin and his expectations and optimal action will change. In this scenario the joint expected payoff is higher in the world with complete revelation and so both agents would prefer complete revelation.

In scenario 2 the rise in the threshold signal value produced by the prospect of complete revelation initiates an investment delay effect. Agents would prefer to be in the world without complete revelation, since their joint expected payoff is higher without the prospect of complete revelation if their signals lie within the range between $\hat{\mu}_1$ and $\hat{\mu}_2$. Without ever having observed any public information, complete revelation can be good or bad for expected joint profits, depending upon the private information which agents have before the start of the social learning process.

FIGURE 15: Expected Payoff Scenarios

- $t = 0$ Agents receive private signals
 $t = 1$ Agents are informed whether complete revelation will occur
 $t = 2$ Agents observe the actions of their rivals in the previous period
 $t = 3$ Complete revelation occurs if due, if not the game ends at $t = 2$

Non-agent, non-scenario specific parameters

Revelation?	α	δ	η	$\bar{\mu}$
No	0.50	0.50	0.25	0.22
Yes	0.50	0.50	0.25	0.27

Scenario 1: $\mu^i = 0.2, i = \{1, 2\}$

Payoffs expected at $t = 1$ for agents 1 and 2, and optimal actions

Revelation?	$E[\pi_1^i]$	$E[\pi_2^i]$	$E[\pi_3^i]$	Action	Joint Payoff
No	0.03	-0.01	N/A	$x_2^i = 1$	0.06
Yes	0.03	-0.02	0.05	$x_3^i = 1$	0.10

\Rightarrow expected joint profit is higher with revelation

Scenario 2: $\mu^i \in [0.23, 0.27], i = \{1, 2\}$

Payoffs expected at $t = 1$ for agents 1 and 2, and optimal actions

Revelation?	$E[\pi_1^i]$	$E[\pi_2^i]$	$E[\pi_3^i]$	Action	Joint Payoff
No	[0.06, 0.10]	[0.00, 0.02]	N/A	$x_1^i = 1$	[0.12, 0.20]
Yes	[0.06, 0.10]	[-0.01, 0.01]	[0.06, 0.07]	$x_3^i = 1$	[0.12, 0.14]

\Rightarrow expected joint profit is lower with revelation over the range $\mu^i \in [0.23, 0.27]$

Having shown that there exist examples where expected profit is both lower and higher for different parameter values, the remark is shown to be true. However, the examples indicate that the possibility of new information damaging welfare is very small. Perhaps a far greater concern is that the information will be of no real use, and it is this possibility which will be examined next. Complete revelation is of great use in a situation of full investment breakdown, but can also be useful in a situation of *partial investment breakdown*.

Definition 9. *Partial investment breakdown is said to occur when a project has positive value and one agent invests, but the other agent fails to invest because of problems of asymmetric information and uncertainty about the true value of the project.*

Consider the following example. A project may be worth a strictly positive amount, with $\mu^1 = 0.7$ and $\mu^2 = -0.65$. Agent 1 will invest at time $t = 1$ and agent 2 will delay. Even after observing agent 1's decision to invest, agent 2 will not invest at time $t = 2$ for a wide range of values of $\delta \in (0, 1)$ and $\alpha \in [0, 1]$. This example can be generalized, so for $i \neq j$ whenever $-\mu^i < \mu^j < -\frac{1}{2}(\bar{\mu} + 1)$ there will be a situation of partial breakdown with agent j failing to invest in a worthwhile project.²²

In this case complete revelation will be useful for agent j though not for agent i .

Proposition 23. *Assume $\varepsilon_t = 0$ for all $t \in \mathbb{N}^{++}$. Complete revelation will only be of any benefit in the fraction of cases given by:*

$$f(\bar{\mu}(\alpha, \delta)) \equiv f(\alpha, \delta) = \frac{2 - 6\bar{\mu} + 32\bar{\mu}^2 - 10\bar{\mu}^3 + \bar{\mu}^4}{50 - 20\bar{\mu} + 2\bar{\mu}^2}$$

Proof. See appendix. □

It should be noted that $\frac{\partial f(\alpha, \delta)}{\partial \alpha} > 0$ and $\frac{\partial f(\alpha, \delta)}{\partial \delta} > 0$, and since $\frac{\partial \bar{\mu}(\alpha, \delta)}{\partial \alpha} > 0$ and $\frac{\partial \bar{\mu}(\alpha, \delta)}{\partial \delta} > 0$, this implies that $\frac{\partial f(\bar{\mu})}{\partial \bar{\mu}} > 0$. Therefore $\max_{\alpha, \delta} f(\alpha, \delta) = f(1, 1) \approx 0.103$. With

²²This term can be derived in the following way. For a project to be worthwhile it must be the case that $\mu^i + \mu^j > 0$ which implies that $-\mu^i < \mu^j$. However, after observing that agent i has invested, agent j can calculate $E[\mu^i | x_1^i = 1] = \frac{1}{2}(\bar{\mu} + 1)$ since it is uniform with support $[\bar{\mu}, 1]$. In order for agent j to fail to invest after observing investment by agent i , it must be the case that $\frac{1}{2}(\bar{\mu} + 1) + \mu^j < 0$ which implies that $\mu^j < -\frac{1}{2}(\bar{\mu} + 1)$.

maximum patience and accuracy weighting complete revelation will only be of use for countering partial investment breakdown 10% of the time. For more reasonable discount factors and accuracy weights this will be much lower, for example $f(0.5, 0.5) \approx 0.059$. Proposition 24 gives a necessary condition for undertaking complete revelation when information gathering has a cost.

Proposition 24. *With a cost of gathering information $C_g > 0$ an ex ante (before signal values are realized) necessary condition for welfare-improving complete revelation when revelation occurs at $t = \tau^*$, is:*

$$C_g < \delta^{\tau^*} \left(\bar{\mu}^3 + \frac{(3+\bar{\mu})(1-\bar{\mu})^3}{2(5-\bar{\mu})^2} \right)$$

Proof. See appendix. □

It should be stressed that this is a necessary condition and a very weak one, based on maximum possible signal values throughout, and it will rarely be sufficient. This implies that the information will only be of any use in a small number of cases. Add to this the fact that when the information is of any use this is exactly when the value of the project is likely to be positive but small, and the total value of the complete revelation is seen to be low. To give some idea of the magnitudes involved consider the following example.

Example 1. *All approximations are to three significant figures. For $\alpha = \delta = 0.5$, $\tau^* = 3$ and $\varepsilon_t = 0$ for all $t \in \mathbb{N}^{++}$, we have $\bar{\mu}(\alpha, \delta) \approx 0.281$. This implies that $f(\alpha, \delta) \approx 0.098$, so by proposition 23 complete revelation is only useful in 9.8% of cases. Now using proposition 24 we have as a necessary condition that the cost of public information gathering C_g must be below 0.006. To put this into a reasonable metric, the maximum possible ex ante project value is 2, so a necessary condition for complete revelation in the case when $\alpha = \delta = 0.5$ is that the cost of information gathering not exceed 0.309% of the maximum value of a given project.*

Note that a low value of α or δ makes the necessary condition even stricter, and as $\tau^* \rightarrow \infty$ a necessary condition for complete revelation to be welfare-improving is that

it has zero cost. The necessary and sufficient condition with $\tau^* \rightarrow \infty$ would actually involve the need for negative costs, since $\kappa > 0$. It is intuitive that as $\tau^* \rightarrow \infty$ the closer is $\hat{\mu}_2 \rightarrow \hat{\mu}_1 = \bar{\mu}(\alpha, \delta, \eta)$, i.e. if complete revelation is known to occur a great many periods in the future it will have a lesser effect on the threshold value, with the effect disappearing completely in the limit. More formally, consider the case where expression 5.6 does not hold, i.e. the threshold signal is $\hat{\mu}_2$. When $t = \tau^*$ the general value matching indifference condition is:²³

$$\hat{\mu} - \frac{1}{3}\alpha - \max \left\{ \delta\hat{\mu} - \frac{1}{12}\alpha\delta - \frac{1}{2}\alpha\delta\hat{\mu} + \frac{1}{4}\alpha\delta\hat{\mu}^2 - \alpha\delta\eta, \delta^{\tau^*-1} E[w_t] \right\} = -\frac{1}{4}\delta(1 + \hat{\mu})(3\hat{\mu} - 1)$$

Now $\tau^* \rightarrow \infty \Rightarrow \tau^* - 1 \rightarrow \infty$. Therefore, since $\delta \in (0, 1)$, we have that $\lim_{\tau^*-1 \rightarrow \infty} \delta^{\tau^*-1} = 0$. Hence in the limit expression the indifference condition becomes:

$$\hat{\mu} - \frac{1}{3}\alpha - \max \left\{ \delta\hat{\mu} - \frac{1}{12}\alpha\delta - \frac{1}{2}\alpha\delta\hat{\mu} + \frac{1}{4}\alpha\delta\hat{\mu}^2 - \alpha\delta\eta, 0 \right\} = -\frac{1}{4}(3\delta\hat{\mu}^2 + 2\hat{\mu} - 1)$$

This is exactly the same expression as when there is no complete revelation. Therefore as $\tau^* \rightarrow \infty$ the threshold value with complete revelation $\hat{\mu}$ does indeed tend to the threshold signal value without complete revelation $\bar{\mu}(\alpha, \delta, \eta)$.

5.4.2.1 A Point about Welfare and Externalities

A final consideration which might lessen the importance of this result is the corollary to these findings: just as the gains from the extra information may be slight, so too the damage to welfare from the investment breakdown may not be great in these marginal cases. In effect the important message is that investment breakdown can occur in various ways and is a fairly reasonable concern, but it will occur at times when the projects are likely to be of marginal worth and these cases may be easy to spot. In cases of great

²³The general value matching indifference condition is a generalization of expression 5.7 which gives the indifference condition for the case when $\tau^* = 3$.

worth it is simply less likely that investment breakdown will occur. The main danger then are cases where social worth differs from the sum total of the two private gains - here it is very possible that investment breakdown may occur where a project is of marginal joint private worth but of considerable social worth - and here we might argue that extra information may be useful. But even here we should note one of the central messages of this chapter: that extra information may well slow down decision-making which could partly offset the gains. The inclusion of any positive gap between the private and social worth of any project will immediately strengthen the results in this chapter, worsening the impact of the investment breakdown problem.

5.4.3 Results

The main result in this section regards the prospect of complete revelation of the true state by some third party. Since there is a chance that some form of investment breakdown might occur, where agents find themselves locked into an informational cascade and fail to invest in a worthwhile project, this information could in theory be useful. Indeed, complete revelation provides a higher theoretical upper bound to profits than the zero upper bound provided by the ability to delay investment indefinitely, since it eliminates the prospect of full or partial investment breakdown. However, the prospect of complete revelation also seems likely to add more delay to decision-making and from remark 3 this is capable of actually damaging joint payoffs. More likely still is the prospect that complete revelation will fail to be of any use in enough cases, and through propositions 23 and 24 in general the profit bound provided by complete revelation may not be worth a great deal. Complete revelation becomes less useful as the point of revelation is pushed forward in time ($\tau^* \rightarrow \infty$), as agents become more patient (δ falls), or as agents place a lower value on accuracy (α falls). Even using the reasonable parameter values of $\alpha = \delta = 0.5$ example 1 produced a very strict necessary condition for welfare improvement which needs to be met by the cost of information gathering. It might be better to risk being caught in an investment breakdown than to engage in public information gathering if this involves even a small cost. Taking into account the points made earlier about externalities, this

produces a second reason to look at projects on a case by case basis.

Note that a surprise revelation by a third party will not involve extra delay - since it will not have been expected. Therefore a case by case examination of whether a third party should intervene combined with a surprise revelation would seem sensible. For example, revelations of the type “the government has just released a report suggesting that the potential market for UK products in country X will be much greater than expected” or “a new C.B.I. report suggests that the demand for product Y is much higher than previously thought” might lead to investment where agents had decided not to invest because of a lack of information, but without creating an extra incentive to wait - hence removing the incentive to wait until the announcements for agents that did wish to produce goods for export to country X or increase production of good Y.

5.5 CONCLUSIONS

Section 5.2 developed a very simple example model which allowed the exploration of the solution method for endogenous-timing models in discrete time: the real options approach. We saw that there would exist a symmetric threshold signal value above which agents will decide to invest in the first period. A decision will be made by the second period in the two agent model with no revelation of further information by a third party. With more than two agents the decision can be delayed as long as at each period one agent decides to invest.

The full model in section 5.3 considered the possibility of a third party revealing the true state, through which the value of information due to be revealed at a pre-determined date might be evaluated. This allowed the consideration of a standard problem in investment theory. With common value multi-agent investment under uncertainty it is easy to foresee a failure of investment even if the true value of the state is strictly positive. This might lead to the suggested solution that a third party, such as the government, a regulator or even a joint body established by the agents to gather information, should attempt to evaluate the true state and correct such an investment breakdown, by revealing

positive value investments to all agents. In fact, it was shown that if complete revelation of the true state occurs then decisions may be delayed, even where they would have occurred were it not for the existence of the third party. There are therefore two effects on joint profits. The *profit bounding effect* means there will be a higher bound on expected profits, as worthwhile investment will always take place. However, the delay caused may, via discounting, actually slightly lower joint profits - the *investment delay effect*. A far more damaging point is that in the great majority of cases complete revelation provides no benefits. The net result is that there is significant doubt about the worth of gathering information when it involves a cost. An agent is likely to prefer to be in a situation where complete revelation of information at some point in the future *does not* take place if the agent has to partly fund information gathering. This suggests that the information will only be worthwhile if it is cheap and brings quick results. Surprise revelation of the true state at zero cost will of course unambiguously increase joint profits. Both agents will behave as in the model without complete revelation, but if there is a failure to invest in a worthwhile project, having seen that the project is in fact worthwhile and with no value to waiting, both agents will invest immediately after revelation occurs.

5.5.1 Possible Extensions

The model in section 5.3 has applications as diverse as anti-inflation policy and price-cap regulation. Consider a regulator with a fixed interval price-review period. If the regulator oversees numerous agents that have to make investment decisions, and it is known that at the price review information held by all agents and the regulator will be revealed, this may result in a delay to decision making. This is true *a fortiori* if a decision is about to be made at a time close to the price-review. It is possible that this decision need not have been delayed and the extra information would not have altered the initially proposed decision. Therefore, if the costs of gathering information are not small, then a regulator who collects comparative information and uses it to suggest new areas of potential investment may actually do so at the cost of welfare. The points about surprise revelation and externalities made in the last section provide an important set of

additional considerations.

There are various areas of potential future work involving models of the type developed here. Strategic concerns can be introduced by making utilities depend upon the actions of others, not just indirectly through information revelation, but directly within the utility function. It might be possible to model the desire to be a trend-setter, where utility is only available if others follow. In this case an agent would want others to follow, alternatively congestion might make an agent not want to follow. Modelling these features more comprehensively would require a multiple choice setting, perhaps involving a continuum of choice, a metric for the closeness of one project to another in terms of strategic complementarity or substitutability, multiple agents and a more complex specification of utility. The results would be likely to depend upon whether the other agents' utilities are strategic complements or substitutes, and also depend upon the proximity of different projects to each other. More complex stochastic processes and the introduction of continuous time would be useful for testing the robustness of the results. Relaxing the irreversibility of decisions might produce an interesting link with the optimal experimentation literature initiated by Rothschild (1974). Agents might face some cost to taking an action, but could change their minds later in the game having gained more information from their own experimentation with one of the choices.

It would appear sensible to test the predictions empirically and assess the validity of the assumptions of this chapter through econometric analysis and laboratory experimentation. The next chapter takes up the challenge of experimentally testing the predictions of endogenous-time herding models, and represents perhaps the first formal attempt to carry out such a test.

APPENDIX FOR CHAPTER 5

Proof of Proposition 16:

First the “if” part will be proven, then the “only if”.

(\Rightarrow) The cost of delay can be seen intuitively as $(1 - \delta)\mu^i$ since the unconditional expectation $E[\mu^j] = 0$.²⁴ This is simply the expected payoff at time 1 minus the expected payoff at time 2. Consider the benefit of delay: the option value. Here we need to consider the possibility of regret, where an investment made at time 1 actually seems less sensible when information made available at time 2 is revealed. Information of this sort comes about if it is observed that agent j did not invest at time 1, therefore revealing that $\mu^j < \bar{\mu}$ which provides some evidence that the state of the world is less likely to merit investment.²⁵ This can be avoided if agent i waits and so provides the option value of waiting which occurs with probability $\Pr[\mu^j < \bar{\mu}]$. The option value can therefore be defined as the expected loss avoided by agent i by not investing at $t = 1$ in the event that agent j does not invest at $t = 1$:

$$-\delta \Pr[\mu^j < \bar{\mu}] \{ \mu^i + E[\mu^j | \mu^j < \bar{\mu}] \} \quad (5.9)$$

We have a condition which leaves the marginal decision-maker indifferent when deciding to invest at time 1: indifference occurs when the option value exactly offsets the delay cost; this is none other than the standard value matching condition for a dynamic programming problem. This condition implicitly defines the value of $\bar{\mu}$ using the properties of the uniform distribution:

$$(1 - \delta)\bar{\mu} = -\delta \Pr[\mu^j < \bar{\mu}] \{ \bar{\mu} + E[\mu^j | \mu^j < \bar{\mu}] \} \Rightarrow \bar{\mu} = \frac{-(4-2\delta) \pm [(4-2\delta)^2 + 12\delta^2]^{\frac{1}{2}}}{6\delta} \quad (5.10)$$

²⁴The symmetry of the uniform $[-1, 1]$ distribution around zero makes the unconditional expectation equal to zero.

²⁵There is an assumption of symmetry here - or alternatively the equilibrium decision rules found could be referred to as the symmetric decision rules. The totally symmetric nature of the problem makes this a natural assumption and a natural equilibrium to seek. Gul and Lundholm (1995) make a strong case for the relevance of the symmetric equilibrium in a decision model of this type.

For $\delta \in (0, 1)$ and $\bar{\mu} \in [-1, 1]$ we can rule out one of these two results, eliminating:

$$\bar{\mu} = \frac{1}{6}\delta^{-1} \left\{ -(4 - 2\delta) - [(4 - 2\delta)^2 + 12\delta^2]^{\frac{1}{2}} \right\} \notin [-1, 1] \text{ for } \delta \in (0, 1) \quad (5.11)$$

This leaves the value of $\bar{\mu}$ uniquely given as:

$$\bar{\mu} = \frac{1}{3} + \frac{2}{3}\delta^{-1} \left[(\delta^2 - \delta + 1)^{\frac{1}{2}} - 1 \right] \quad (5.12)$$

Equation 5.12 is well defined for $\delta \in (0, 1)$ and gives a range of values for $\bar{\mu}$ of $\bar{\mu} \in (0, \frac{1}{3}]$, that can be roughly approximated by the linear function $\bar{\mu} = \frac{1}{3}\delta$ over the relevant range of values of δ . It has been shown that there exists a unique value of $\bar{\mu}$ given in equation 5.12 such that if $\mu^i > \bar{\mu}$ the cost of delay is strictly offset by the option value of waiting. We have $>$ since the cost of delay is rising in μ^i (and falling in δ) which therefore defines the optimal decision rule for agent i at time 1. The assumption of a positive option value to delay implies that $\bar{\mu} > 0$.

(\Leftarrow) Consider what the value μ^i must take if agent i has optimally decided to invest at time 1. Optimally deciding to invest implies that the delay cost is strictly offset by the option value:

$$(1 - \delta)\mu^* < -\delta \Pr[\mu^j < \mu^*] \{ \mu^* + E[\mu^j | \mu^j < \mu^*] \} \quad (5.13)$$

where μ^* implicitly defines the value of μ required for this inequality relation to hold. But this is exactly the value $\bar{\mu}$ we defined in the first part of the proof.

Proof of Proposition 17:

We are given that agent i did not invest at time $t = 1$. If this was so we know from proposition 16 that $\mu^i < \bar{\mu}$. Investment will benefit agent i if $E[\pi_2^i] > 0$. Two rationales for delay at $t = 1$ are possible and are considered in turn.

(a) If $\mu^i \in (-1, 0]$ and therefore $E[\pi_2^i] < 0$, only if new information suggested a rise in $E[\pi_2^i]$ would it be rational to decide to invest. Agent i must have observed one of two possible histories: $x_1^j = 1$ or $x_1^j = 0$. Only if he observed $x_1^j = 1$ would he raise his

expectation of π_2^i :

$$E[\pi_2^i | x_1^j = 1] = \mu^i + E[\mu^j | \mu^j > \bar{\mu}] = \mu^i + \frac{1+\bar{\mu}}{2} > \mu^i = E[\pi_1^i] \quad (5.14)$$

$$E[\pi_2^i | x_1^j = 0] = \mu^i + E[\mu^j | \mu^j < \bar{\mu}] = \mu^i - \frac{1-\bar{\mu}}{2} < \mu^i = E[\pi_1^i] \quad (5.15)$$

Since this is a symmetric problem the same is true for agent j if $\mu^j \in (-1, 0]$, therefore if agent j did not invest at $t = 1$ then he too would only raise his expectation if $x_1^i = 1$. If neither invest then no increase in expectation occurs at $t = 2$ and so neither invest at $t = 2$, and hence no rise in expectation occurs at $t = 3$, etc. Therefore we have shown that if one agent does not observe investment from the other he will not invest and the next period will look much like the second, so the decision not to invest becomes permanent. If either agent invested the other would increase his expectation, but only once (since the other player may never move again) and will therefore raise his expectation so $E[\pi_2^i | x_1^j = 1] > 0$ and invest at $t = 2$ or despite the increase it will be the case that $E[\pi_2^i | x_1^j = 1] < 0$ because his signal was so low, and no investment will take place at $t = 2$ or ever.

(b) If $\mu^i \in (0, \bar{\mu})$ and $E[\pi_2^i] > 0$ then he was delaying despite expecting positive profit because of the positive option value to delay. This option value has however been expended. If $x_1^j = 1$ then he would have been better off investing at $t = 1$ and would have done so had he realized that agent j would definitely invest. He will invest at $t = 2$ since there will be no further revelations as agent j has *de facto* left the game. Now if $x_1^j = 0$ agent i will lower his payoff expectation as will agent j therefore if it was optimal for them to delay at $t = 1$ it is optimal to delay at $t = 2$ *a fortiori* and so it will be optimal not to invest at $t = 2, 3, 4$, etc. We have shown that in all cases, agent i will either invest at $t = 1$, invest at $t = 2$, or never invest and so we have proven part (i) of the proposition. Furthermore in all the cases examined it is only optimal to invest at $t = 2$ if agent j invested at $t = 1$ and therefore we have also proven part (ii).

Proof of Proposition 18:

We need to show that if there is no investment at an arbitrary time, $t = \tau$, then there will be no investment at time $t = \tau + 1, \tau + 2, \dots$. We know from proposition 17 that if there is investment at time $t = \tau$ then agent i will not alter his optimal decision not to invest, and by symmetry this will be the case for all i . The only additional information revealed at time $t = \tau + 1$ lowers expected payoffs so as in proposition 17 agents will either go from a position where $\mu^i \in (-1, 0] \Rightarrow E[\pi_{\tau+1}^i] < 0$ and will then certainly not invest at time $t = 2$, or $\mu^i \in (0, \bar{\mu}) \Rightarrow E[\pi_{\tau+1}^i] > 0$ and they will have decided optimally to delay because of a positive option value, and it will remain optimal to delay *a fortiori* just as in the two agent case. At time $t = \tau + 3$ agent i is in an identical position to the position at time $t = \tau + 2$, since no agents have invested once more, so there is no additional information at all being revealed, and this will clearly be the case for $t = \tau + 4, \tau + 5, \tau + 6, \dots$. Therefore there will be no reason for any agent to change his optimal decision not to invest.

Proof of Proposition 22:

There is no option value at time $t = \tau^*$. Since the state of the world is now known with certainty, both w_{τ^*} and δ are known to agent i at time $t = \tau^*$, so there is no longer any need to consider the actions or information of agent j . Therefore a very simple decision rule is optimal: $x_{\tau^*}^i = 1 \Leftrightarrow \pi_{\tau^*}^i = \delta^2 w_{\tau^*} > 0$. Since we are dealing with a symmetric Markov process once we consider periods after $t = 3$ the agents will once again be forced to consider expectations of the true state of the world, but this expectation will be that the true state of the world is as it was at $t = 3$ and utilities will look much as they did at time $t = 3$ but with the addition of a positive MSE. The net result is that $E[\pi_4^i] < \pi_3^i$. Therefore if $\pi_3^i \geq 0$ there is no reason to delay beyond the period in which information revelation takes place. If $\pi_3^i < 0$ then investment is not profitable now and will be even more unprofitable beyond $t = 3$. In general, we can say that since $E[\pi_{\tau^*+1}^i] < \pi_{\tau^*}^i$ where $t = \tau^*$ is the time of full information revelation, then $\pi_{\tau^*}^i \geq 0 \Leftrightarrow x_{\tau^*}^i = 1$. Alternatively if $\pi_{\tau^*}^i < 0$ then $E[\pi_{\tau^*+1}^i] < 0 \Rightarrow E[\pi_{\tau^*+2}^i] < 0$ etc. therefore $x_t^i = 0$ for all $t \geq \tau^*$. Thus the solution to the decision problem is fully determined by time $t = \tau^*$.

Proof of Proposition 23:

Since $\varepsilon_t = 0$ for all $t \in T_{++}$ the only importance of time is through the discount rate. Complete revelation is only of any use if, in the world before the prospect of revelation, the agents' signal values were such that full or partial investment breakdown would have occurred. This requires that μ^i and μ^j are both in the region $[-1, \bar{\mu}(\alpha, \delta)]$ and that $w = \mu^i + \mu^j > 0$. The distribution of the value of the project at t , below the threshold value, is the sum of two uniform distributions with support $[-1, \bar{\mu}]$ and is therefore triangular with support $[-2, 2\bar{\mu}]$. Denote the probability of investment breakdown as $g(\bar{\mu}(\alpha, \delta)) \equiv g(\alpha, \delta)$. Using the properties of the triangular distribution $g(\alpha, \delta)$ is given by:

$$\begin{aligned} g(\alpha, \delta) &= \Pr[w > 0 \mid \mu^i < \bar{\mu} \ \& \ \mu^j < \bar{\mu}] \Pr[\mu^i < \bar{\mu} \ \& \ \mu^j < \bar{\mu}] \\ &= 2 \left(\frac{\bar{\mu}}{\bar{\mu}+1} \right)^2 \left(\frac{\bar{\mu}+1}{2} \right)^2 = \frac{1}{2} \bar{\mu}^2 \end{aligned} \quad (5.16)$$

Denote the probability of partial investment breakdown as $h(\bar{\mu}(\alpha, \delta)) \equiv h(\alpha, \delta)$.

$$h(\alpha, \delta) = 2 \Pr \left[-\mu^i < \mu^j < -\frac{\bar{\mu}+1}{2} \right] = 2 \Pr \left[\mu^j < -\frac{\bar{\mu}+1}{2} \right] \Pr \left[\mu^i + \mu^j > 0 \mid \mu^j < -\frac{\bar{\mu}+1}{2} \right]$$

The second probability is in fact just the probability that a drawing from a triangular distribution with support $[-2, \frac{1}{2}(1 - \bar{\mu})]$ is strictly positive. Using the characteristics of the uniform and triangular distributions this yields:

$$h(\alpha, \delta) = \left(\frac{1-\bar{\mu}}{2} \right) \left[2 \left(\frac{1-\bar{\mu}}{5-\bar{\mu}} \right)^2 \right] = \frac{(1-\bar{\mu})^3}{(5-\bar{\mu})^2} \quad (5.17)$$

Now since $f(\alpha, \delta) = g(\alpha, \delta) + h(\alpha, \delta)$, combining equations 5.16 and 5.17 yields the

required fraction of time when complete revelation is useful as required:

$$f(\alpha, \delta) = \frac{1}{2}\bar{\mu}^2 + \frac{(1-\bar{\mu})^3}{(5-\bar{\mu})^2} = \frac{2-6\bar{\mu}+32\bar{\mu}^2-10\bar{\mu}^3+\bar{\mu}^4}{50-20\bar{\mu}+2\bar{\mu}^2}$$

Proof of Proposition 24:

From remark 3 the investment delay effect causes a small but strictly positive loss in joint payoffs of $\kappa > 0$. From proposition 23 complete revelation will be useful in countering full investment breakdown for the fraction of cases $\frac{1}{2}\bar{\mu}^2$. In these cases the maximum potential gain in profit is the sum of the two highest signal values which still lie in the investment breakdown signal region, i.e. the two highest signals for which $\mu^i \in [0, \bar{\mu}(\alpha, \delta)]$, $\mu^j \in [0, \bar{\mu}(\alpha, \delta)]$ and $\mu^i + \mu^j > 0$. This produces the maximum possible combined signal value of $2\bar{\mu}$. Furthermore the return will only occur at the point of complete revelation, therefore the payoffs must be discounted up to that point. Hence we have a maximum possible gain from countering full investment breakdown of $2\delta^{\tau^*}\bar{\mu}(\frac{1}{2}\bar{\mu}^2)$. Complete revelation is also useful in countering partial investment breakdown which occurs in the fraction of cases $(5-\bar{\mu})^{-2}(1-\bar{\mu})^3$. In these cases the maximum possible gain is $1 + \frac{1}{2}(\bar{\mu} + 1)$, which must again be discounted up to the point of complete revelation, and so we have a maximum possible gain from countering partial investment breakdown of $\delta^{\tau^*}\frac{1}{2}(3+\bar{\mu})(1-\bar{\mu})^3(5-\bar{\mu})^{-2}$. Combining all of this we have a necessary condition on the cost of information gathering:

$$C_g \leq 2\delta^{\tau^*}\bar{\mu}\left(\frac{1}{2}\bar{\mu}^2\right) + \delta^{\tau^*}\frac{1}{2}(3+\bar{\mu})(1-\bar{\mu})^3(5-\bar{\mu})^{-2} - \kappa$$

So we have as a weaker necessary condition:

$$C_g < \delta^{\tau^*}\left(\bar{\mu}^3 + \frac{(3+\bar{\mu})(1-\bar{\mu})^3}{2(5-\bar{\mu})^2}\right)$$

CHAPTER 6

AN EXPERIMENT IN ENDOGENOUS TIME

6.1 INTRODUCTION

Although there is some recent experimental work examining sequential herding models, there is as yet no literature on the experimental testing of endogenous-time herding models of the type examined in the last chapter.²⁶ This chapter presents new experimental findings relating to such models and begins the task of evaluating the practical application of existing endogenous-time models. In such models, agents receive their private signals and may wait for as long as they wish before choosing a particular action. Agents must counterbalance their desire to observe the actions of others, infer further useful information, and then make a more informed choice, against the discounting of payoffs which occurs over time. The decisions to be made are complex, and therefore, as with most experiments, the central notion to be examined is (Bayesian) rationality. Here this comes through exploring whether people respond appropriately to private signals and publicly observed actions.

6.1.1 Overview

This chapter begins in section 6.2 with a brief examination of the literature detailing the experimental testing of herding theory. Section 6.3 details the experimental design, which

²⁶This chapter is based on SgROI (2000b), and was presented at the Experimental Economics Workshop, Department of Economics, University of Oxford, December 1999.

following Anderson and Holt (1997) is kept as simple as possible. Section 6.4 examines optimal behaviour within this design and gives the optimal strategies as a function of the initial signal and observed actions. Section 6.5 presents some analysis of the results which are in general very supportive of rationality and herding. In all cases decisions were made quickly with delay very close to suggested optimal levels. Despite rationality, the occasional incorrect herd (or reverse cascade) could not be avoided. Section 6.6 introduces an alternative condition: the subjects are now informed what the correct action was after each game is played. This should make no difference if all subjects behave rationally. In a reverse cascade many subjects disregard their own good information to follow those with bad information, and it is just after such an event that revealing the correct choice is the most devastating. Having discovered that an incorrect choice was made despite following the prevalent view many subjects respond by delaying decisions in future games and moving away from rationality. It seems here that sub-optimal actions derived from rational play can seed irrationality in the minds of players, possibly in response to the failure of rational play to produce success. In another sense certain well researched cognitive biases might be seen to be at work as subjects, having recently witnessed an unlikely event, exaggerate its importance for future decision-making. Section 6.7 concludes by stressing that the chapter does support the general rationality of subjects which as predicted by the theory will produce herds and in some cases even herding on the wrong action.

6.2 THE EXPERIMENTAL LITERATURE

The herding literature in its present form dates back to Banerjee (1992) and Bikchandani, Hirshleifer, and Welch (1992), and endogenous timing dates back to Chamley and Gale (1994) and Gul and Lundholm (1995). Given the recent nature of the most seminal papers, it is not surprising that very little experimental testing has been published.

In a discussion paper, Allsopp and Hey (1998) test a simple version of Banerjee (1992), and find that herding does indeed occur, but does so less frequently than predicted by

models of the type in chapter 1. Anderson and Holt (1997) produce results again showing that subjects are very willing to learn from the actions of their predecessors, more or less as predicted by the theory. No published paper yet deals with an experimental test of the predictions of the endogenous-timing herding models. Despite this the simple experimental design in Anderson and Holt (1997) is worth detailing here as it will form the basis of the more complex experimental design in the next section.

Consider two urns, A and B, each containing three balls. Urn A contains two red and one white ball, and urn B contains two white and one red ball. The contents of one urn (randomized with odds 50:50) are emptied into a container. This process is not seen by test subjects, but they are made fully aware of the original contents of the urns and that with 50:50 odds one urn has been emptied into the container. Next we allow a sequence of test subjects to arrive at the container and select one ball from the container, note its colour and replace it. After selecting a ball they must predict which urn was used, winning a prize for a correct answer. A red ball is suggestive of urn A and a white ball of urn B, but neither is completely revealing. All signals are therefore of the same quality. Once subjects have made their decision it is noted on a board which is clearly visible by all the subjects in the room. Anderson and Holt found that herding occurred consistently in the laboratory where other sociological incentives to go along with the crowd can be controlled. Some decision sequences resulted in reverse cascades where initially misrepresentative signals started chains of incorrect decisions not broken by more representative signals gained later. Cascades were roughly split between reverse and normal cascades. In 12 sessions cascades formed in 87 periods of 122 in which they were possible. Individuals generally used information efficiently and followed the decisions of others when it was rational. They did find that there were errors which tended to make subjects rely more on their own private signals. Anderson and Holt felt they could explain this by factoring in the positive probability of an error in decision-making. Subjects would then slightly favour their own signals to the possibly erroneous decisions of others. They also found that the most prevalent systematic bias was the tendency for some test subjects to rely on the simple counting of signals rather than

the use of Bayes' rule for updating where these implied different decisions. The main conclusion to be taken from Anderson and Holt is that the most basic herding literature does seem to have some predictive power in the laboratory, and that this justifies further experimental investigation of the later theoretical work in herding.

The Anderson and Holt experiment was based entirely on sequential discrete models in the style of chapter 2. As a result they did not consider endogenous-time or discounting, which are examined in this chapter.

6.3 EXPERIMENT DESIGN

The experiment was designed to be as simple as possible and yet still capture the main themes of endogenous timing in herding models. It was essential to avoid any form of collusion, and to this end throughout the experiment no form of communication between subjects was allowed. The first part of the appendix gives the full experimental text. 54 subjects took part, each playing four games, for a grand total of 216 different games. The data is fully described in section 6.5.

6.3.1 First Step: Signals

There are two urns, one *red* and one *white*. The *red* urn contains two red and one white ball and the *white* urn contains two white and one red ball, and this is known to the test subjects. The contents of one urn is emptied into a container, the probability of this being the *red* or the *white* urn standing at 50%. Again, this is known to the test subjects. Subjects win a prize for correctly guessing which urn was emptied. The subjects arrive in sequence at the container and draw two balls, one at a time, *replacing* the ball after each draw. This gives them a signal as to which urn was emptied, a red ball suggesting the *red* urn, a white ball suggesting the *white* urn. Since they have two draws they have signals of different quality. There are three sorts of signal: a strong signal in favour of the *red* urn (two red); a strong signal in favour of the *white* urn (two white); and a neutral signal conveying no information (one red, one white). Subjects then return to their seats,

and without communication wait until all other subjects have drawn their signals.

6.3.2 Second Step: The Prize

Each subject is told that he or she will be awarded a prize for guessing correctly which urn was emptied. They are also told that they may wait as long as they wish before coming to a decision, but that for every minute waited their potential prize will fall steadily over time. The reward scheme is explained to the subjects, based on the following scheme: (i) each subject in a game is certain to receive £1.50; (ii) each subject also receives a prize of between £0 and £3.50; (iii) the prize is £0 if a decision is incorrect; (iv) if correct the decision will yield £3.50 minus 25 pence for every minute waited beyond the first. To make all this clear the subjects are given a table which allows them to check easily what their payoff will be at every time interval, and the table is also publicly displayed. Ten seconds before each minute elapses the time is announced. A large timer is clearly visible throughout the experiment. When everyone is ready, and the rules are thoroughly understood, the timer is started.

6.3.3 Third Step: Decision-Making

The experiment then enters the decision-making step. To capture the notion of discrete time the subjects each receive private forms upon which to make their choices and at the end of each period (a minute in length) the game pauses and the form is viewed by the experimenter or an assistant. The choices are: *red*, *white* or *wait*. After noting any positive choices (*red* or *white*) on a board clearly visible to all, the next period is initiated and play continues, pausing again at the end of the second period, and so on. After making a choice other than *wait* subjects leave the main experiment room. When all have made a decision other than *wait*, all subjects return to the main experiment room and the correct urn is revealed. The game definitely ends after 15 periods and this is made clear to the subjects, so if waiting continues to period 14 all will know that they must decide *red* or *white* in period 15 or receive no prize.

At the end of this stage the debriefing starts and payments are made based upon

subjects' performance. In practice, in each session 4 games were played in a period of between 90 minutes and 2 hours. This gives an average expected payment of approximately £8 per hour.

6.4 OPTIMAL BEHAVIOUR

This section considers what the optimal behaviour resulting from the experimental design might be expected to look like.

The notation XY is used to denote the two draws by a subject where X and Y can be R (red) or W (white). The italicized *red* and *white* refer to the red urn and white urn respectively. Consider any given signal actually representing the probability that the correct urn is *red*. In this case we have three signals: RR is an 80% signal; RW is a 50% signal; and WW is a 20% signal. All signals are now on the interval $(0, 1)$. Define a signal strength μ_R such that if $\mu^i \geq \mu_R$ a subject i will definitely go for *red* in the first period. Similarly define a signal strength $\mu^i \leq \mu_W$ such that a subject i will choose *white*. The area in the middle is a zone of uncertainty such that our subject will wait. Our task is to find out if the available signal strengths produce clear cut optimal actions in these regions.

Consider choosing *red* in the first period on the basis of a RR signal. This signal is strong, providing an expected payoff of £4. Opting for white would provide an expected payoff of £1. Waiting is a more complex case. If you wait you are likely to obtain more information. Even if the other subjects do nothing, they still reveal that they do not have strong enough signals to choose *red* or *white* in the first period. For example, if the subject waits until the second time period and observes all 8 other subjects choosing *white* this would clearly imply that the 8 other subjects did not have RR signals, and our subject might *regret* selecting *red*.

Start with the assumption that movement on the first period requires a strong signal,

so we have the following equilibrium mappings:

Signal $RR \mapsto$ Action *red*

Signal $WW \mapsto$ Action *white*

Signal WR or $RW \mapsto$ Action *wait*

We need to check that a subject would be correct to follow this course of action given that the other subjects also do so (a Nash check). We remove the certain payoff of £1.50 per game, since this is guaranteed and consider the bonus for a correct guess. The *cost of delaying* an action until the second minute, given the observed signal RR , is 20 pence, that is the difference between choosing *red* now and *red* next period, multiplied by the probability that *red* is actually the correct choice. This probability is 0.8 given an observed signal of RR . Now we must look at the *benefit of delay* to determine if the cost exceeds the benefit of delay when the initial signal is RR . We will go through one calculation, before looking at the general calculation. Having elected to choose *red*, the subject now gets to observe the first period decisions of the other 8 players, and there are 45 different possibilities. Of these 45, certain combinations will induce regret about having chosen *red*. For example, if the subject should observe all 8 other subjects selecting *white*. From the hypothesized equilibrium this implies a total set of inferable draws of 16 white balls and 2 red balls, that is a net 14 white balls. This provides a probability in favour of the urn being *white* of $\frac{2^{14}}{2^{14}+1}$, with only a probability of $\frac{1}{2^{14}+1}$ in favour of *red*. Therefore subject i will now see his expected payoff bonus shrink from $0.8 \times \text{£}3.50 = \text{£}2.80$, to 0.0214 pence. He should certainly feel considerable regret. The benefit of delay in this case is just under £3.25. However, he will not expect to see all 8 other players go for *white*! In fact, if he receives an initial RR signal he will expect to see very few people acting as if they had WW signals. In terms of *real options theory*, this

notion of regret is none other than the real option value of the decision to wait, destroyed through choosing positive action. This is similar to the real option in an optimal stopping problem; see Dixit (1993).

The second part of the appendix calculates the general benefit of delay incorporating this option value and shows that for a subject with a RR or WW signal the cost of delay is 20 pence, whereas the benefit of delay is only 0.5 pence, therefore the subject should choose *red* or *white* respectively in the first period. By contrast, the appendix also shows that the cost of delay for a subject with an initially neutral signal (RW or WR) is 12.5 pence whereas the benefit of delay is 38.5 pence. Since this exceeds the cost of delay, the subject with a neutral signal should wait. Therefore, we see that the candidate equilibrium is actually reasonable, since the full mapping from signal to strategy is consistent with optimal behaviour. With this candidate equilibrium shown to form an optimal set of (Nash) actions, we can say immediately that all subjects with strong signals should move immediately and all subjects with neutral signals should wait one period. Having observed all the strong signals after one period all the subjects with neutral signals should then make their decision in the second period to avoid further discounting, since there is no further benefit to waiting.

To summarize, optimal expected symmetric actions are: all with RR choose *red* in the first period; all with WW choose *white* in the first period; and all with neutral signals of R and W should wait one period, then select based on the majority of first-period choices. In a tie they should simply randomize in the second period. There is no benefit from waiting any further since all the useful information has been revealed. This is the unique symmetric equilibrium, and the strategies which form this equilibrium are therefore considered the expected actions of the experimental subjects.

6.5 RESULTS

This section evaluates the raw data to be found in tabular form in the final part of the appendix. Glancing at the data a number of points stand out:

- (1) The games ended quickly, in two or at most three periods.
- (2) Almost all players waited at least one period when their signals were not very conclusive.
- (3) Initial movement was almost always based on strong signals.
- (4) In almost all cases what occurred looked like herding.
- (5) Most “herds” were on the right action, though four were on the wrong action.

Based on the calculated optimal strategies in section 6.4 we can assign optimal actions to the various observed signals in the experiment.

To address one immediate concern, it is fairly clear from the optimal behaviour calculations in the second part of the appendix that the problem is not a trivial one. Drawing RR from the urn and choosing sensibly involves finding the cost of delay and then comparing this with the likelihood of seeing any useful information if you wait. The size of the calculation indicates that it might be unreasonable to expect a subject to work this out without a calculator in a single minute. However, the scale of the cost-benefit differential means that precise measures are not necessary. Since 20 pence is 40 times larger than 0.5 pence it should be fairly obvious to anyone with a basic mathematical ability that moving immediately is the right thing to do if you have a strong signal. Since 38 pence is three times larger than 12.5 pence it should also be reasonable to expect subjects to wait if they receive a mixed signal. In fact, even using a simple rule of thumb, it is fairly clear that a completely uninformative signal should not produce immediate action with such a mild discount rate. Since the optimal actions are so clear cut we could reasonably hope for optimal behaviour despite the complexity of the calculations required.

This section splits the optimal actions into two benchmarks. Initially we consider a thought experiment in which we simply forced *all* subjects to play according to the optimal strategy set given their signals. This generates a set of *ex ante* optimal actions and a corresponding set of payoffs. Secondly, we consider each player's actions and compare these with the expected optimal actions *given the observed actions of others*. In this way if one player has deviated a second player might choose a rational option but this might diverge from the *ex ante* set of optimal actions when we assumed no

such deviation. In this sense we have a set of *ex post* optimal actions to compare with observed actions. We can also calculate the payoffs associated with the *ex ante* and *ex post* optimal actions and the observed payoffs.

6.5.1 Ex Ante Predicted Choices

Here we concentrate on the actions that would be taken if all subjects were to act according to the optimal strategies in section 6.4. It is interesting to note that in only one case did a subject exceed the expected *ex ante* optimal payoff, but here one of the optimal rules involves randomization and so there is a 50% chance that a subject following the optimal strategy would have done better. Figure 16 summarizes the actual payoffs compared with the payoffs that would have ensued if *all* subjects behaved optimally.

FIGURE 16: Average Subject Payments, £

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Overall
Observed Actual π	14.23	18.69	16.06	17.56	16.00	16.11	16.44
<i>Ex Ante</i> Optimal π	15.44	18.86	18.00	17.64	16.89	16.35	17.20
Absolute % difference	7.84	0.90	10.78	0.45	5.27	1.47	4.42

The average *ex ante* optimal payoff for day 1 was £15.44, which is a little higher than the actually observed average of £14.13. On day 2 the optimal average would have been £18.86, whereas the observed actual average was £18.69. For day 3 the observed figure was £16.06 and the *ex ante* optimal figure was £18, and similarly for days 4, 5 and 6 observed averages were slightly below the optimal figure. Overall, the average *ex ante* optimal payoff would have been £17.20, whereas the observed average was a little lower at £16.44. The percentage difference is only 4.4%.

FIGURE 17: Percentage of Predicted Choices

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Overall
Colour	81	100	86	100	94	97	93
Time	89	86	89	92	89	94	90
Time (+1 period) ¹	94	100	97	100	100	100	99
Time (+2 periods) ¹	100	100	100	100	100	100	100

Note: ¹ +X period(s) refers to a choice of time within X period(s) of the predicted time.

In order to examine how close the subjects got to the predicted *ex ante* behaviour we split the data into two parts, the correct colour choice and the correct time of action. Within the optimal action sets there exists randomization at points of indifference. It is clearly difficult to determine whether subjects used randomization, so with this in mind, any colour choice at a point of indifference is taken to be optimal (since either colour would be optimal). Details are given in figure 17. What we find is that overall 93% of colour choices were *ex ante* optimal reactions to signals. In terms of time choices, 90% of decisions were made when predicted, 9% of decisions were delayed by one extra period and only 1% of decisions were delayed by an extra two periods, with no delays beyond two extra periods.

What we have is a picture of day 1 as departing mildly from *ex ante* optimality in terms of colour choice, and having a close to the *ex ante* optimal pattern of timing decisions. Day 2 looks much like optimality in terms of colour choice and in terms of timing. This seems to provide some evidence for the ability of subjects to choose correctly but a little more slowly than if fully optimal. More than 90% of decisions made by subjects were exactly as predicted by time and colour which seems very high given the complexity of the decision. This rises to 93% when we add a lag of a single period. Days 3 through to 6 follow a similar pattern.

6.5.2 Rational Responses

We no longer *impose* rationality on *all* subjects to find our benchmark. Instead we require that the subjects *respond rationally* given the assumption that all others are rational.

This is perhaps the more natural way to examine the data. For example, given the early decision by some with good signals to wait in game 2 of day 1 and the resulting decision to go for the wrong choice by those with neutral signals, the behaviour of later movers within a reverse-cascade is not an irrational phenomenon. We would still expect a strong signal (RR or WW) to result in optimal first period movement in the relevant colour. However, a neutral signal (WR or RW) would result in waiting until the second period and then basing a decision on the majority choice in the first period. This would factor in observed behaviour, rather than *ex ante* optimal behaviour, and help solve the problem of one-off irrational choices rendering the decisions of all later movers sub-optimal, when they are actually responding rationally to this deviation. In this sense we are looking at *ex post* optimality. The second part of the appendix also provides the actions which rational agents should have played in response to the observed actions of the other subjects, while this subsection examines some aggregate findings based a comparison of the raw data with the *ex post* optimal actions.

FIGURE 18: Average Subject Payments, £

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Overall
Observed Actual π	14.23	18.69	16.06	17.56	16.00	16.11	16.44
<i>Ex Post</i> Optimal π	14.51	18.86	16.92	16.91	16.17	15.99	16.56
Absolute % difference	1.96	0.90	5.06	(-)3.84	1.03	(-)0.77	2.26

As before we can compare the optimal payoff, with our alternative measure of optimality, to the observed payoff, and this is done in figure 18. Since we are considering a weaker form of optimality as expected the percentage difference is on average lower, standing at only 2.26%. This figure is an average of the absolute percentage difference; the simple average is lower at 0.72%. Note that where the optimal strategy calls for a coin flip, and where that flip results in the correct actions being taken, it is easily possible to exceed the *expected* optimal payoff.

FIGURE 19: Percentage of Predicted Choices

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Overall
Colour	92	100	97	100	100	100	98
Time	89	86	89	92	89	94	90
Time (+1 period) ¹	94	100	97	100	100	100	99
Time (+2 periods) ¹	100	100	100	100	100	100	100

Note: ¹ +X period(s) refers to a choice of time within X period(s) of the predicted time.

Figure 19 suggests that the choice of colour and time of action was very close to our candidate set of rational actions. As before no one was more than 2 periods slower than they should have been and almost all were at most 1 period slower. Despite the complexity of the decision 90% of subjects behave exactly as predicted by time and colour and 98% behave as predicted with only one extra period of delay.

6.6 AN ALTERNATIVE CONDITION

This section details the results of making a slight change in the experimental design. For two days of the experiment the subjects were told what the correct choice was after each game was played. This alternative should make no difference to optimal play.

6.6.1 The Availability Hypothesis

Despite the assertion that the revelation of the correct option after each game should make no difference to optimal play, and the fact that subjects seemed to act close to optimally, the minor change in experiment design did in fact have a significant impact. In four days of the experiment reverse cascades were experienced, and in all cases games were played after the reverse cascades. In the games with the alternative condition in which the incorrect choices of subjects was made clear, there seems to have been a dramatic result. Directly after a reverse cascade, subjects slow their decision-making right down and seem to trust their own signals less. While subjects were not told that a reverse cascade occurred, for many this will have been obvious on discovering that they chose incorrectly

despite following the majority view of which choice to select. In effect the subjects seem to behave more optimally as long as unlikely events do not take place. A reverse cascade is sufficiently unlikely to shake their belief in their own optimality. This might well be related to certain commonly observed cognitive biases, such as the *availability heuristic* which seems to be a common feature in experimental psychology, for example see Akerlof and Yellen (1987). In general, this naturally occurring heuristic seems to result in subjects relying too heavily on salient information which is easily retrievable from memory. This might result in subjects over-exaggerating the likelihood of unlikely events having seen them occur recently. Many who are not confident in their own ability to correctly determine the relevant probabilities might set aside their beliefs and rely on observed outcomes, putting aside probabilities in favour of frequencies. According to the *representativeness heuristic*, subjects act as if stereotypes are more common than they should. This might also apply in combination with the availability heuristic to exaggerate the dangers of reverse cascades.

6.6.2 Examining the Data

The data given in figures 27 and figure 28 in the appendix detail the results of the experiment for the alternative condition. Figure 20 below summarizes the data for days on which reverse cascades were obtained.

FIGURE 20: Observed Behaviour Before and After Reverse Cascades

Day	Correct Choice Revealed?	Average Deviation from Optimum	
		Before	After
1	No	+0.11 periods	+0.06 periods
3	No	+0.11 periods	+0.11 periods
7	Yes	+0.06 periods	+0.44 periods
8	Yes	+0.06 periods	+0.39 periods

There is clearly a remarkable difference between the results for the alternative condition. There is a deviation from optimum of an extra 6-11% delay in most cases, but this

rises dramatically to about 40% after subjects have been informed of their payoff just after a reverse cascade. There is no appreciable change if they are not informed, as in the standard condition. This is an interesting result and not easily explained. A reverse cascade is a low probability event with unfortunate consequences for payoffs and should subjects incorrectly exaggerate the likelihood of this event there are clearly many possible irrational responses. Should subjects react by ignoring the signals of others, they would be expected to reduce delay and simply act immediately based on their private signal, but this was clearly not the case. Alternatively they might respond by doubting the significance of even their own private signal which might lead to a more random pattern of actions. Glancing at the raw data, this was also clearly not the case. The actual response seems to have been simply extra delay. This might be based on a reduction in confidence in signals resulting in a desire to wait longer and attempt to gain more information from other subjects. While this is clearly not a rational action if other subjects are sticking with the optimal set of strategies and indeed is also not rational if others also wait, it is a feasible irrational response. It might then be possible to conclude that the revelation to agents that they chose incorrectly, based on entering a low probability reverse cascade might, when combined with the commonly observed availability heuristic, result in a deviation from rational play.

It is interesting to note that when a reverse cascade occurred in game 2 on day 8, by game 4 the average deviation from optimum fell back to +0.22 periods. This is the only example of a day in which a reverse cascade occurred early enough in the sequence of four games to see any change in behaviour more than one game later and where subjects were told the correct choices between games. This suggests, though it is by no means conclusive, that the deviation from rationality which occurs after realizing a reverse cascade occurred might well be a short-run phenomenon.

6.7 CONCLUSIONS

Some general findings should be stressed. The experimental games should have ended in two periods if the players were fully rational. In fact they ended in three periods in a few cases, but often did end in two periods. Strictly this implies less than rational behaviour; however given the complexity of the required mental calculations this is close enough to suggest something like rational behaviour. In fact 85%-90% of the time subjects did just as expected, and this figure rises to almost 100% with the addition of up to 1 period of lag above the optimal time of action. It should be emphasized that it was made clear to the subjects that they had up to 15 periods in which to make a decision, so there was clearly a good deal of thought involved in acting so rapidly. Virtually all subjects decided to wait a period if they received an uninformative signal. This suggests that despite the discount factor virtually all subjects appreciated the value of waiting to observe others behaviour. The results certainly seem to support the theoretical literature with the proviso that some slight extra delay might be seen in practice, perhaps when a subject fears that others may not be rational. In this sense the results here do not conflict with the results in Anderson and Holt for the simpler sequential case.

Herds are a reality. This is clearly confirmed by the data, which even produced reverse cascades. The cascades occurred at times slightly more slowly than they should with fully optimal behaviour, but more often did occur by the second period as predicted. The reverse cascade was a particularly interesting case since it is difficult to explain this kind of phenomenon without reference to social learning. This could only reasonably happen if subjects were paying close attention to the behaviour of their peers and making a serious effort to update probabilities. This is also a clear warning that sensible observation can still result in poor payoffs.

It seems reasonable to suggest that if laboratory subjects with small prizes behave much as their should in theory then large firms or governments with so much more at stake are even more likely to learn from their peers and herd as a result. This is of course a subjective point.

An interesting additional finding comes from a close examination of the alternative condition given in section 6.6. It seems to be the case that when subjects realize they have chosen incorrectly despite having followed the majority view, they respond through extra delay in the following game. The delay seems to fall back a little in any later game. Therefore, when faced with the realization of a low probability reverse cascade subjects appear to respond with a higher degree of irrationality. Herding theory itself tells us that very sub-optimal choices may be made despite rational decision-making. The experimental findings add to this problem, by suggesting that when agents have recently faced a reverse cascade they might respond with a burst of irrationality which clearly increases the chance of further sub-optimal choices being made in future games. This would seem to make it even more important that policy makers keep a close watch for the dangers of reverse cascades in situations where herds are possible.

6.7.1 Rules of Thumb and Bounded Rationality

One criticism has arisen from the recent findings of Huck and Oechssler (2000). Cascades do seem ubiquitous, and more casual results from experimental economists seem to support the role of herds and even reverse cascades in the laboratory or the class-room. However, with the desire to keep experimental design simple, to allow a quick understanding of the rules of play, comes a danger. There are often simple rules of thumb which produce herding outcomes, but which are not reliant on a strict observance of Bayes' rule. Huck and Oechssler concentrate their criticism on sequential herds, but even in the case of this chapter their critique may apply. Consider the following rule of thumb: *follow your own signal unless it is inconclusive; in that case wait and follow the majority.*

The rule of thumb would produce an outcome almost indistinguishable from that found by following a strict observance of Bayes' rule. Furthermore the simplicity of rules of thumb of this type may make them attractive to subjects with limited time and resources. This chapter adds to the growing body of interest in herding experimentation, and is like all previous papers, subject to this criticism, yet the response of this chapter is rather to embrace the role of rules of thumb. The fact that there exists a rule of thumb

which very closely duplicates the optimal outcome can explain the success of the herding hypothesis in this and other papers, but more generally, demonstrates that herding as a general concern arises not simply out of a strict application of Bayesian updating, but can also arise in more informal settings.

One area for future research might be to answer the question: will there always exist a simple rule of thumb which generates an outcome which closely matches that predicted by strict Bayesian updating. This requires strict definitions of *simple* and *closely matches* and the obvious way forward is in the bounded rationality literature. To take this further we need a general framework in which to capture the notion of rules which can produce the optimal outcome but are not themselves optimal. The next chapter develops a way of modelling bounded rationality, based on *neural networks*, and produces a player which is in fact likely to follow such, close to rational, rules of thumb. This is formalized with reference to rules which locally minimize error by getting decisions right in a subclass of games, rather than globally minimizing error by always making the correct choice.

APPENDIX FOR CHAPTER 6

1. Subject Instructions

“Thank you for attending today’s experiment. I shall start by laying down a simple rule about silence, then detailing the experiment and finally explaining the prizes you can win. Feel free to ask questions. We shall then run a practice and you will have another opportunity to ask questions. When I am satisfied that everyone is ready we will begin the main experiment which will consist of what I shall call a game, repeated four times.

I would ask that you do not talk during the experiment except when I invite questions. You can raise your hand at any time to attract the attention of myself or one of my assistants who can deal with minor problems such as mislaying your pen. Otherwise please remain silent. Any attempt to do anything other than following the experimental instructions will be penalized through a reduction in your prize.

I shall ask you to perform a number of tasks. I shall explain some of my aims at the end of the experiment, but leave you to consider the significance as you see fit. Much of the action will revolve around the simple observation of coloured balls and decisions you will have to make. I shall not suggest any methodology: how you make your decisions is very much up to you.

These are the devices I shall use: a green urn and six balls, three red, three white. I am now placing two red balls and one white ball into the red bag. This bag is coloured red to help you remember that it contains a majority of red balls. Similarly I am now placing the remaining balls, two white and one red, into the white bag. When the first game begins I shall randomly choose one of these bags and empty it into the urn outside this room. You will not be able to see which bag was emptied. I shall then call you in sequence and you will leave this room and enter the adjoining corridor where you will stand with your back to the urn. The urn will be covered but I shall raise the cover when I am satisfied that you cannot see the contents of the urn and your hand will be directed into the urn. You will then take a single ball from the urn and you may remember the colour of that ball. You will then drop the ball back into the urn. I shall shake the urn

and you will, as before, draw once again from the urn. You will then be asked to leave the corridor, and be taken into the larger room. In the larger room you will be taken to a seat by an assistant and before you will be three items: a pen, a form and a table. The form must be completed as follows. You will see a space for your name, and your signal. By signal I mean the colour of the balls which you saw. The form also includes a list of time periods with a space next to each period in which you will be asked to write your choice. At this point you will simply wait until all the other subjects have drawn from the urn and entered the larger room. When all of you have entered the larger room I shall enter and remind you to write your name and signal on the form. I shall also remind you to keep silent and not attempt to move out of your seat until you are asked to leave. You will note that you are not able to see any of the other subjects' forms and you should not attempt to do so. The table lists the prize you could win based upon your actions. I shall detail this later.

After I have re-stated these rules I shall initiate the first time period. You will have one full minute in which to write one of three alternatives on your form alongside the first time period listed there. You will be told 10 seconds before the minute ends and a clock is clearly visible. The timing on this clock will be final. You may write: "red", "white", or "wait". A colour indicates you believe the urn contains the contents of the respective bag (red or white). If you write wait you are indicating that you do not yet wish to make a decision and wish to try again in the next period. An assistant will come to each of you and observe your form. If a colour is written there you will be asked to leave, taking with you your form. You will hand that form to me as you leave. I shall then note the colour choices that were made on the clearly visible board. For example, if two of you choose to write a colour, one red and one white, you will be asked to leave. You will give me your forms as you leave and I shall write on the board alongside the first period, "W, R", to indicate the two colour choices. Those who write "wait" will simply stay in their seats, but may observe the board, consult their tables, re-read their forms and do whatever else they wish, in silence, until the next period is announced. When the room has been emptied of those who made their decision I shall start the timer again and as

before you will have a minute to make your decision. This will continue until either all have decided or 15 minutes have elapsed. Then the game will end and I shall go with the last remaining subjects back to the small room. Those who left earlier will have been directed to the small room by an assistant and asked to wait there quietly. There will be another assistant there to ensure silence is observed.

When we all find ourselves back in the little room I shall rerun the game again. I shall once again go into the corridor and randomize which bag is emptied into the urn and we will go on as before. We will do this four times. After this I shall announce what the correct choices were, calculate the prizes you are due and for those who are interested, explain what I expected to observe.

Now I shall explain the structure of the prizes. After each game has ended I shall have a note of your four choices and when you made them. I shall also have a note of the correct choice, i.e. the bag which was emptied into the urn. This will totally determine your prize. You will get a basic reward of £1.50 for participation in each game, and a further prize of between £0 and £3.50 based on your performance. If you guess correctly in the first period you will get £3.50. If you wait until the second period and then guess correctly you will get £3.25. The prize will continue to fall by 25 pence every period. Therefore if you guess correctly in the final period, period 15, you will get only the basic £1.50 for taking part. If you guess incorrectly at any time you will also get just the basic £1.50 for taking part. This is all detailed on the tables in the larger room and you can consult this at any time. It is also listed on the board in this room and in the larger room. There are also copies of the tables in this room alongside your seats.

As the game is repeated four times you will receive a definite £6 and potentially up to £20. We have up to 2 hours to complete the experiment including the time I have used in explaining these rules. Before we start the experiment we will run through one practice game. This will be run just as I have described the main games except for two changes. You will not receive a prize for your actions, and we will have another opportunity for questions at the end of the practice. It will also give me the opportunity to double-check that you understand what is expected of you. You now have some time to examine the

tables and ask any questions.”

2. The Costs and benefits of Delay

This appendix calculates the costs and benefits of delay for a subject in the experiment with identical discount factors. Start by assuming the subject has already observed a private *RR* signal. This then leads him to expect others to be more likely to have also observed signals from the *red* urn, since his own signal biases him in that direction. He should consider all 45 alternative sets of signal observations, calculate the likelihood of each one taking place (which will be biased by his initial beliefs based on his own *RR* signal) and then evaluate the expected value given the possible signal. This will in some cases produce regret, where the subject feels he should really have gone for *white*. The total regret is the total of all the expected values of opting for *red*, given that *white* seems more likely after waiting one period, weighted by the likelihood of observing these pro-*white* sequences of signals. The expression begins with $\frac{1}{3}$ to the power equal to the number of possible *R* signals and $\frac{2}{3}$ to the power equal to the number of *W* signals. This is then multiplied by the number of combinations which yield this set of signals and by the probability of *white* being correct, remembering we have already observed a *RR*. This set of calculations is then summed and the total is multiplied by the second period payoff.

The actual sets of possible actions which would lead to regret are:

$$\begin{aligned} &\{8WW\}, \{7WW, 1RR\}, \{6WW, 2RR\}, \{5WW, 3RR\}, \\ &\{7WW, 1WR\}, \{6WW, 1RR, 1WR\}, \{5WW, 2RR, 1WR\}, \{6WW, 2WR\}, \\ &\{5WW, 1RR, 2WR\}, \{4WW, 2RR, 2WR\}, \{5WW, 3WR\}, \{4WW, 1RR, 3WR\}, \\ &\{4WW, 4WR\}, \{3WW, 1RR, 4WR\}, \{3WW, 5WR\}, \{2WW, 6WR\}. \end{aligned}$$

Summing each possibility respectively produces:

$$\begin{aligned} &\frac{1}{3} \frac{16}{2^{14+1}} + \frac{1}{3} \frac{14 \cdot 2^2}{3} 8 \frac{2^{10}}{2^{10+1}} + \frac{1}{3} \frac{12 \cdot 2^4}{3} 28 \frac{2^6}{2^6+1} + \frac{1}{3} \frac{10 \cdot 2^6}{3} 56 \frac{2^2}{2^2+1} + \frac{1}{3} \frac{15 \cdot 2^1}{3} 8 \frac{2^{12}}{2^{12+1}} + \frac{1}{3} \frac{13 \cdot 2^3}{3} 56 \frac{2^8}{2^8+1} \\ &+ \frac{1}{3} \frac{11 \cdot 2^5}{3} 168 \frac{2^4}{2^4+1} + \frac{1}{3} \frac{14 \cdot 2^2}{3} 28 \frac{2^{10}}{2^{10+1}} + \frac{1}{3} \frac{12 \cdot 2^4}{3} 168 \frac{2^6}{2^6+1} + \frac{1}{3} \frac{10 \cdot 2^6}{3} 420 \frac{2^2}{2^2+1} + \frac{1}{3} \frac{13 \cdot 2^3}{3} 56 \frac{2^8}{2^8+1} \end{aligned}$$

$$+\frac{1}{3}\frac{11}{3}\frac{2^5}{3}280\frac{2^4}{2^4+1} + \frac{1}{3}\frac{12}{3}\frac{2^4}{3}70\frac{2^6}{2^6+1} + \frac{1}{3}\frac{10}{3}\frac{2^6}{3}280\frac{2^2}{2^2+1} + \frac{1}{3}\frac{11}{3}\frac{2^5}{3}56\frac{2^4}{2^4+1} + \frac{1}{3}\frac{10}{3}\frac{2^6}{3}28\frac{2^2}{2^2+1}$$

All multiplied by £3.25, yields 0.5 pence which is below the cost of delay of 20 pence so the subject acting optimally should choose *red* if he observes a *RR* signal. By symmetry, it is also the case that a subject with the signal *WW* should choose *white* in the first period.

For a neutral signal we follow a similar procedure except we have a mixed signal of *W* and *R* as our initial signal. The subject knows that the distribution is skewed in favour of the correct choice, so anticipates a $\frac{2}{3}$ probability of the correct ball being selected from the urn. We simply assume *white* is correct for this calculation, since this will give the same result as assuming *red* is correct by symmetry. For the cost of delay we instead assume no useful information so any choice (*red* or *white*) will result in a $\frac{1}{2}$ probability of the correct action. This yields $\frac{1}{2}(\text{£}3.50 - \text{£}3.25) = 12.5$ pence.

The sets which lead to regret in this case are slightly different:

$$\begin{aligned} &\{8WW\}, \{7WW, 1RR\}, \{6WW, 2RR\}, \{5WW, 3RR\}, \\ &\{7WW, 1WR\}, \{6WW, 1RR, 1WR\}, \{5WW, 2RR, 1WR\}, \{4WW, 3RR, 1WR\}, \\ &\{6WW, 2WR\}, \{5WW, 1RR, 2WR\}, \{4WW, 2RR, 2WR\}, \{5WW, 3WR\}, \\ &\{4WW, 1RR, 3WR\}, \{3WW, 2RR, 3WR\}, \{4WW, 4WR\}, \{3WW, 1RR, 4WR\}, \\ &\{3WW, 5WR\}, \{2WW, 1RR, 5WR\}, \{2WW, 6WR\}, \{1WW, 7WR\}. \end{aligned}$$

Summing each possibility respectively produces:

$$\begin{aligned} &\frac{2}{3}\frac{16}{3}\frac{2^{16}}{2^{16}+1} + \frac{2}{3}\frac{14}{3}\frac{1^2}{3}8\frac{2^{12}}{2^{12}+1} + \frac{2}{3}\frac{12}{3}\frac{1^4}{3}28\frac{2^8}{2^8+1} + \frac{2}{3}\frac{10}{3}\frac{1^6}{3}56\frac{2^4}{2^4+1} + \frac{2}{3}\frac{15}{3}\frac{1^1}{3}8\frac{2^{14}}{2^{14}+1} \\ &+\frac{2}{3}\frac{13}{3}\frac{1^3}{3}56\frac{2^{10}}{2^{10}+1} + \frac{2}{3}\frac{11}{3}\frac{1^5}{3}168\frac{2^6}{2^6+1} + \frac{2}{3}\frac{9}{3}\frac{1^7}{3}280\frac{2^2}{2^2+1} + \frac{2}{3}\frac{14}{3}\frac{1^2}{3}28\frac{2^{12}}{2^{12}+1} + \frac{2}{3}\frac{12}{3}\frac{1^4}{3}168\frac{2^8}{2^8+1} \\ &+\frac{2}{3}\frac{10}{3}\frac{1^6}{3}420\frac{2^4}{2^4+1} + \frac{2}{3}\frac{13}{3}\frac{1^3}{3}56\frac{2^{10}}{2^{10}+1} + \frac{2}{3}\frac{11}{3}\frac{1^5}{3}280\frac{2^6}{2^6+1} + \frac{2}{3}\frac{9}{3}\frac{1^7}{3}560\frac{2^2}{2^2+1} + \frac{2}{3}\frac{12}{3}\frac{1^4}{3}70\frac{2^8}{2^8+1} \end{aligned}$$

$$+\frac{2^{10}1^6}{3^3}280\frac{2^4}{2^4+1} + \frac{2^{11}1^5}{3^3}56\frac{2^6}{2^6+1} + \frac{2^91^7}{3^3}168\frac{2^2}{2^2+1} + \frac{2^{10}1^6}{3^3}28\frac{2^4}{2^4+1} + \frac{2^91^7}{3^3}8\frac{2^2}{2^2+1}$$

All multiplied by £3.25, yields 38.5 pence to the nearest half-penny. Since this exceeds the cost of delay, which equals 12.5 pence, the subject with a mixed signal of R and W should wait.

3. Raw Data

This appendix details the raw data taken over the course of the experiment. The subjects are indexed by day and actual payoff. For example, the highest payoff subject on day two is given the index 2.1.

FIGURE 21: Optimal and Actual Actions, Day 1¹

Subject	Signals	Actions ²			Payoffs		
		Ex Ante	Ex Post	Actual	Ex Ante	Ex Post	Actual
1.1	WR, RR, RR, WW	?2, r1, r1, w1	w2, r1, r1, w1	w2, w3, r1, w1	£18.125	£19.75	£16.25
1.2	WW, RW, WR, WW	w1, r2, ?2, w1	w1, w2, ?2, w1	w1, w2, r2, w1	£17.875	£14.625	£16.25
1.3	RR, RR, RR, WW	r1, r1, r1, w1	r1, r1, r1, w1	w2, w3, r1, w1	£16.50	£16.50	£16.25
1.4	RR, RR, WR, RW	r1, r1, ?2, w2	r1, r1, ?2, w2	r1, r1, r2, w2	£14.625	£14.375	£15.00
1.5	WW, WW, RW, WW	w1, w1, ?2, w1	w1, w1, ?2, w1	w1, w1, w2, w1	£14.625	£14.625	£13.00
1.6	WW, WW, RW, WR	w1, w1, ?2, w2	w1, w1, ?2, w2	w1, w1, w2, w2	£14.375	£14.375	£12.75
1.7	RR, RW, WR, WW	r1, r2, ?2, w1	r1, w2, ?2, w1	r1, w2, r2, w1	£14.375	£11.125	£12.75
1.8	RW, RW, WW, WW	?2, r2, w1, w1	w2, w2, w1, w1	w2, w2, w1, w2	£14.375	£12.75	£12.50
1.9	WR, WR, WW, WR	?2, r2, w1, w2	w2, w2, w1, w2	w2, w2, w1, w2	£14.125	£12.50	£12.50

Notes: 1 Correct choices: w,r,r,w; 2 For example, r1 is red in period 1, ?2 is randomize in period 2.

FIGURE 22: Optimal and Actual Actions, Day 2¹

Subject	Signals	Actions ²			Payoffs		
		Ex Ante	Ex Post	Actual	Ex Ante	Ex Post	Actual
2.1	WW, WW, RR, RR	w1, w1, r1, r1	w1, w1, r1, r1	w1, w1, r1, r1	£20.00	£20.00	£20.00
2.2	WW, WW, RR, RR	w1, w1, r1, r1	w1, w1, r1, r1	w1, w1, r2, r1	£20.00	£20.00	£19.75
2.3	WW, WW, WR, RW	w1, w1, r2, r2	w1, w1, r2, r2	w1, w1, r2, r2	£19.50	£19.50	£19.50
2.4	WW, WW, RW, RW	w1, w1, r2, r2	w1, w1, r2, r2	w1, w2, r2, r2	£19.50	£19.50	£19.25
2.5	WR, RW, RW, RR	w2, w2, r2, r1	w2, w2, r2, r1	w2, w2, r2, r1	£19.25	£19.25	£19.25
2.6	RW, RW, RR, RW	w2, w2, r1, r2	w2, w2, r1, r2	w2, w2, r1, r2	£19.25	£19.25	£19.25
2.7	RW, WW, WR, RR	w2, w1, r2, r1	w2, w1, r2, r1	w2, w2, r2, r2	£19.50	£19.50	£19.00
2.8	WW, WW, RR, WW	w1, w1, r1, w1	w1, w1, r1, w1	w1, w2, r1, w1	£16.50	£16.50	£16.25
2.9	RR, WW, RR, RW	r1, w1, r1, r2	r1, w1, r1, r2	r1, w2, r1, r2	£16.25	£16.25	£16.00

Notes: 1 Correct choices: w,w,r,r; 2 Actions are denoted as in the previous table.

FIGURE 23: Optimal and Actual Actions, Day 3¹

Subject	Signals	Actions ²			Payoffs		
		Ex Ante	Ex Post	Actual	Ex Ante	Ex Post	Actual
3.1	WW, WW, RR, RR	w1, w1, r1, r1	w1, w1, r1, r1	w1, w1, r1, r1	£20.00	£20.00	£20.00
3.2	WW, RW, RR, RR	w1, w2, r1, r1	w1, ?2, r1, r1	w1, w2, r1, r1	£19.75	£18.125	£19.75
3.3	WW, WR, WR, RR	w1, w2, r2, r1	w1, ?2, r2, r1	w1, w2, r2, r1	£19.50	£17.875	£19.50
3.4	WW, WR, RW, WR	w1, w2, r2, r2	w1, ?2, r2, r2	w1, r2, r2, r2	£19.25	£17.625	£16.00
3.5	WW, WR, RW, RR	w1, w2, r2, r1	w1, ?2, r2, r1	w2, r2, r2, r2	£19.50	£17.875	£15.75
3.6	WR, WW, RW, RW	w2, w1, r2, r2	w2, w1, r2, r2	w2, r3, r3, r2	£19.25	£19.25	£15.50
3.7	RR, RW, RR, RR	r1, w2, r1, r1	r1, ?2, r1, r1	r1, r2, r1, r1	£16.25	£14.625	£13.00
3.8	RW, RR, WW, WR	w2, r1, w1, r2	w2, r1, w1, r2	w2, r1, w1, r2	£12.50	£12.50	£12.50
3.9	RR, RW, RW, RR	r1, w2, r2, r1	r1, ?2, r2, r1	r1, r2, r2, r1	£16.00	£14.375	£12.50

Notes: 1 Correct choices: w,w,r,r; 2 Actions are denoted as in the previous table.

FIGURE 24: Optimal and Actual Actions, Day 4¹

Subject	Signals	Actions ²			Payoffs		
		Ex Ante	Ex Post	Actual	Ex Ante	Ex Post	Actual
4.1	RR, WW, RR, WW	r1, w1, r1, w1	r1, w1, r1, w1	r1, w1, r1, w1	£20.00	£20.00	£20.00
4.2	RR, WW, RR, WW	r1, w1, r1, w1	r1, w1, r1, w1	r1, w1, r1, w1	£20.00	£20.00	£20.00
4.3	RR, WW, WR, WR	r1, w1, r2, w2	r1, w1, r2, w2	r2, w1, r2, w2	£19.50	£19.50	£19.50
4.4	RW, WW, WR, WR	r2, w1, r2, w2	?2, w1, r2, w2	r3, w1, r2, w2	£19.25	£17.625	£19.00
4.5	RW, WW, WR, WR	r2, w1, r2, w2	?2, w1, r2, w2	r3, w1, r2, w2	£19.25	£17.625	£19.00
4.6	WW, WW, RW, WW	w1, w1, r2, w1	w1, w1, r2, w1	w1, w1, r2, w1	£16.25	£16.25	£16.25
4.7	WW, WR, WR, RW	w1, w2, r2, w2	w1, w2, r2, w2	w1, w2, r2, w2	£16.00	£16.00	£16.00
4.8	RW, RR, RR, RW	r2, r1, r1, w2	?2, r1, r1, w2	r3, r1, r1, w2	£16.00	£14.275	£15.75
4.9	RW, RW, WW, RR	r2, w2, w1, r1	?2, w2, w1, r1	r2, w2, w1, r1	£12.50	£10.875	£12.50

Notes: 1 Correct choices: r,w,r,w; 2 Actions are denoted as in the previous table.

FIGURE 25: Optimal and Actual Actions, Day 5¹

Subject	Signals	Actions ²			Payoffs		
		Ex Ante	Ex Post	Actual	Ex Ante	Ex Post	Actual
5.1	RR, WW, RR, RR	r1, w1, r1, r1	r1, w1, r1, r1	r1, w1, r1, r1	£20.00	£20.00	£20.00
5.2	RR, WW, RW, RR	r1, w1, r2, r1	r1, w1, ?2, r1	r1, w1, r2, r1	£19.75	£18.125	£19.75
5.3	RW, RW, RR, RR	r2, w2, r1, r1	r2, w2, r1, r1	r2, w2, r1, r1	£19.50	£19.50	£19.50
5.4	RR, WR, WR, RR	r1, w2, r2, r1	r1, w2, ?2, r1	r1, w2, r2, r1	£19.50	£17.875	£19.00
5.5	RW, RW, RR, RW	r2, w2, r1, r2	r2, w2, r1, r2	r2, w3, r2, r2	£19.25	£19.25	£18.75
5.6	RR, WW, WR, WR	r1, w1, r2, r2	r1, w1, ?2, r2	r2, w2, w2, r2	£19.50	£17.875	£15.75
5.7	WW, WR, WR, WR	w1, w2, r2, r2	w1, w2, ?2, r2	w1, w2, w2, r2	£15.75	£14.125	£12.50
5.8	RR, RR, WW, WW	r1, r1, w1, w1	r1, r1, w1, w1	r1, r1, w1, w1	£9.50	£9.50	£9.50
5.9	WW, WR, WW, WW	w1, w2, w1, w1	w1, w2, w1, w1	w1, w2, w1, w1	£9.25	£9.25	£9.25

Notes: 1 Correct choices: r,w,r,r; 2 Actions are denoted as in the previous table.

FIGURE 26: Optimal and Actual Actions, Day 6¹

Subject	Signals	Actions ²			Payoffs		
		Ex Ante	Ex Post	Actual	Ex Ante	Ex Post	Actual
6.1	WW, WW, WW, WR	w1, w1, w1, r2	w1, w1, w1, r2	w1, w1, w1, r2	£19.75	£19.75	£19.75
6.2	WW, WR, WW, RR	w1, w2, w1, r1	w1, w2, w1, r1	w1, w2, w1, r1	£19.75	£19.75	£19.75
6.3	WW, WW, WR, RR	w1, w1, ?2, r1	w1, w1, ?2, r1	w2, w1, w2, r1	£18.125	£18.125	£19.50
6.4	WR, WW, RW, WR	w2, w1, ?2, r2	?2, w1, ?2, r2	w3, w1, w2, r2	£17.625	£16.00	£19.00
6.5	RR, WW, WR, RR	r1, w1, ?2, r1	r1, w1, ?2, r1	r1, w1, w2, r1	£14.625	£14.625	£16.25
6.6	WW, WR, RR, RW	w1, w2, r1, r2	w1, w2, r1, r2	w1, w2, r1, r2	£16.00	£16.00	£16.00
6.7	RW, WW, RR, RW	w2, w1, r1, r2	?2, w1, r1, r2	r2, w1, r1, r2	£16.00	£14.375	£12.75
6.8	RR, RW, WR, RR	r1, w2, ?2, r1	r1, w2, ?2, r1	r1, w2, r2, r1	£14.375	£14.375	£12.75
6.9	RR, RR, WR, RW	r1, r1, ?2, r2	r1, r1, ?2, r2	r1, r1, r2, r2	£10.875	£10.875	£9.25

Notes: 1 Correct choices: w,w,w,r; 2 Actions are denoted as in the previous table.

The next two figures were based on the alternative condition in which the subjects are told what the correct answer is after each game.

FIGURE 27: Optimal and Actual Actions, Day 7¹

Subject	Signals	Actions ²			Payoffs		
		Ex Ante	Ex Post	Actual	Ex Ante	Ex Post	Actual
7.1	WW, RW, RR, RR	w1, w2, r1, r1	w1, w2, r1, r1	w1, w2, r1, r1	£19.75	£19.75	£19.75
7.2	WW, WW, WW, RR	w1, w1, w1, r1	w1, w1, w1, r1	w1, w1, w1, r1	£16.50	£16.50	£16.50
7.3	WR, WW, WR, RW	w2, w1, r2, r2	w2, w1, w2, r2	w2, w1, w2, r2	£19.25	£16.00	£16.00
7.4	WW, RW, WR, RW	w1, w2, r2, r2	w1, w2, w2, r2	w2, w2, w2, r2	£19.25	£16.00	£15.75
7.5	WR, WR, WR, RW	w2, w2, r2, r2	w2, w2, w2, r2	w2, w2, w2, r3	£19.00	£15.75	£15.50
7.6	WR, WR, RR, RW	w2, w2, r1, r2	w2, w2, r1, r2	w2, w2, w3, r3	£19.25	£19.25	£12.50
7.7	WR, WW, RW, WW	w2, w1, r2, w1	w2, w1, w2, w1	w2, w1, w2, w1	£16.00	£12.75	£12.50
7.8	RR, WR, RR, RR	r1, w2, r1, r1	r1, w2, r1, r1	r1, w2, w2, r2	£16.25	£16.25	£12.50
7.9	RW, RR, WW, RW	w2, r1, w1, r2	w2, r1, w1, r2	w2, r1, w1, r3	£12.75	£12.75	£12.50

Notes: 1 Correct choices: w,w,r,r; 2 Actions are denoted as in the previous table.

FIGURE 28: Optimal and Actual Actions, Day 8¹

Subject	Signals	Actions ²			Payoffs		
		Ex Ante	Ex Post	Actual	Ex Ante	Ex Post	Actual
8.1	RR, WW, RR, WR	r1, w1, r1, w2	r1, w1, r1, w2	r1, w1, r1, w2	£16.25	£16.25	£16.25
8.2	RW, RW, WR, WW	r2, r2, r2, w1	r2, w2, r2, w1	r2, w2, r2, w1	£19.25	£16.00	£16.00
8.3	RR, RW, WR, RW	r1, r2, r2, w2	r1, w2, r2, w2	r1, w2, r2, w2	£19.25	£16.00	£16.00
8.4	RR, WR, WR, RW	r1, r2, r2, w2	r1, w2, r2, w2	r1, w2, r2, w2	£19.25	£16.00	£16.00
8.5	WR, WW, RW, WW	r2, w1, r2, w1	r2, w1, r2, w1	r2, w1, r3, w1	£16.00	£16.00	£15.75
8.6	WW, RR, WR, RW	w1, r1, r2, w2	w1, r1, r2, w2	w1, w3, r3, w2	£16.00	£16.00	£12.25
8.7	RR, WR, RR, WR	r1, r2, r1, w2	r1, w2, r1, w2	r2, w2, r2, w3	£19.50	£16.25	£15.50
8.8	WR, RR, RW, WR	r2, r1, r2, w2	r2, r1, r2, w2	r2, w3, r3, w3	£19.25	£19.25	£15.25
8.9	WR, RR, WW, RR	r2, r1, w1, r1	r2, r1, w1, r1	r2, r1, w2, r1	£12.75	£12.75	£12.50

Notes: 1 Correct choices: r,r,r,w; 2 Actions are denoted as in the previous table.

This data was collected over eight days between May 1999 and November 1999. In all cases subjects had no previous experience of similar experiments and no prior knowledge of herding theory, or game theory in general. Graduate students in economics and related fields were excluded from consideration as subjects. The age range was from late teens to late twenties and there was a roughly even gender mix.

PART III
LEARNING BY EXAMPLE

CHAPTER 7

NEURAL NETWORKS AND BOUNDED-RATIONAL LEARNING

7.1 INTRODUCTION

That people do not behave fully rationally is not a controversial assertion.²⁷ How best to model bounded rationality is, however, a difficult question. So difficult that most game theorists and economists opt for one of two extremes: the assumption of full rationality implying that players will play Nash equilibria, and correctly Bayes update if the decision involves uncertainty; or the assumption of total ignorance with the hope that players will eventually achieve some form of rationality through a gradual learning process. Both alternatives have fair records, and in many cases an assumption of rationality still allows suboptimal social situations to be well modelled, as in the first two parts of this thesis. The second approach is winning more followers in the form of evolutionary game theory and this too has its successes, for example, Young (1993), or Kandori, Mailath, and Rob (1993). Some work has occurred between these two approaches, most notably automata theory has attempted to model a form of bounded-rational play based on the finite memories of players as in Neyman (1985), Abreu and Rubinstein (1988).

This chapter also opts for the middle ground, and attempts to model bounded-rational play in games, but does so in an innovative way. Here we focus on the way in which players learn and what they learn. The assumption is that they learn by example, and what they learn is a means of playing, not a single game, but all games. They learn

²⁷This chapter is based on the first half of Zizzo and Sgroi (2000).

a method of play which they can apply again and again to never before seen games. They do not ignorantly guess a solution, or simply make the correct decision every time, rather they use their previous experiences and make an intelligent, though imperfect, “bounded-rational” choice.

7.1.1 Motivation

Consider a naive agent thrown into a world of Nash equilibrium players. For example, he may be an infant who, first, starts playing games with parents and other relatives, and later with a larger and larger circle of people; he also observes other people playing games, either in front of him, or (say) on television. He does not always face the same game: on the contrary, he either plays or observes other subjects playing a wide range of games. Will he learn to play Nash strategies right away as he grows up, when facing games never encountered before? If yes, at what success rate? If the success rate is less than 100% but higher than chance, is it because the agent has *endogenously* learnt rules of thumb allowing him to perform *in a satisficing way* when playing new games? The rules of thumb would be learnt endogenously insofar as the agent’s behaviour would have always been reinforced according to Nash (he always plays with Nash players). This would be an example of emergent bounded-rational behaviour as in Simon (1955), Simon (1959) or Rubinstein (1998): rather than playing the optimal strategy, the agent achieves a “good enough” solution (hence, he is satisficing, in Simon’s terminology). Moreover, this would be a model of bounded-rational behaviour in games in which, differently from other models, for example, Osborne and Rubinstein (1998), rules of thumb emerge endogenously as a result of the learning process rather than being exogenously superimposed on the agent.

The issue of the learnability of Nash equilibria play is also relevant in its own right. Playing a Nash strategy in a game has long been thought a bare minimum requirement for rational play in games, and is treated as such in countless theoretical and applied papers. The challenge is to provide “a compelling argument for Nash equilibrium” (Mailath (1998), p. 1351). In the behavioural reinforcement paradigm, what we typically have

are agents capable of reaching Nash equilibrium (or otherwise) in specific games, after a feasible long-run dynamic path as in Roth and Erev (1995) and Roth and Erev (1998). In the evolutionary game theory paradigm, a good deal of recent work has gone into justifying Nash equilibrium as a stable point in a dynamic learning or evolutionary process, for example, Young (1993) or Kandori, Mailath, and Rob (1993); yet, by itself it does not provide structural models of learning as Mailath (1998) is at pains to point out. This has not gone unquestioned. Furthermore, what much of the work in both of these paradigms has in common is an emphasis on learning to play a Nash strategy in a particular game. Start with an arbitrary strategy in a game and determine whether agents converge to playing Nash strategies, and thus a Nash equilibrium, in the long-run. This chapter is very different, with the emphasis being placed on how reasonable Nash behaviour is in play *across a number of games*, or throughout a player's economic life. An implication of this is that these models typically make no prediction, or predict naive behaviour, when *new games* are faced.²⁸ Conversely, here the stress will be on the ability of the 'grown-up' agent to play in a non-naive way in games never encountered before, and even belonging to classes of games never encountered before.

7.1.2 Neural Networks

Neural networks are now a standard way to model psychological behaviour in computer science, engineering and cognitive psychology. As yet, however, relatively little has been done within game theory to capitalize on this research. Notable exceptions do exist, for example Rubinstein (1993) which uses perceptrons to capture notions of complexity in a model of monopolistic competition. Some work has been done using neural networks as models of learning in single games, in which case they can be seen as simple extensions of Roth and Erev (1995) style learning processes, for example Hutchins and Hazelhurst (1991). In this chapter, economic situations are modelled through a special class of normal form games, and the main player of interest is modelled as a neural network. The network is exposed to a series of example games where the Nash choice is highlighted as

²⁸Stahl (1998) provides one example of a notable exception.

the optimal action. Then the network, having been *trained*, faces a new series of games and is left to choose what strategies it sees fit.

What will happen next is the main concern of this chapter. To present a brief and simplified summary: a neural network with a powerful enough learning dynamic *could* in theory learn to play games never seen before in a fully rational way, picking Nash equilibria at the very first attempt, having been trained on a set of example games. However, the most standard biologically plausible learning method used by neural networks, *back-propagation*, is a gradient descent algorithm, and as such is at best likely to find only a local minimum of the error function it aims to minimize. This can be made more formal with the application of *algorithm complexity theory*, and we see that the problem faced by a network player using any form of gradient descent to train itself is *NP-hard*. So we come to the conclusion that a neural network is much more likely to learn a *local error-minimizing algorithm* (LMA) that, if followed, minimizes the network's error in a subset, but only in a subset, of cases. LMAs are interesting because they correspond to one or more behavioural heuristics that the bounded-rational agent has endogenously picked up to perform in a satisficing way on the decision problem, and look like rules of thumb, developed endogenously within the learning process.

This chapter therefore presents a theoretical account of the neural network as a plausible form of bounded-rational learning, able to develop an approach to new games well beyond random guessing, but well below the perfect selection of a Nash equilibrium at the first attempt. The next chapter moves on by asking the related question: if it is not learning to play Nash, what is it that a neural network is actually doing? If we believe that a neural network is the most biologically plausible model of bounded-rational learning then this question can be rephrased as: what could an economic actor be reasonably expected to do when facing a new situation?

7.1.3 Overview

The next section provides an introduction to neural networks, and details the features they exhibit which make them well suited to the task of modelling bounded-rational

players. In particular the focus is on neural networks as biologically plausible models of human behaviour, and on the scope for learning by example which they offer. Section 7.3 presents the framework into which our neural network player will be placed, a sequence of 3×3 normal form games in which the network player faces a series of players who always play Nash strategies. Section 7.4 formally defines the neural network player, and details the scope for the network to “learn by example”. The neural network’s task is defined as the pursuit of a general algorithm which allows the correct choice of a Nash equilibrium in never before seen games. Sections 7.5 and 7.6 present what is both a problem and a great success in the study of neural network learning: the failure of the learning process to provide a perfect success rate, but the expectation that the network will develop a method which will be partially successful. This provides the potential for neural network learning to successfully model *bounded rationality* not in terms of memory (as in automata theory) but in terms of the form of processing used by the network and possibly even biological brains. Section 7.5 begins by first detailing the certain existence of a set of network parameters which *will* solve the problem of finding a Nash algorithm. The section then goes on to stress the difficulties of avoiding local minima in the space which the network’s learning algorithm searches for the most successful error-minimizing algorithm. Section 7.6 develops an alternative way of characterizing the neural network’s difficulty as a *NP*-hard problem, a term which comes from *algorithm complexity* theory. A brief primer to the relevant area in complexity theory is provided which shows what the network’s learning can and cannot be expected to accomplish in *polynomial time*. Section 7.7 re-interprets the difficulty faced by the network in terms of its decision to accept rules of thumb, more formally defined as *local minimizing algorithms*, which do achieve a degree of success but only in a subset of the space of all possible games. Section 7.8 offers some conclusions.

7.2 NEURAL NETWORKS AND LEARNING

Neural networks can be loosely defined as *artificial intelligence models inspired by analogy with the brain and realizable in computer programs*. They typically learn by exposure to a series of examples (a training set), and adjustment of the strengths of the connections between its nodes. They are then able to do well not only on the original training set, but also when facing problems never encountered before. Unlike behavioural theories which assume stimulus-response mechanisms for learning better play in specific games such as Roth and Erev (1995), neural networks can learn not only how to choose a better response to the specific example it has been exposed to, but also how to choose a better response in a new problem. As such, neural networks may model how agents actually become rational by exposure to, and generalization from, a series of examples.

7.2.1 What is a Neural Network?

A feature of biological brains is that the connections between neurons are of different strengths, and that they can either increase or decrease the firing rate of the receiving neuron. In neural networks, this is modelled by attaching a connection weight to each connection. This weights the input from the sending node to the receiving node. Since the weight can be either positive or negative, the activation of a node will either increase or decrease the activation of the receiving node.

Networks can be usefully thought of as agents that receive external stimuli, process them, and produce an output. A typical network has an input layer of nodes receiving stimuli from the outside (as real numbers). This input is then transmitted to the nodes that the input layer is connected to, multiplied by the respective connection weights. Each node on the downstream layer receives input from many nodes. The sum is then transformed according to the activation function and transmits the result to the nodes in the further downstream layer, and so on. In such a way, the network processes the input until it reaches the output nodes (the output layer), in the form of new real-valued numbers. The activation level of the output nodes expresses the outcome of network

processing, i.e., the network's decision.

In other words, what the network does is a complex non-linear transformation mapping the input (a vector of numbers) into an output (another vector of numbers). Our network for economic decision-making receives the payoff values of the game as inputs. As output, it produces a strategy. We can then determine whether the network's choice is optimal, and thus rational, or not. Since we are dealing with normal form games of full information, the optimal choice will be the Nash equilibrium strategy. Learning to be rational will entail learning a set of connection weights such that the network is able to perform well not only on the examples in the training set, but also on new games.

Generally, the optimal parameter or set of parameters cannot be calculated analytically when the model is non-linear, and so must rely on a form of numerical optimization. The network adjusts connection weights during training following a basic learning algorithm (effectively a numerical optimization technique) called *backpropagation* developed by Rumelhart, Hinton, and Williams (1986). Backpropagation compares actual with desired output relative to the example, and adjusts connection weights to try to minimize the difference. The basic idea is to *adjust backwards* connection weights from output connections through hidden layers to the input layer's connections. Adjustment is weighted by a *learning rate*. A second learning parameter, the *momentum*, makes connection changes smoother by introducing positive autocorrelation in the changes in connection weights when successive examples are presented.

Intuitively, the economic decision-maker tries to learn how to perform better in the task. If a player faces a Prisoner's Dilemma, and cooperates while the other player defects, he will not be very willing to repeat the experience in the future. In the training set, the *correct* answer will be that dictated by the Nash equilibrium. The greater the difference between the network's behaviour and the optimal (Nash equilibrium) strategy, the greater the adjustment that the learning algorithm will trigger.

7.2.2 Learning by Example

Neural networks capture what is perhaps one of the most prevalent forms of learning in the real world - learning by example. The players within this chapter are not assumed to have perfect access to models of the real world, or even assumed to be able to manipulate Bayes' Rule, rather they are simply subjected to a sequence of example games, and then asked to assimilate what general knowledge they can from these examples to play new, never before seen games.

Although in some cases agents may receive a direct statement from their peers which expresses the optimal action, in many cases they will be forced to generalize based on simple examples and early observations in life. Both are well captured by the neural network methodology (see for example Zizzo (2000a) for more on this). So, as in Roth and Erev (1995), the basic idea is that of psychological reinforcement. However, here reinforcement does not entail direct adjustments to economic behaviour. Rather, it operates directly on connection weights, and only indirectly on behaviour.

The difference may appear subtle but is crucial. The behavioural learner learns how to behave better in an economic situation, but will be completely naive as soon as he faces a new one: knowing how to perform well in a coordination game tells him nothing about how to perform optimally in a Prisoner's Dilemma. Instead, given enough exposure to examples, the neural network learner is able to find a set of connection weights that enables it to perform optimally a majority of times even in economic situations never encountered before. In other words, it learns how to generalize its economic know-how.

The nearest alternative theory is undoubtedly evolutionary game theory which suggests that play may approach Nash play over time in specific games, but does not deal with the reasonable objection that surely experience of any game will enable players to play better! Stahl and Wilson (1994) find that subjects facing a set of games never encountered before gave correct answers an average of 59.6% of the time. This is a much better performance than chance, and yet one-shot games do not allow the scope for a learning process to bring them to a Nash solution. Clearly we do not need to see multiple examples of every game before we can correctly choose a Nash strategy. In many cases all

that is required to generate Nash equilibria is repeated exposure to examples, which is not a hard requirement for consumers, workers, firms and other economic agents. Such repeated exposure enables the decision-maker not only to perform well on the examples, which would hardly be surprising, but also on new problems never encountered before. The neural network model developed in this chapter and tested in the next chapter will be able to play rationally on newly faced economic problems, of the kind used by Stahl and Wilson (1994), over 60% of the time. Rationality is here an emergent psychological property of the network: the network learns to be rational. It learns how to play Nash equilibria in games it has never seen before.

7.2.3 Prototype vs. Exemplar-Based Categorization

The problem of a network with n connections is to find an appropriate configuration of its connection weights in the n -dimensional space of their possible combinations. If there are only a few examples, the network will assimilate novel cases to the most similar one, producing a similar output; if there are many examples, the network implements *prototypical categorization* as detailed in Way (1997). Prototypes are the results of a process that extracts specific complexes of features (the statistical central tendency information) from a set of examples. Prototypical categorization is forced upon the network by its property that knowledge tends to be distributed across various nodes and connection weights, i.e. different examples tend to be coded over the same units; see for example, Van Gelder (1991). It follows that the features in which the different examples are common tend to be reinforced, whereas their differences tend to cancel out. New cases are assimilated to the nearest prototype, and the economic behaviour associated to that prototype follows.

7.2.4 Empirical Success

Neural networks are considered the best model available of the human brain. By using such a model we are attempting to capture learning in as realistic a manner as possible, moving one step beyond simpler reinforcement dynamics that while trivial to comprehend, are not based on any detailed examination of the way people actually learn.

Neural networks are here treated as psychological models of how agents actually face, and learn to face, problems never encountered before. There is certainly evidence that children learn by example - either by direct experience or by observation of instances - as they grow up (see Bandura (1977)). More importantly, they are able not only to learn the examples observed (for example, to understand or utter words or sentences) but also to generalize from those examples (for instance, learn to talk: Plunkett and Sinha (1992)). In cognitive science, neural networks have been used as models, among other things, for how agents actually face pattern recognition and categorization, as in Taraban and Palacios (1994), for child development, as in Elman, Bates, Johnson, Karniloff-Smith, Parisi, and Plunkett (1996), for animal learning, as in Schmajuk (1997), and even arithmetic learning, as in Anderson (1998).

What networks learn in practice is also what makes neural networks useful for psychological modelling. For example, a model of arithmetic learning that would predict the absence of mistakes is unlikely to be plausible when dealing with human subjects as shown in the work on arithmetic learning in Anderson (1998). It follows that the fact that in practice the network converges to a solution algorithm with mistakes is therefore of greater interest than if it were to converge to an algorithm achieving a 0% mistake rate.

Neural network agents are likely to locally minimize the costs of irrational behaviour, resulting in the use of *rules of thumb*, and algorithms which are close to Nash. The player will then converge to playing Nash strategies in some percentage of games, significantly higher than with random play, but significantly lower than 100% of the time. The next chapter will include simulations to give a flavour of what might be possible using games for which random play would give 33% success, and shows that a neural network player can learn an algorithm capable of raising that percentage to 60-70% success. The neural network is able to learn to choose a Nash strategy when facing economic situations never encountered before, at success rates that are shown to fit experimental data with human subjects.

The scope for rules of thumb to be used in laboratories is well known; however, less

well-known is how close these simple rules can get to optimality in terms of outcomes. The final part of chapter 6 demonstrated this point with reference to a rule of thumb which closely reproduces the outcome that would be expected if all players correctly applied Bayesian updating. As we shall see in this chapter a rule of thumb which in large part results in correct decisions being made can be closely modelled as a local error-minimizing algorithm, and it is these which neural networks seek out.

7.2.5 Many Games

Most evolutionary games, or learning dynamics, are based on the assumption that only one game is of interest. They then look at whether a player can converge to Nash behaviour within that game. If we then consider another game we have to reset the dynamic and start again, forgetting the long process of learning to play Nash completely, or else we have to assume that having mastered one game the player will simply pick a Nash equilibrium perfectly without any further need to learn. This chapter goes well beyond these two simplifications. The decision algorithm used by the player is formed out of a series of observed examples, the results being a decision-rule in which the emphasis is on learning how to play games *in general*.

7.2.6 A New Approach

Neural networks provide a new modelling tool that can be of interest to economists wishing to model learning by example, and generalization from these examples, in a range of applications. One such case is to model the formation of expectations, a crucial issue when dynamic optimization is involved, as it is today in much macroeconomics and industrial organization based on dynamic games; see for example, Cho and Sargent (1996). Another example of interest to evolutionary game theorists may be to use neural networks coupled with an evolutionary mechanism to simulate learning processes in economic decision-making in a more sophisticated way: examples of this kind (the learning of conventions) already exist (for example, Hutchins and Hazelhurst (1991) and Macy (1996)). Rubinstein (1993) uses a simple neural network model to capture the notion

of a limit to forward-looking behaviour by competing monopolists. Indeed, neural networks can be more generally applied everywhere automata models of bounded-rational behaviour have been used before; see for example Neyman (1985) or Abreu and Rubinstein (1988). Zizzo (2000b) shows how, through repeated exposure to different economic environments, networks can help in the modelling of the endogenous determination of preferences. These examples are not exhaustive. Rather, they are meant to convey the idea that neural networks might be useful to economists willing to model learning by example and generalization of learning in a variety of economic contexts.

7.3 THE MODEL

This section is devoted to a sequence of definitions, and a thorough explanation of the neural network model. The notion of *random games* will also be introduced, the key feature of which is that each game in a sequence will not have been played before by any player. This captures the notion of play in a new and changing environment and rules out standard theories of evolutionary learning in which the game to be repeatedly played is essentially unchanging.

7.3.1 Basic Definitions

For the purposes of this chapter, a *random game*, G , is defined as a 3×3 normal form game of full information with a unique pure strategy Nash equilibrium and randomly determined payoffs taken from a uniform distribution with support $[0, 1]$. More formally we can define the simple game by a list $G = \langle N, \{A_i, u_i\}_{i \in N} \rangle$. We will restrict the number of players, indexed by i , to $N = 2$. A_i describes player actions available to a player in role i , with realized actions given by $a_i \in A_i$. In each game we consider the set of feasible actions available to each player to be of size 3. Feasible action combinations are given by $A = A_1 \times A_2$. Payoffs for both players are given by $u_i : A_i \mapsto \mathbb{R}$ which is a standard von Neumann-Morgenstern utility function. Payoffs are bounded, so $\exists Q \geq 0$ such that $|u_i(a)| \leq Q$ for all a . More specifically we consider the payoffs to be randomly drawn

from a uniform $(0, 1)$ and then revealed to the players before they select an action, so $\forall_{i,a}, \sup u_i(a_i) = 1$. We will place one further restriction on the game, by requiring the existence of a single unique pure strategy Nash equilibrium. To summarize:

Definition 10. *A random game is a list $G = \langle N, \{A_i, u_i\}_{i \in N} \rangle$, such that $N = 2$ players meet to play a game playing realized action $a_i \in A_i$, where three action choices are allowed. The value of the payoffs, $u_i : A_i \mapsto \mathbb{R}$ are randomly drawn from a uniform $(0, 1)$, made known to the players before the game starts, and form a standard von Neumann-Morgenstern bounded utility function.*

Now consider a population of Q players playing a series of random games, indexed by $t \in \mathbb{N}^{++}$. We consider a pair to be drawn from our population and then forced to play a given random game. Define the unique pure strategy Nash strategy to be $\alpha_i \in A_i$ where a Nash strategy is defined as the unique choice of pure strategy by player i which, when taken in combination with the Nash strategy chosen by the other member of the drawn pair, form the unique Nash equilibrium in the random game. So defining the specific Nash strategy for player i in a given random game to be α_i , and indexing the second player by j we have:

$$u_i(a_i = \alpha_i \mid a_j = \alpha_j) > u_i(a_i \neq \alpha_i \mid a_j = \alpha_j) \quad (7.1)$$

We can say immediately that:

Proposition 25. *Player i , taken from the population of size Q , wishing to maximize u_i , must play the unique Nash strategy when drawn to play in G , and therefore the outcome will be the unique Nash equilibrium in G .*

This is trivial given the definition of a Nash equilibrium. To say more we must first define an *evolutionary stable strategy*:

Definition 11. *Let x and y be two mixed strategies from the set of mixed strategies in G . Now let $u(x, y)$ define the utility associated with the play of strategy x given the play of*

strategy y by the other player. The strategy x is said to be an evolutionary stable strategy (ESS) if $\forall y \exists \varepsilon_y > 0$ s.t. when $0 < \varepsilon < \varepsilon_y$:

$$u(x, (1 - \varepsilon)x + \varepsilon y) > u(y, (1 - \varepsilon)x + \varepsilon y) \quad (7.2)$$

Now we can show that:

Proposition 26. *The unique Nash strategy of G is an evolutionary stable strategy (ESS).*

Proof. This proof is simply a restatement of the well-known result that local superiority implies evolutionary stability. Firstly, G has a unique Nash strategy by definition. Call this strategy α_i for player i . Now we know that $u(\alpha_i | a_j = \alpha_j) > u(a_i | a_i \neq \alpha_i, a_j = \alpha_j)$ so by the uniqueness of α_i we know that any mix of α_i with $a_i \neq \alpha_i$ will reduce $u(\alpha_i, a_j)$ where $u(a_i, a_j)$ is the payoff to player i from action $a_i \in A$ given player j plays $a_j \in A$. Therefore the Nash equilibrium of G must be strict. By strictness we know that $u(\alpha_i, a_j) > u(\beta_i, a_j)$ where $\beta_i \neq \alpha_i$. This in turn implies local superiority, so:

$$\lim_{\varepsilon \rightarrow 0} \{(1 - \varepsilon)u(\alpha_i, \alpha_j) + \varepsilon u(\alpha_i, \beta_j)\} > \lim_{\varepsilon \rightarrow 0} \{(1 - \varepsilon)u(\beta_i, \alpha_i) + \varepsilon u(\alpha_i, \beta_j)\}$$

By linearity of expected utility in probabilities this implies:

$$u(\alpha_i, (1 - \varepsilon)\alpha_j + \varepsilon\beta_j) > u(\beta_i, (1 - \varepsilon)\alpha_j + \varepsilon\beta_j) \text{ for } \varepsilon \rightarrow 0$$

Which is simply a restatement of the definition of an ESS given in definition 11. \square

We have a simple notion of a population of players playing Nash against each other and performing well in terms of their payoffs. We now consider the mutation of a proportion γ of the population into *neural network players*. By proposition 26 we know the remaining population will continue to play Nash strategies, if we let $\gamma \rightarrow 0$. We can retain this property by letting $Q \rightarrow \infty$ and setting γ to be fixed at a number strictly above zero, but finite. In particular, we can consider a single member of the population to be a

neural network player, but let $Q \rightarrow \infty$ to bring γ arbitrarily close to zero. We can now examine the actions of this single neural network player content in the knowledge that all other players will continue to play Nash strategies. Therefore, we can be assured that the neural network's best reply to the population will be to play a Nash strategy.

7.4 THE NEURAL NETWORK PLAYER

Let us start with an intuitive account of the single *neural network player* in G . Consider a young economic agent with no prior experience of play in any game. This agent will, however, have a base of experience derived from a prior period spent learning “how to play”. We might imagine a student's time at school or observation of how older members of the population play certain games. This agent has a store of observed example games, none of which will necessarily fit exactly with any game he will face in the future, but which might share certain similarities. We capture this notion of learning by example prior to entry into the economy, or marketplace of games, through the use of a *neural network*. We first “train” the agent by subjecting him to a series of example games with given actions and payoffs, we require him to be a utility maximizer, and then use a form of backpropagation to allow him to develop a form of pattern recognition which will enable him to “learn by example”, and attempt to learn to play the Nash strategy in order to maximize his payoff. The question we ask is: can we introduce the network to a marketplace modelled by a sequence of random games filled with Nash players and expect the network to learn to play Nash? The key point is that he will be most unlikely to ever play the same game twice and will be most unlikely ever to have seen the identical game in his training period, so will his pattern recognition abilities be sufficient for him to intuitively recognize the Nash strategy in an entirely new game and play it? We are now switching out of the evolutionary framework and focusing on the play of a single player rather than the behaviour of the population in general. We first need to add some formal definitions.

7.4.1 Defining the Network

Consider a neural network, or more simply C to be a machine capable of taking on a number of states, each representing some computable functions mapping from input space to output space, with two *hidden* layers of further computation between input and output.²⁹ The following definition formalizes this.

Definition 12. *Define the neural network as $C = \langle \Omega, X, Y, F \rangle$ where Ω is a set of states, $X \subseteq \mathbb{R}^n$ is a set of inputs, Y is a set of outputs and $F : \Omega \times X \mapsto Y$ is a parameterized function. For any ω the function represented by state ω is $h_\omega : X \mapsto Y$ given by $h_\omega(x) = F(\omega, x)$ for an input $x \in X$. The set of functions computable by C is $\{h_\omega : \omega \in \Omega\}$, and this is denoted by H_C .*

Put simply, when the network, C , is in state ω it computes the function h_ω providing it is computable. In order to reasonably produce answers which correspond to a notion of correctness (in this case the unique Nash strategy in a 3×3 game), we need to *train* the network. First we will consider the form of the network in practice, and start by defining an activation function.

Definition 13. *An activation function for node i of layer k in the neural network C is of the logistic form*

$$a_i^k = \frac{1}{1 - \exp\left(-\sum_j w_{ij}^k u_{ij}^{k-1}\right)} \quad (7.3)$$

where u_{ij}^k is the output of node j in layer $k - 1$ sent to node j in layer k , and w_{ij} is the weight attached to this by node i in layer k . The total activation flowing into node i , $\sum_j w_{ij}^k u_{ij}^{k-1}$, can be simply defined as t_i .

Consider a set of 18 input nodes each recording and producing as output a different value from the vector $x_k = (x_k^1, \dots, x_k^{18})$. This neatly corresponds to the payoffs of a 3×3

²⁹Hidden layers can be thought of as intermediate layers of computation between input and output. Since we see the input go in, and the output come out, but do not directly see the activity of intermediate layers, they are in some sense *hidden*.

game. Now consider a second set of 36 nodes (the first hidden layer). Each node in this second layer receives as an input the sum of the output of all 18 input nodes transformed by the activation function of node i in layer 2. All of the nodes in the second layer send this output a_i^2 to all nodes in the second hidden layer, which weights the inputs from all i of the first hidden layer, by the activation function to produce a_i^3 . These numbers are sent to the final layer of two nodes to produce an output y which forms a 2-dimensional vector which represents the choice of strategy in a 3×3 game. To explain this representation of a strategy in a 3×3 game for the row player, the vector $(1, 0)$ would imply the pure Nash strategy is the top row $(0, 1)$ would imply the middle row, and $(0, 0)$ the bottom row.

7.4.2 Training the Network

Training essentially revolves around finding the set of weights that is most likely to produce the desired output. During training C receives a sequence of random games until some stopping rule determines the end of the training at some round T (discussed below). The *training sample* consists of M random games. If $T > M$, then (some or all of) the random games in M will be presented more than once.

Let us formally define the training sample. The vector $x_k = (x_k^1, \dots, x_k^{18})$ consists of 18 real-valued numbers drawn from a uniform $(0, 1)$ representing the payoffs of a 3×3 game. It is recorded in the first set of input nodes, and then sent and transformed by the two hidden nodes before an output y , a two-dimensional vector, is produced and represented in the final layer. This is repeated M times with a new set of inputs x_k and outputs y_k . Assume that each vector x_k is chosen independently according to a fixed probability distribution P_T on the set X . The probability distribution is fixed for a given learning problem, but it is unknown to C , and for our purposes will be taken to be a uniform $(0, 1)$. The information presented to C during training therefore consists only of several sequences of numbers.

Definition 14. For some positive integer m , the network is given a training sample:

$$x^M = ((x_1^1, \dots, x_1^{18}), (x_2^1, \dots, x_2^{18}), \dots, (x_M^1, \dots, x_M^{18})) = (x_1, x_2, \dots, x_M) \in X^M$$

The labelled examples x_i are drawn independently according to the probability distribution P_T . A random training sample of length M is an element of X^M distributed according to the product probability distribution P^M .

Assume that $T > M$. In this case, training might be *sequential*: after $q \times M$ rounds (for any positive integer q s.t. $q \times M < T$), M is presented again, exactly in the same order of games. If training is *random with replacement*, it is less restricted to the extent that the order in which the random games are presented each time is itself random. If training is *random without replacement*, in each round the network is assigned randomly one of the random games in M , until round T .

Having selected a sample sequence of inputs, x , and determined the unique Nash strategy associated with each, α , we need to consider how C learns the relationship between the two, to ensure that its output y will approach the Nash strategy. The method used is backpropagation. First let us define the error function.

Definition 15. Define the network's root mean square error ε as the root mean square difference between the output y and the correct answer α over the full set of $q \times M$ games where individual games are indexed by i , so our error function is:

$$\varepsilon \equiv \left(\sum_{i=1}^{q \times M} (y_i - \alpha_i)^2 \right)^{\frac{1}{2}}$$

The aim is to minimize the error function by altering the set of weights w_{ij} of the connections between a typical node j (the sender) and node i (the receiver) in different layers. These weights can be adjusted to raise or lower the importance attached to certain inputs in the activation function of a particular node. Backpropagation is a form

of numerical analysis akin to gradient descent search in the space of possible weights. Following Rumelhart, Hinton, and Williams (1986) we use a function of the form:

$$\Delta w_{ij} = -\eta \frac{\partial \varepsilon}{\partial w_{ij}} = \eta k_{ip} o_j \quad (7.4)$$

where w_{ij} is simply the weight of the connection between the sending node j and receiving node i . As ε is the neural network's error, $\partial \varepsilon / \partial w_{ij}$ measures the sensitivity of the neural network's error to the changes in the weight between i and j . There is also a *learning rate* given by $\eta \in (0, 1]$ which is a parameter of the learning algorithm and must not be chosen to be too small or learning will be particularly vulnerable to local error minima. Too high a value of η may also be problematic as the network may not be able to settle on any stable configuration of weights. Define $\partial \varepsilon / \partial w_{ij} = -k_{ip} o_{jp}$ where o_{jp} is the degree of activation of the sender node o_{jp} . The higher o_{jp} is, the more the sending node is at fault for the erroneous output, so it is this node we wish to correct more. k_{ip} is the error on unit i for a given input pattern p , multiplied by the derivative of the output node's activation function given its input. Calling g_{ip} the goal activation level of node i for a given input pattern p , in the case of the output nodes k_{ip} can be computed as:

$$k_{ip} = (g_{ip} - o_{ip}) f'(t_{ip}) = (g_{ip} - o_{ip}) o_{ip} (1 - o_{ip}) \quad (7.5)$$

since the first derivative $f'(t_{ip})$ of the receiving node i in response to the input pattern p is equal to $o_{ip}(1 - o_{ip})$ for a logistic activation function. Now assume that a network has N layers, for $N \geq 2$. As above, we call layer 1 the input layer, 2 the layer which layer 1 activates (the first hidden layer), and so on, until layer N the output layer which layer $N - 1$ activates.

We can now define the backpropagation learning process.

Definition 16. *Using backpropagation, we first compute the error of the output layer (layer N) using equation 7.5, and update the weights of the connections between layer N and $N - 1$, using equation 7.4. We then compute the error to be assigned to each node*

of layer $N - 1$ as a function of the sum of the errors of the nodes of layer N that it activates. Calling i the hidden node, p the current pattern and β an index for each node of layer N (activated by i), we can use:

$$k_{ip} = f'(t_{ip}) \sum_{\beta} k_{\beta p} w_{\beta i} \quad (7.6)$$

to update the weights between layer $N - 1$ and $N - 2$, together with equation 7.4. We follow this procedure backwards iteratively, one layer at a time, until we get to layer 1, the input layer. A variation on standard backpropagation would involve replacing equation 7.4 with a momentum function of the form:

$$\Delta w_{ij}^t = -\eta \frac{\partial \varepsilon^t}{\partial w_{ij}^t} + \mu \Delta w_{ij}^{t-1} \quad (7.7)$$

where $\mu \in [0, 1)$ and $t \in \mathbb{N}^{++}$ denotes the time index (an example game, vector x , is presented in each t during training). Momentum makes connection changes smoother by introducing positive autocorrelation in the adjustment of connection weights in consecutive periods. The connection weights of the network are updated using backpropagation until round T . T itself can be determined exogenously by the researcher, or it can be determined endogenously by the training process, i.e. training may stop when the network returns the correct output with ε lower than a given target value.

7.5 INADEQUATE LEARNING

We have now a clear idea of what the neural network is and the game that the neural network will face. The training set is simply a sequence of vector pairs (x, y) where the inputs $x \in X$ correspond to the set of actions A_i for N players in M random games, and the outputs to the payoffs $u_i : A \mapsto \mathbb{R}$ for N players for each of the actions. We set $M = 2000$, $N = 2$ and restrict the action set by assuming a 3×3 normal form game. This restriction is done without loss of generality: potentially any finite normal form could be modelled in a similar way, while 2×2 , 2×3 and 3×2 games count as a

subclass of 3×3 .³⁰ We then allow the network to play 2000 further random games never encountered before, selecting a single input and recording a single output with the aim of maximizing the payoff in each game. Since we force each game to contain a unique Nash equilibrium in pure strategies and we restrict the network's choice to be in pure strategies, we can then check the network's *success rate* as defined by the proportion of times the network selected the Nash strategy to within a given threshold of mean squared error (as defined in definition 15). For example if the correct output is $(1, 0)$ and the neural network returns $(0.99, 0)$ it easily meets an $\varepsilon = 0.05$ threshold).

7.5.1 Incomplete Neural Network Learning

We now need to examine how well a neural network can do in theory, and fortunately various results exist in the literature to which we can refer. One of the most well-known results comes from Hornik, Stinchcombe, and White (1989) reprinted in White (1992). Hornik, Stinchcombe, and White (1989) show that standard *feedforward* networks with only a single hidden layer can approximate any continuous function uniformly on any compact set and any measurable function arbitrarily well in the p_μ metric, which they define as follows (in slightly amended form):

Definition 17. (Hornik, Stinchcombe, and White (1989)). Let \mathbb{R} be the set of real numbers, and \mathbb{B}^r the Borel σ -field of \mathbb{R}^r . Let K^r be the set of all Borel measurable functions from \mathbb{R}^r to \mathbb{R} . Given a probability measure μ on $(\mathbb{R}^r, \mathbb{B}^r)$ define the metric p_μ from $K^r \times K^r$ to \mathbb{R}^+ by $p_\mu(f, g) = \inf \{ \varepsilon > 0 : \mu \{ x : |f(x) - g(x)| > \varepsilon \} < \varepsilon \}$.

The main result, summarized in theorem 2.4 in Hornik, Stinchcombe, and White (1989), effectively concerns the *existence* of a set of weights which allow the perfect

³⁰While the neural network model was designed with 3×3 games in mind, since all payoff values are drawn from a uniform $[0, 1]$, it is straightforward to train the current network on 2×2 games, by putting zeros on all the values in the third row and third column of each game. Similarly we could construct 2×3 and 3×2 games by placing zeros in the appropriate row or column. The choice of a 3×3 is therefore much more general than the choice of 2×2 , and can include several interesting and well-known 2×2 games.

emulation of the algorithm that the neural network is attempting to learn. There are three potential areas for failure:

- (1) *Inadequate learning*, or a failure of the learning dynamic to reach the global error-minimizing algorithm.
- (2) *Inadequate network size*, or insufficient hidden units.
- (3) The presence of a stochastic rather than a deterministic relation between input and target.

Problem 3 can be extended to include poly-random functions (which cannot be distinguished from random functions by any polynomial approximation) but is still not a problem for the class of normal form games G . Problem 2 introduces a parallel mechanism for examining bounded-rational behaviour. Along similar lines to the automata literature, we might restrict the number of hidden units in order to force bounded-rational behaviour upon our player.³¹ However, regardless of any attempts to raise the network size to cover for any potential problems, we cannot reasonably deal with problem 1, and it is the problem of inadequate learning and the nature of the learning algorithm which is the focus of the rest of this section and indeed the whole chapter.

7.5.2 Learning and Learnability

A *learning algorithm* takes random training samples and acts on these to produce a hypothesis $h \in H$ that, provided the sample is large enough is with probability at least $1 - \delta$, ε -good (with ε defined as above) for P_T . It can do this for each choice of ε , δ and P_T . To define this more formally:

Definition 18. *Suppose that H is a class of functions that map $X \mapsto Y$. A learning algorithm L for H is a function $L : \cup_{M=1}^{\infty} Z^M \mapsto H$ from the set of all training samples to H , with the following property: for any $\forall_{p \in (0,1), \delta \in (0,1)} \exists$ an integer $M_0(\varepsilon, \delta)$ s.t. if*

³¹Abreu and Rubinstein (1988) or Neyman (1985) provide examples of finite automata used in this way. In particular, the finite states of the automata proxy for the limited memories of bounded-rational players.

$M \geq M_0(\varepsilon, \delta)$ then, for any probability distribution P_T on $Z = X \times Y$, if z is a training sample of length M drawn randomly according to the product distribution P^M , then, with probability at least $1 - \delta$, the hypothesis $L(z)$ output by L is such that $er_p(L(z)) < opt_p(H) + \varepsilon$. More compactly, for $M \geq M_0(\varepsilon, \delta)$, $P^M \{er_p(L(z)) < opt_p(H) + \varepsilon\} \geq 1 - \delta$.

To restate in slightly different terms we can define a function L as a learning algorithm, if \exists a function $\varepsilon_0(M, \delta)$ s.t. \forall_{M, δ, P_T} , with probability at least $1 - \delta$ over $z \in Z^M$ chosen according to P^M , $er_p(L(z)) < opt_p(H) + \varepsilon_0(M, \delta)$, and $\forall_{\delta \in (0, 1)}$, $\varepsilon_0(M, \delta) \rightarrow 0$ as $M \rightarrow \infty$. This definition stresses the role of $\varepsilon_0(M, \delta)$ which we can usefully think of as an estimation error bound for the algorithm L . A closely related definition is:

Definition 19. *We say that H is learnable if \exists a learning algorithm for H .*

The function h_w can be thought of as representing the entire processing of the neural network's multiple layers, taking an input vector x and producing a vector representation of a choice of strategy. Over a large enough time period we would hope that C will return a set of optimal weights which will in turn produce the algorithm h_w which will select the Nash strategy if \exists a learning algorithm for selecting Nash equilibria (H in this case). Or alternatively if we wish to attain some below perfect success rate, we can do so using a finite training sample, and the success rate will grow as the number of examples increases. This all crucially rests on the ability of backpropagation to pick out the globally error-minimizing algorithm for finding Nash equilibria.³² This now allows us to tightly define the learning problem faced by C .

Definition 20. *C , using the learning algorithm given by definition 16, faces a training sample of size $M \times q$. The Nash problem is to find an algorithm as defined in definition 18 for which $\varepsilon_0(M, \delta) \rightarrow 0$ as $M \rightarrow \infty$ where the error function ε is as defined in definition 15.*

³²The exact algorithm for calculating the unique pure Nash strategy in a 3×3 normal form game with a single pure Nash equilibrium is given in the appendix to the chapter.

7.5.3 Finding the Global Minimum

Having established the problem we now need to verify that the algorithm which successfully collapses $\varepsilon_0(M, \delta)$ to zero is indeed learnable. While backpropagation is undoubtedly one of the most popular search algorithms currently used to train feedforward neural networks, it is a gradient descent algorithm and therefore this approach leads only to a local minimum of the error function (see for example, Sontag and Sussmann (1989). White (1992), p. 160, makes the point: "...Hoornik *et al* (1989) have demonstrated that sufficiently complex multilayer feedforward networks are capable of arbitrarily accurate approximations to arbitrary mappings... An unresolved issue is that of "learnability", that is whether there exist methods allowing the network weights corresponding to these approximations to be learned from empirical observation of such mappings." White (1992), chapter 9, theorem 3.1 provides a theorem which summarizes the difficulties inherent in backpropagation: he proves that backpropagation can get stuck at local minima or saddle points, can diverge, and cannot even be guaranteed to get close to a global minimum. Generally, however, this is hardly surprising as backpropagation, for all its biological plausibility, is after all a gradient descent algorithm.

The problem is exacerbated in the case of our neural network C as the space of possible weights is so large. Auer, Herbster, and Warmuth (1996) have shown that the number of local minima for this class of networks can be exponentially large in the number of network parameters. Sontag (1995) gives upper bounds for the number of such local minima, but the upper bound is unfortunately not tight enough to lessen the problem. In fact as the probability of finding the absolute minimizing algorithm (the Nash algorithm) is likely to be exponentially small, the learning problem faced by C falls into a class of problems known in algorithm complexity theory as *NP*-hard.

7.6 ALGORITHM COMPLEXITY AND INTRACTABILITY

To fully understand the concept of *NP*-hardness first requires a primer in algorithm complexity theory. This section provides such a primer, and then moves on to give a

proposition concerning the intractability of the Nash problem given in definition 20.³³

7.6.1 A Formal Treatment

We will start with one of the most standard machine intelligences in computer science, the *deterministic (one-tape) Turing machine* (or DTM). This is finite in the set of states (say Q), including one distinguished start state (q_0), and two halt states (q_Y and q_N), each corresponding to a “yes” or “no” answer to a decision-problem. The DTM will halt after it comes to a decision, either q_Y or q_N . We next need to consider a *language*.

Definition 21. *For any finite set Σ of symbols, denote by Σ^* the set of all finite strings of symbols from Σ . If Λ is a subset of Σ^* , we say that Λ is a language over the alphabet Σ .*

Call an abstract decision-problem Π , with a set D_Π of instances and a subset $Y_\Pi \subseteq D_\Pi$ of yes-instances. The correspondence between decision-problems and languages is brought about by *encoding schemes* used for specifying the problems instances whenever we intend to compute them.

Definition 22. *An encoding scheme e for a problem Π provides a way of describing each instance of Π by an appropriate string of symbols over some fixed alphabet Σ .*

Thus the problem Π and the encoding scheme e for Π partition Σ^* into three classes of strings: those that are not encodings of instances of Π , those that are encodings of instances of Π for which the answer is “no”, and those that encode instances of Π for which the answer is “yes”. This third class of strings is the language we associate with Π and e , setting:

$$\Lambda[\Pi, e] = \left\{ \sigma \in \Sigma^* : \begin{array}{l} \Sigma \text{ is the alphabet used by } e, \text{ and } \sigma \text{ is the} \\ \text{encoding under } e \text{ of an instance of } I \in Y_\Pi \end{array} \right\}$$

³³For even more detail see the classic Garey and Johnson (1979), or more recently Cormen, Leiserson, and Rivest (1992).

If a result holds for the language $\Lambda[\Pi, e]$, then it holds for the problem Π under the encoding scheme e .

Returning to our DTM, we say that a DTM program or algorithm L with input alphabet Σ accepts $\sigma \in \Sigma^*$ if and only if L halts in state q_Y when applied to input σ . We say a language Λ_M recognized by the program M is given by:

$$\Lambda_M = \{\sigma \in \Sigma^* : L \text{ accepts } \sigma\}$$

Now we are ready to consider the first complexity class, P :

$$P = \{\Lambda : \text{there is a polynomial time DTM program } L \text{ for which } \Lambda = \Lambda_M\}$$

This requires some explanation. Polynomial time programs *run* in polynomial time, that is a program or algorithm of inputs of size n is bounded by a running time of $O(n^k)$ for some finite constant k . This is clearly restrictive in the sense that exponential time would provide a far more generous bound for large n and any give k , but in all other senses seems a generous notion of time. Returning to the set P we will say that a decision-problem Π belongs to P under the encoding scheme e if $\Lambda[\Pi, e] \in P$, that is if there is polynomial time DTM program M that “solves” Π (i.e. reaches a decision) under encoding scheme e . It is standard within mathematical complexity theory and computer science to omit further references to e as there is a high level of equivalence between most encoding schemes.

Now consider a *non-deterministic Turing machine* (NDTM). This differs from a DTM in one major way. The DTM essentially follows a program L (or algorithm) in an attempt to find a solution (decision) for the problem Π . However, a NDTM can instead “guess” a solution, and then follow a stage similar to the DTM’s program stage, by attempting to verify whether the guess was correct. The NDTM program is defined in the same way starting with q_0 and continuing until a halting stage is reached, but this time the aim is to verify (or refute) the guess rather than record a solution. The NDTM is said to record an *accepting computation* if it ends with the outcome q_Y , so the guess is verified. All

the other notation from the DTM carries over, including the notion of languages, and recognition. In a similar way we can therefore define the complexity class, NP :

$$NP = \{\Lambda : \text{there is a polynomial time NDTM program } L \text{ for which } \Lambda_M = \Lambda\}$$

It should be noted that $P \subseteq NP$ since any deterministic algorithm L could be used as the checking algorithm in a non-deterministic computation. So $\Pi \in P \Rightarrow \Pi \in NP$. The relationship in the other direction is somewhat more complex. It is widely believed that $P \neq NP$, so NP is a larger set than P , but this has yet to be proven. The rationale behind what is a standard belief in computer science and mathematical complexity theory comes about because of the existence of a number of problems belonging to the theorized set $NP - P$. Consider the abstract problem Π^* , which cannot be solved by a DTM in polynomial time, so $\Pi^* \notin P$, and consider such a problem to be solvable in polynomial time by a NDTM, so $\Pi^* \in NP$. Furthermore, let us strengthen this by defining Π^* to be the *hardest possible problem* solvable in polynomial time by a NDTM, but not solvable by a DTM. Hardness in this sense has a very simple meaning: every other problem in NP can be polynomially reduced to Π^* . Therefore, if Π^* could be solved by a DTM then so could every other problem in NP and NP would collapse to P . In this way we could consider the set containing Π^* to be the set of the hardest problems in NP , so hard that if solvable by a DTM in polynomial time we could say without hesitation that $NP = P$; we call this set the set of NP -complete problems. We have one member of NP -complete, but there is no reason to believe there would only be one such hard problem; it might be that there exist many problems which have the property that their solution by a DTM would collapse the NP set. As it happens there are literally dozens of such problems, and Garey and Johnson (1979) list numerous examples. Although a major research effort exists in computer science to attempt to solve these polynomial-equivalent problems, none have ever been solved by a DTM in polynomial time, which is why the concept of NP -completeness is such a strong one within complexity theory.

The last concept we need is that of *NP*-hardness. We say something is *NP*-hard if it is equivalent in hardness to a member of the set of *NP*-complete problems. So it is a member, or rewritten version of a member, of the set of *NP*-complete problems. Perhaps a simpler way of saying this is that the problem is essentially intractable.

7.6.2 *NP*-hardness

Finally we now have the tools we need to deal with the learnability of the problem faced by C . Theorem 25.5 from Anthony and Bartlett (1999) succinctly states the following (in a slightly amended form):

Theorem 4. (*Anthony and Bartlett, 1999*). *The problem given in definition 20 faced by the class of networks encompassing C is *NP*-hard.*

Anthony and Bartlett (1999) chapters 23 to 25, provides various forms of this theorem for different types of network including the feedforward class of which C is a member. The following proposition is simply a restatement of theorem 4 with reference to the particular problem faced by C .

Proposition 27. *C supplemented by the backpropagation learning dynamic will not be able to learn the Nash algorithm in polynomial time.*

Proof. Backpropagation as a form of gradient descent is a DTM-algorithm. Therefore if the problem to be faced is *NP*-hard, by the definition of *NP*-hardness any DTM will therefore fail to find the minimizing algorithm L in polynomial time. A direct application of theorem 4 demonstrates that the problem is in fact *NP*-hard and we have our proof. \square

Gradient descent algorithms attempt to search for the minimum of an error function, and backpropagation is no exception. However, given the prevalence of local minima, a DTM cannot consistently solve the problem in definition 18 and find an absolute minimum. The basins of attraction surrounding a local minimum are simply too strong for a simple gradient descent algorithm to escape, so looking back to definition 18 we cannot expect $\varepsilon_0(M, \delta) \rightarrow 0$ as $M \rightarrow \infty$, and in turn we cannot consider the task facing the

network to be *learnable* in the sense of definition 19.³⁴ However, if we were to supplement the algorithm with a guessing stage, i.e. add something akin to grid search or one of several theorized additions or alternatives to backpropagation, then we might expect to find the absolute minimum in polynomial time.³⁵ To restate this in terms of the search for an algorithm capable of providing Nash equilibria in never before seen games, backpropagation cannot do this perfectly, while other far less biologically plausible methods involving processor hungry guess and verify techniques, can produce consistent results.

So our player will find a decision-making algorithm that will retain some error even at the limit, or to put this an alternative way, we may have to be content with an algorithm which is effective in only a subclass of games, so it optimizes network parameters only in a small subspace of the total space of parameters. In the case of normal form games we can summarize this section as: *our player will almost surely not learn the globally error-minimizing algorithm for selecting Nash equilibria in normal form games.* However, we can reasonably assume that some method will be learned, and this should at least minimize error in some subset of games corresponding to the domain of some local error-minimizing algorithm.

7.7 RULES OF THUMB

This brief section attempts to map the intractability and learning problems of a neural network into an alternative phrase which has a long tradition in economics: *rules of*

³⁴This is intuitively seen to be reasonable with reference to two results. Fukumizu and Amari (2000) shows that local minima will always exist in problems of this type, and Auer, Herbster, and Warmuth (1996) show that the number of local minima for this class of networks is exponentially large in the number of network parameters. In terms of the theory of *NP*-completeness, we might say the solution could be found in exponential time, but not in polynomial time. For any network with a non-trivial number of parameters, such as *C*, the difference is great enough for the term intractable to be applied to such problems.

³⁵White (1992), chapter 10, discusses a possible method which can provide learnability, using an application of the theory of sieves (see Grenader (1981)). However, White (1992), p. 161, stresses: "The learning methods treated here are extremely computationally demanding. Thus, they lay no claim to biological or cognitive plausibility." The method is therefore useful for computing environments and practical applications, rather than the modelling of decision-making. In other words his method is an application of an NDTM method not a DTM method.

thumb. The main tool to be used is the concept of a local error-minimizing algorithm.

7.7.1 Local Error-Minimizing Algorithms

Given that backpropagation will find a local minimum, but will not readily find an absolute minimizing algorithm in polynomial time, we are left with the question, what is the best our neural network player can hope to achieve? If we believe the neural network with a large, but finite training set nicely models bounded-rational economic agents, but cannot flawlessly select Nash strategies with no prior experience of the exact game to be considered, this question becomes: what is the best a bounded-rational agent can hope to achieve when faced with a population of fully rational agents?

In terms of players in a game, we have what looks like bounded-rational learning or satisficing behaviour: the player will learn until satisfied that he will choose a Nash equilibrium strategy sufficiently many times to ensure a high payoff. We label the outcome of this bounded-rational learning as a local error-minimizing algorithm (LMA).

More formally, consider the learning algorithm L , and the ‘gap’ between perfect and actual learning, $\varepsilon_0(M, \delta)$. Recall that Z^M defines the space of possible games as perceived by the neural network.

Definition 23. *If \exists a function $\varepsilon_0(M, \delta)$ s.t. \forall_{M, δ, P_T} , with probability at least $1 - \delta$ over all $z \in Z^M$ chosen according to P^M , $er_p(L(z)) < opt_p(H) + \varepsilon_0(M, \delta)$, and $\forall_{\delta \in (0, 1)}$, $\varepsilon_0(M, \delta) \rightarrow 0$ as $M \rightarrow \infty$ then this function is defined the global error-minimizing algorithm (GMA).*

This simply states that for all possible games faced by the network, after sufficient training, the function L will get arbitrarily close to the Nash algorithm given in the appendix, collapsing the difference to zero. This clearly requires an algorithm sufficiently close to Nash to pick a Nash equilibrium strategy in almost all games.

Definition 24. *A local error-minimizing algorithm (LMA) will select the same outcome as a global error-minimizing algorithm for some $z \in Z^M$, but will fail to do so for all $z \in Z^M$.*

LMAs can be interpreted as examples of rules of thumb that a bounded-rational agent is likely to employ (for example, Simon (1955) and Simon (1959)). They differ from traditionally conceived rules of thumb in two ways. First, they do select the best choice in some subset of games likely to be faced by the learner. Second, they are learned *endogenously* by the learner in an attempt to maximize the probability of selecting the best outcome. The ‘best’ outcome can be determined in terms of utility maximization or a reference point, such as the Nash equilibrium.

7.7.2 Rules of Thumb: A Simple Example

Recall the experimental design from chapter 6. The last chapter finds a rule of thumb, which provides optimal actions in the game faced by the subjects of the endogenous-time herding experiment detailed in that chapter: *follow your own signal unless it is inconclusive; in that case wait and follow the majority.*

This rule is not a global optimizer, and would clearly fail in many other settings, but in a subclass of games this will provide for a high payoff relative to the global optimizer which is to apply full Bayesian updating upon the observation of any action (or inaction) taken by other players in an endogenous-time herding game. The next chapter focuses on testing some of these ideas for a class of normal form games, but there is nothing that restricts the fact that the best available model of the human brain is likely to be happy with what might appear to be a rule of thumb.

It should be noted that the nature of neural network learning does not lend credence to the argument that any rule of thumb is reasonable. A rule of thumb will only be reasonable under the conditions of this chapter if such a rule of thumb is a local minimizer, and this is strictly defined in terms of minimizing error in a strict locality of the space of all possible games. In this sense many *ad hoc* rules of thumb would not be reasonable resting places for a neural network learning algorithm. To emphasize: *only rules of thumb which provide close to optimal behaviour in a subclass of games likely to be faced by the player during his formative period of play, are possible resting places for his long-term behaviour.*

7.8 CONCLUSIONS

As we have seen neural networks can be used to model bounded-rational behaviour in a normal form game framework. Potentially any finite normal form could be modelled in this way, though the chapter concentrated on 3×3 games, noting that 2×2 , 2×3 and 3×2 games count as a subclass of 3×3 . The inclusion of a neural network player in a population of Nash players would not change the behaviour of the Nash players, and the neural network player, having seen a sufficiently large sample of example games in which the Nash outcome was high-lighted, *could* potentially learn the Nash algorithm.

However, the task the neural network player faces heavily relies on the nature of the learning algorithm it uses to find the weights necessary to emulate the Nash algorithm. Following a gradient descent search algorithm like backpropagation results in the network facing an *NP*-hard problem. Loosely speaking this problem can be considered to be as hard as a number of standard (*NP*-complete) problems that are thought to be unsolvable by a polynomial time algorithm. Though a set of weights exists, as ensured by the existence theorems of Hornik, Stinchcombe, and White (1989) and others, this set of weights provides just one in a number of possible locally error-minimizing sets of weights which grows exponentially with the number of network parameters. In the language of this chapter, a neural network is likely to find some locally error-minimizing algorithm, getting sufficiently close to the Nash algorithm in a large enough number of cases to leave the network satisfied that it has found a suitable way of playing new games. Once this occurs we would expect the network to reach a good level of success in finding a Nash equilibrium in a never before seen game, though it would not achieve 100% success. So our neural network will learn, but will only ever achieve what might best be described as bounded rationality.

APPENDIX FOR CHAPTER 7

The Nash Algorithm

This appendix examines the algorithm which the neural network is attempting to learn: the Nash algorithm for 3×3 normal form games with a unique Nash strategy.

Consider a 2 player 3×3 normal form game G with payoff matrix X ,

$$X = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ x_7 & x_8 & x_9 & x_{10} & x_{11} & x_{12} \\ x_{13} & x_{14} & x_{15} & x_{16} & x_{17} & x_{18} \end{pmatrix}$$

Where payoffs are in pairs, so x_1 and x_2 are the payoffs to players 1 and 2 respectively from the realized outcome (top, left). Strategies are $a_i \in \{a_i^1, a_i^2, a_i^3\}$ for players $i = 1, 2$. More specifically, for the row player ($i = 1$), $a_1^1 = \text{"top"}$, $a_1^2 = \text{"middle"}$, and $a_1^3 = \text{"bottom"}$, and for the column player ($i = 2$), $a_2^1 = \text{"left"}$, $a_2^2 = \text{"middle"}$ and $a_2^3 = \text{"right"}$. Now let the vector y denote the choice of strategy for player 1 so $y = (1, 0) \Leftrightarrow a_1 = a_1^1$, $y = (0, 1) \Leftrightarrow a_1 = a_1^2$ and $y = (0, 0) \Leftrightarrow a_1 = a_1^3$. Now express the unique Nash strategy for player 1 as y^* .

The Nash algorithm for 3×3 normal form games with a unique Nash strategy is then:

$$y^* = (1, 0) \text{ if } \begin{cases} (x_1 > \max\{x_7, x_{13}\} \ \& \ x_2 > \max\{x_4, x_6\}) & \text{or} \\ (x_3 > \max\{x_9, x_{15}\} \ \& \ x_4 > \max\{x_2, x_6\}) & \text{or} \\ (x_5 > \max\{x_{11}, x_{17}\} \ \& \ x_6 > \max\{x_2, x_4\}) \end{cases} \quad (7.8)$$

$$y^* = (0, 1) \text{ if } \begin{cases} (x_7 > \max\{x_1, x_{13}\} \ \& \ x_8 > \max\{x_{10}, x_{12}\}) & \text{or} \\ (x_9 > \max\{x_3, x_{15}\} \ \& \ x_{10} > \max\{x_8, x_{12}\}) & \text{or} \\ (x_{11} > \max\{x_5, x_{17}\} \ \& \ x_{12} > \max\{x_8, x_{10}\}) \end{cases} \quad (7.9)$$

$$y^* = (0, 0) \text{ if } \begin{cases} (x_{13} > \max\{x_1, x_7\} \ \& \ x_{14} > \max\{x_{16}, x_{18}\}) \ \text{or} \\ (x_{15} > \max\{x_3, x_9\} \ \& \ x_{16} > \max\{x_{14}, x_{18}\}) \ \text{or} \\ (x_{17} > \max\{x_5, x_{11}\} \ \& \ x_{18} > \max\{x_{14}, x_{16}\}) \end{cases} \quad (7.10)$$

The restrictions on G ensure that $y^* \neq \emptyset$ and at most one of expressions 7.8, 7.9 and 7.10 are true.

CHAPTER 8

NEURAL NETWORKS IN PRACTICE

8.1 INTRODUCTION

This chapter moves forward from chapter 7 by presenting a simulated neural network model that plays (up to) 3×3 games.³⁶ The simulation of a working neural network allows us to check whether, after training on games with unique pure Nash equilibria, the network remains entirely naive in playing never before seen games, whether it learns to play Nash, or, as was argued in the last chapter, whether the network learns some *local error-minimizing algorithm* (or LMA) allowing it to perform well, but not perfectly, when attempting to select Nash equilibria.

The basic proposition of the last chapter was that a set of theoretical weights exist which if located by a neural network's search algorithm, could in theory allow it to learn to play never before seen games in a fully rational way, picking Nash equilibria at the very first attempt. However, we saw that the backpropagation training process used by the network finds it too hard (in the language of algorithm complexity theory it finds it *NP-hard*) to search for global error-minimizing sets of weights; rather it must content itself with a *local error-minimizing algorithm*. Such an algorithm, if followed, will produce outcomes which are equivalent to those produced by the *globally error-minimizing algorithm* in a subset of cases. While it may be too much to expect a neural

³⁶This chapter is based on the second half of Zizzo and Sgroi (2000). Daniel Zizzo is to be especially thanked for much of the work detailed here, in particular the design and execution of the simulations and econometric tests which make up the bulk of the chapter.

network to be able to find the Nash algorithm in any game after training, it will get close, and its decision-making may provide us with a good model of bounded-rational play.

Locally error-minimizing algorithms, as detailed in the last chapter (see definition 24) are more tightly defined than *rules of thumb* in two very specific ways:

(1) They are learnt endogenously by the learner in an attempt to maximize the probability of selecting the best outcome. The “best” outcome can be determined in terms of utility maximization or a reference point, such as the Nash equilibrium.

(2) They do select the best choice in some subset of games likely to be faced by the learner; for example, if the full iterated deletion of strictly dominated strategies results in a final strategy, then it will be a Nash equilibrium. Therefore a selection method (LMA) based on dominance of this type will select Nash equilibria in a subset of games. In others it will fail to select a Nash equilibrium. The crucial thing is that it will select the best choice in a subset of cases.³⁷

In terms of normal form games the globally error-minimizing algorithm would pick out a Nash equilibrium strategy and insist on playing such a strategy.

In fact, according to the simulations in this chapter, the trained network is able to find pure Nash equilibria of games never encountered before approximately 60% of the time, an empirically sound success rate (for supporting experimental evidence see Stahl and Wilson (1994)). We show that, however, criteria based on iterated deletion of strictly dominated strategies (including rationalizability) outperform Nash whenever they yield a unique solution. We find and discuss the game features that, as part of its LMA, the network has learnt to recognize to address the Nash solving problem in a satisficing way. The network displays some strategic awareness, but this decreases in the level of iterated deletion required. The network goes for high payoff values. It takes into account the temptation of the other player of deviating from Nash. It plays better in higher stakes

³⁷A “rule of thumb” can be much more general than this. It is often imposed exogenously from outside the model or learning process, and may never select the optimal action.

games, particularly if there is more conflict of interests between itself and the other player.

We also find that the trained network's behavioural heuristics allow it to play in a meaningful way not just on new games, but on new classes of games, namely games with multiple and zero pure Nash equilibria. Moreover, networks trained on different sets of games (all with a unique pure Nash equilibrium) display focal points when encountering games with multiple equilibria.

8.1.1 Overview

The standard way to test neural networks in the cognitive and computer sciences is through the development of a working network. This chapter reports on the construction and testing of such a network. Section 8.2 details testing designed to check the robustness of the network to different parameter values, in particular the convergence level, and finds that as argued in the last chapter the network will achieve a success rate in picking Nash equilibria in never before seen games at well below 100%, but also well above that possible through random guessing (33%). In fact it achieves a 60% success record which nicely equates with the (human) experimental findings of Stahl and Wilson (1994).

In sections 8.3 and 8.4 the focus is on trying to find the resting place for the neural network's learning dynamic. We want to find out, first, whether the network's behaviour can be better described by algorithms other than Nash. A variety of candidate rules of thumb are examined as possible answers to the bounded-rational network's attempt to do its best given the limits to its computational power. The answer to this question yields another avenue of enquiry, namely which game features the network has learnt to pick when categorizing games and choosing the appropriate response. The framework for all of this was made clear in the last chapter: 3×3 normal form games with a unique pure Nash equilibrium.

The training stage will remain as in chapter 7; once more the network will face games with unique Nash equilibria, but in sections 8.5 and 8.6, the network will be made to face 3×3 normal form games with multiple pure Nash equilibria, and then no pure

Nash equilibria. The analysis will provide a rationale for focal points in games with multiple equilibria and also show how, in both sets of games, the acquired knowledge of the network yields meaningful behavioural predictions, even though the network has never faced a game with multiple or zero Nash equilibria in the training stage.

Section 8.7 concludes, finding that neural networks may be a promising tool to study bounded-rational behaviour in normal form games, with rules of thumb emerging endogenously as a result of the learning process rather than being exogenously super-imposed on the agent.

8.1.2 Learning to Play Nash

We start by using the same terminology as in chapter 7, so we are dealing with a neural network player $C = \langle \Omega, X, Y, F \rangle$ as defined in definition 12, with utility function $u(Y)$ facing a sequence of 3×3 normal form games, each given by $G = \langle N, \{A_i, u_i\}_{i \in N} \rangle$, having developed a method of play through a learning algorithm L attempting to minimize the gap between the “correct” choice, Nash equilibrium strategy α , and the actual choice, a . This learning algorithm L , and the “gap” between perfect and actual learning, $\varepsilon_0(M, \delta)$, is given in the last chapter in definition 18. Recall that Z^M defines the space of possible games as perceived by the neural network. Furthermore we will be heavily reliant on the term “local error-minimizing algorithm” introduced in definition 24 from the last chapter.³⁸

We will call the simulated network C^* once trained to a given convergence level. The training set consisted of $M = 2000$ games with unique pure Nash equilibria (henceforth PNE). Training was random with replacement, and continued until the error ε converged below 0.1, 0.05 and 0.02, i.e. three *convergence levels* γ were used: more than one convergence level was used for the sake of performance comparison. Convergence was checked every 100 games, a number large enough to minimize the chance of too early an end to training: clearly, even an untrained or poorly trained network will get an

³⁸The appendix to chapter 7 details an algorithm which is globally error-minimizing in finding Nash equilibria in the class of games examined here.

occasional game right, purely by chance. The computer determined initial connection weights and order of presentation of the games according to some ‘random seed’ given at the start of the training. To check the robustness of the analysis, C was trained 360 times, that is once for every combination of 3 learning rates η (0.1, 0.3, 0.5), 4 momentum rates μ (0, 0.3, 0.6 and 0.9) and 30 (randomly generated) random seeds. Convergence was always obtained, at least at the 0.1 level, except for a very high momentum rate.³⁹

The trained C (from now on C^*) was tested on a set of 2000 games with unique Nash equilibria never encountered before. We considered an output value to be correct when it was within some range from the exact correct value. If both outputs were within the admissible range, then the answer could be considered correct. The ranges considered were 0.05, 0.25 and 0.5. An answer satisfying the tight 0.05 error tolerance criterion can be considered exactly correct. An answer satisfying the 0.5 criterion is correct only in the sense that, if we interpreted choice as an alternative between three strategies only rather than in probabilistic terms, the network would gravitate around the correct one.

Figure 29 overleaf displays the average performance of C^* classified by γ , η and μ . It shows that C^* trained until $\gamma = 0.1$ played exactly (i.e., within the 0.05 range) the Nash equilibria of 60.03% of the 2000 3×3 games never encountered before. This fits well with the 59.6% average success rate of human subjects newly facing 3×3 games in the experiment detailed in Stahl and Wilson (1994). With an error tolerance of 0.25 and 0.5, the correct answers increased to 73.47 and 80%, respectively.

³⁹A combination of a very low learning rate (0.1) and momentum (0 or 0.3) slowed learning considerably; still, 0.1 convergence was always achieved. A very high learning rate and momentum were also detrimental; since such a high momentum seems very hard to justify psychologically, this can be reasonably disregarded. Otherwise, C converged robustly across different combinations. A detailed table can be found in Zizzo (2000b).

FIGURE 29: Percentage of Games Never Encountered Before for which the Network Chose a Partially or an Entirely Correct Strategy

Convergence Level	At Least 1 Correct Output Error tolerance criterion			Correct Answer Given Error tolerance criterion		
	0.05	0.25	0.5	0.05	0.25	0.5
0.1	85.12	91.76	94.31	60.03	73.47	80
0.05	87.26	92.24	94.31	64.12	74.75	80.09
0.02	88.52	92.51	94.25	66.66	75.47	79.96

Learning Level	At Least 1 Correct Output Error tolerance criterion			Correct Answer Given Error tolerance criterion		
	0.05	0.25	0.5	0.05	0.25	0.5
0.1	86.88	92.3	94.42	63	74.48	80.22
0.3	86.75	92.06	94.25	63.3	74.42	79.89
0.5	86.81	92.04	94.2	63.66	74.54	79.94

Momentum Level	At Least 1 Correct Output Error tolerance criterion			Correct Answer Given Error tolerance criterion		
	0.05	0.25	0.5	0.05	0.25	0.5
0	86.87	92.36	94.5	62.86	74.63	80.47
0.3	86.77	92.27	94.42	62.89	74.49	80.22
0.6	86.91	92.09	94.23	63.73	74.53	79.9
0.9	86.6	91.52	93.8	64.05	74.05	79.04

Notes: The level of convergence simply measures how correct we ask C^* to be: the smaller it is, the stricter the criterion. The learning rate is a coefficient that determines the speed of the adjustment of the connection weights when the network fails to play the Nash equilibrium behaviour. A positive momentum rate introduces autocorrelation in the adjustments of the connection weights when successive examples are presented. The error tolerance criterion measures how much error is tolerated, i.e. how close the answer given by the network must be to the correct answer in order to consider the answer right. The smaller the error tolerance criterion, the tighter it is. The numbers given under the heading "At least 1 Correct Output" are the percentages of cases in which at least 1 of the two output nodes is correct. The numbers given under the heading "Correct Answer Given" are the percentages of case in which both output nodes are correct, and so a correct answer is given.

Further training of C^* improves its performance on exactness - the 0.02-converged C^* plays exactly the Nash equilibria of a mean 66.66% of the games - but *not* on “rough correctness” (the 20% result appears robust). This suggests (and indeed further training of the network suggests) that there is an upper bound on the performance of the network.

Figure 29 also shows that, once C^* converges, the degree to which it makes optimal choices is not affected by the combination of parameters used: the average variability in performance across different learning rates is always less than 1%, and less than 2% across different momentum rates. This is an important sign of robustness of the analysis.

We compared C^* 's performance with three null hypotheses of zero rationality. Null1 is the performance of the entirely untrained C : it checks whether any substantial bias towards finding the right solution was hardwired in the network. Null2 alternates among the three pure strategies: if C 's performance is comparable to Null2, it means all it has learnt is to be decisive on its choice among the three. Null3 entails a uniformly distributed random choice between 0 and 1 for each output: as such, it is a proxy for zero rationality. Figure 30 overleaf compares the average performance of C^* with that of the three nulls. Formal t tests for the equality of means between the values of C^* and of each of the nulls (including Null2) are always significant ($P < 0.0005$). C^* 's partial learning success is underscored by the fact, apparent from figures 29 and 30, that when C^* correctly activates an output node it is very likely to categorize the other one correctly, while this is not the case for the nulls.

So it appears that C^* has learnt to generalize from the examples and to play Nash strategies at a success rate that is significantly above chance. Since it is also significantly below 100%, the next question we must address is how to characterize the LMA achieved by the trained network.

FIGURE 30: Average Performance of the Trained Network versus Three Null Hypotheses

	At Least 1 Correct Output Error tolerance criterion			Correct Answer Given Error tolerance criterion		
	0.05	0.25	0.5	0.05	0.25	0.5
Trained C	86.82	92.13	94.29	63.31	74.48	80.02
Null1 (Untrained C)	0	0.005	67	0	0	0
Null2 (Strategy Switcher)	78.85	78.85	78.85	22.8	22.8	22.8
Null3 (Random)	0.091	43.5	75.1	0.003	0.06	25.5

Notes: The smaller the error tolerance criterion, the tighter the criterion used to consider the network's strategy choice correct. The numbers given under the heading "At least 1 Correct Output" are the percentages of cases in which at least 1 of the two output nodes is correct. The numbers given under the heading "Correct Answer Given" are the percentages of cases in which both output nodes are correct, and so a correct answer is given. The strategy switcher alternates between pure strategies.

8.2 ALTERNATIVES TO NASH

Our first strategy in trying to characterize the LMA employed by the trained network is to ask ourselves whether there are simple alternatives to Nash capable of describing what the network does better than Nash, on the games over which they are uniquely defined. Given the robustness of our analysis in the previous section to different combinations of η and μ , in this and the next sections we just focus on the case with $\eta = 0.5$ and $\mu = 0$.⁴⁰ Hence, for testing we used the 30 networks trained with the 30 different random seeds but with the same learning (0.5) and momentum (0) rates. Using these 30 networks, we tested the average performance of the various algorithms on the same testing set of 2000 new games with unique PNE considered in the previous section.

In all cases, this note will consider such algorithms, by testing whether algorithms other than Nash are able to predict the network's choices better than Nash does, even though they have not been taught to the network. We will consider the following algorithms in turn:

- (1) Minmax
- (2) Rationalizability
- (3) "0-level strict dominance" (OSD)
- (4) "1-level strict dominance" (ISD)
- (5) "pure sum of payoff dominance" (PSPD)
- (6) "maximum payoff dominance" (MPD)
- (7) "nearest neighbour" (NNG)

8.2.1 Minmax

Minmax is often considered a form of reservation utility, and can be defined as:

Definition 25. *Consider the game G as defined in definition 10, and the neural network*

⁴⁰The robustness of the results with different parameter combinations ensures that *this particular choice* is not really relevant. In any event, it was driven by two considerations: 1) any momentum greater than 0 has hardly any real psychological justification, at least in this context; 2) given $\mu = 0$, a learning rate of 0.5 had systematically produced the quickest convergence.

player C as defined in definition 12. Index the neural network by i and the other player by j . The neural network's minmax value (or reservation utility) is defined as:

$$r_i = \min_{a_j} \left[\max_{a_i} u_i(a_i, a_j) \right]$$

The payoff r_i is literally the lowest payoff player j can hold the network to by any choice of $a \in A$, provided that the network correctly foresees a_j and plays a best reply to it. It therefore requires a particular brand of pessimism to have been developed during the network's training on Nash equilibria. An algorithm which looks for the minmax payoff is of course a local error-minimizing algorithm in the sense of definition 24 when the subclass of games faced is zero-sum, or by a minor reworking of utility, constant sum. So we might consider this a candidate if the learning dynamic, designed to pick up Nash equilibria, has faced several purely competitive games of the constant sum variety. This is of course possible, though unlikely when payoffs are independently drawn from a uniform distribution with support $[0, 1]$.

8.2.2 Rationalizability and Related Concepts

Rationalizability is widely considered a weaker solution concept compared to Nash equilibrium, in the sense that every Nash equilibrium is rationalizable, though every rationalizable equilibrium need not be a Nash equilibrium. Rationalizable sets and the set which survives the iterated deletion of strictly dominated strategies are equivalent in two player games: call this set S_i^n for player i after n stages of deletion. To give a simple intuitive definition, S_i^n is the set of player i 's strategies that are not strictly dominated when players $j \neq i$ are constrained to play strategies in S_j^{n-1} . The network will delete strictly dominated strategies and will assume other players will do the same, and this may reduce the available set of strategies to be less than the total set of actions for i , resulting in a subset $S_i^n \subseteq A_i$. Since we are dealing with only three possible strategies in

our game G , the subset can be adequately described as $S_i^2 \subseteq A_i$ with player j restricted to $S_j^1 \subseteq A_j$.

The algorithm OSD checks whether all payoffs for the neural network (the row player) from playing an action are strictly higher than those of the other players, so no restriction is applied to the action of player $j \neq i$, and player i 's actions are chosen from $S_i^0 \subseteq A_i$. 1SD allows a single level of iteration in the deletion of strictly dominated strategies: the row player thinks that the column player follows OSD, so chooses from $S_j^0 \subseteq A_j$, and player i 's action set is restricted to $S_i^1 \subseteq A_i$. Both of these algorithms can be viewed as weakened versions of iterated deletion, which are certainly easier to calculate. In this terminology 2SD would be full rationalizability or the full iterated deletion of strictly dominated strategies as defined above.⁴¹

8.2.3 Payoff Dominance

PSPD and MPD are different ways of formalizing the idea that the agent might try to go for the largest payoffs, independent of strategic considerations.

In the case of PSPD, an action $a_i = a_{PSPD}$ is chosen by the row player according to:

$$a_{PSPD} = \arg \max_{a_i \in A_i} \{a_i \mid a_j = \Delta_j\}$$

Where Δ_j is defined as a perfect mix over all available strategies in A_j . Put simply, a_{PSPD} is the strategy which picks a row by calculating the payoff from each row, based on the assumption that player j will randomly select each column with probability $\frac{1}{3}$, and then chooses the row with the highest payoff calculated in this way.

MPD is even more lowly in its required level of rationality, and the neural network, if following this algorithm, learns a set of rules designed to spot the highest conceivable payoff for itself. The network then picks the row containing the highest payoff, hoping

⁴¹OSD and 1SD are related to, but narrower concepts than player types based on levels of reasoning detailed in Stahl and Wilson (1994) and Stahl and Wilson (1995).

the other player will pick the corresponding column.

Both PSPD and MPD, by making very arbitrary assumptions about the other player, are strategically unsophisticated as candidate LMAs. Nevertheless, they represented often cited rules of thumb, and it will be interesting to see whether they describe any part of the network's play.

8.2.4 Nearest Neighbour

The NNG is an algorithm that draws its roots from cognitive science. As we discussed in chapter 7, neural networks are considered models of the categorization process, of how decision-makers classify input in order to produce the correct output. We also mentioned that one possible way in which networks may do this, when the sample is small enough, is by behaving like an exemplar-based model of categorization: that is, they may assimilate the instance to the nearest instance (example) encountered in the past, and solve the decision problem accordingly.

The best way to formalize this approach is not in terms of expected utility theory, but rather in terms of an alternative paradigm which emphasizes the role of past experiences in decision-making. Gilboa and Schmeidler (1995) provide such a paradigm in the form of *case based decision theory*. Consider G^P and A to be finite, nonempty sets, of games and strategies (or actions) respectively. X is a set of realized outcomes. x_0 is included within X as the result “this strategy was not chosen”. The set of *cases* is $C \equiv G^P \times A \times X$, which lists all conceivable combinations of games, strategies and realized outcomes. To interpret, when a player considers a current game, that player does so in terms of the game itself, possible strategies, and realized outcomes which spring from particular choices of strategy. Importantly, the player is not asked to consider hypothetical outcomes from untried strategies. Rather, the player considers only past experiences of realized outcomes. To make this concrete, given a subset of cases $C_s \subseteq C$, denote its projection P by H . So,

$$H = H(C_s) = \{q \in G^P \mid \exists a \in A, x \in X, \text{ such that } (q, a, x) \in C_s\}$$

Where H denotes the history of games, and $C_s \subseteq C$ denotes the subset of cases recorded as memory by the player, such that (i) for every $q \in H(C_s)$ and $a \in A$, \exists a unique $x = x_{C_s}(q, a)$ such that $(q, a, x) \in C_s$, and (ii) for every $q \in H(C_s)$, \exists a unique $a \in A$ for which $x_{C_s}(q, a) \neq x_0$. Memory is therefore a function which assigns realized outcomes to pairs of the form (game, strategy). For every memory C_s , and every $q \in H = H(C_s)$, there is one strategy that was actually chosen at q with an outcome $x \neq x_0$, and all other potential strategies are assigned the outcome x_0 . So, our agent has a memory of various games, and potential strategies, where one particular strategy was chosen. This produced an outcome $x \neq x_0$, with other strategies having never been tried, so given the generic outcome x_0 .

Our agent, when faced with a problem, will examine his memory, C_s , for some similar problems encountered in history, H , and assign these past problems a value according to a similarity function $s(p, q)$. These past problems each have a remembered action with a given result, which can be aggregated according to the summation $\sum_{(q,a,x) \in C_s} s(p, q) u(x)$ where $u(x)$ evaluates the utility arising from the realized outcome x . Decision-making is simply a process of examining past cases, assigning similarity values, summing, and then computing the strategy a to maximize $U(a) = U_{p,C_s}(a) = \sum_{(q,a,x) \in C_s} s(p, q) u(x)$.

Under this framework, the NNG algorithm examines each new game from G^P , and attempts to find the specific game $p \subseteq G^P$ from the training set (which proxies for memory) with the highest valued similarity function. In practice, in this chapter, similarity is computed by summing the square differences between each payoff value of the new game and each corresponding payoff value of each game of the training set. This sum of squares is a measure of dissimilarity between the new game and each game in the training set; if, in the limit, there were two identical games (which is never the case), the value would be exactly equal to 0. The game with the lowest dissimilarity index is defined as the

nearest neighbour. The NNG algorithm looks for the game with the lowest dissimilarity index and chooses the unique pure NE corresponding to the nearest neighbour. This is somewhat different from the case based decision theory optimization which involves a summation over all similar problems, but serves as a first approximation, and could be further generalized to incorporate a more complex similarity function.

In case based decision theory, nearest neighbour considerations would be incorporated into maximizing behaviour. In fact the NNG algorithm correctly specifies the optimal action, when there exists only one similar game for each new problem, so the function $s(p, q)$ becomes an indicator function, selecting the unique nearest neighbour. Nevertheless, in terms of game theory and Nash equilibrium, the NNG algorithm is closer to a rule of thumb.

8.2.5 Existence

Since, by construction, all games in the training set have a unique PNE, we are virtually guaranteed to find a NNG solution for all games in the testing set.⁴² A unique solution, or indeed any solution, may not exist with other algorithms, such as rationalizability, OSD and 1SD. A unique solution may occasionally not exist with other algorithms, such as MPD, because of their reliance on strict relationships between payoff values.

Figure 31 lists the number and percentage of games (out of 2000) for which a unique solution does not exist according to each algorithm, averaged out across the 30 neural networks trained with different random seeds, under a learning rate of 0.5 and momentum of 0.

⁴²The only potential (but unlikely) exception is if the nearest neighbour is not unique, because of two (or more) games having exactly the same dissimilarity index. The exception never arose with our game samples.

Non Nash Algorithm	Non Answerable Games	% of Total
0 Level Strict Dominance	1634	81.7
1 Level Strict Dominance	1205	60.25
Rationalizability	813	40.65
Pure Sum of Payoff Dominance	0	0
Maximum Payoff Dominance	8	0.4
Minmax	5	0.25
Nearest Neighbour	0	0

FIGURE 31: Existence with a Unique Solution

8.2.6 Algorithm Performance

Figure 32 overleaf describes how well the various algorithms fare on the testing set. Algorithms are classified into two groups, according to their performance. MPD, PSPD, minmax and, most interestingly, NNG fare worse than Nash on the data. We should not be surprised by the fact that the NNG still gets about half of the games right according to the 0.02 convergence level criterion: it is quite likely that similar games will often have the same Nash equilibrium. The failure of the NNG algorithm relative to Nash suggests that, at least with a training set as large as the one used in the simulations ($n = 2000$), the network does not reason by simply working on the basis of past examples. One must recognize, though, two limitations to this result. Firstly, partial nearest neighbour effects cannot be excluded in principle on the basis of figure 32. Secondly, perhaps C^* does not simply rely on the nearest neighbouring game; perhaps it relies on a weighted average of most similar games, although placing the most weight on the nearest game: this more sophisticated version of exemplar-based categorization would not be fully captured by the NNG.

Explanatory Variables	Model 1				Model 2				Model 3			
	Coef.	S.E.	Prob.	Sig.	Coef.	S.E.	Prob.	Sig.	Coef.	S.E.	Prob.	Sig.
MPD Existence	0.037	0.061	0.549									
Minmax Existence	0.067	0.077	0.382									
Same As PSPD	-0.095	0.011	0	****	-0.098	0.01	0	****	-0.098	0.01	0	****
Same As MPD	-0.037	0.01	0	****	-0.034	0.01	0	****	-0.034	0.01	0	****
Same As Minmax	-0.005	0.011	0.615									
Same As Nearest Neighbour	-0.085	0.008	0	****	-0.085	0.008	0	****	-0.086	0.008	0	****
Own NE Payoff	-0.339	0.063	0	****	-0.391	0.029	0	****	-0.392	0.029	0	****
Other's NE Payoff	-0.237	0.049	0	****	-0.202	0.034	0	****	-0.201	0.034	0	****
Positive Payoff Ratio	-0.025	0.013	0.057	*	-0.018	0.011	0.101		-0.017	0.011	0.108	
Negative Payoff Ratio	0.058	0.062	0.35									
Own Generic Temptation	0.104	0.048	0.031	**	0.105	0.048	0.03	**	0.108	0.048	0.025	**
Other's Generic Temptation	0.317	0.055	0	****	0.317	0.054	0	****	0.32	0.054	0	****
Game Harmony Index	-0.311	0.66	0.638									
Mean	-0.326	0.314	0.3		-0.237	0.208	0.255		-0.347	0.135	0.01	**
Game Harmony * Mean	-0.667	0.779	0.392		-0.857	0.507	0.091	*	-0.551	0.131	0	****
Deviation from Avg. Mean	0.192	0.147	0.19		0.174	0.13	0.181		0.213	0.121	0.077	*
Standard Deviation	0.587	0.602	0.33		0.883	0.349	0.012	**	1.13	0.201	0	****
Deviation from Avg. SD	-0.147	0.274	0.591		-0.027	0.018	0.125		-0.026	0.018	0.14	
Skewness	-0.03	0.018	0.096	*	-0.032	0.014	0.023	**	-0.025	0.013	0.05	*
Kurtosis	-0.033	0.014	0.02	**	0.056	0.026	0.032	**	0.055	0.026	0.035	**
Deviation from Avg Skewness	0.055	0.026	0.034	**	0.016	0.014	0.264					
Deviation from Avg Kurtosis	0.019	0.015	0.224									
NE Action 1	0.024	0.01	0.012	**	0.025	0.01	0.01	**	0.025	0.01	0.011	**
NE Action 2	0.037	0.01	0	****	0.038	0.01	0	****	0.038	0.01	0	****
Strict Dominance: Level 0 Sufficient	-0.079	0.014	0	****	-0.082	0.013	0	****	-0.081	0.013	0	****
Strict Dominance: Need for Level 1	-0.058	0.011	0	****	-0.059	0.011	0	****	-0.059	0.011	0	****
Strict Dominance: Need for Level 2	-0.049	0.011	0	****	-0.05	0.011	0	****	-0.05	0.011	0	****
Max Own Temptation	0.288	0.026	0	****	0.289	0.026	0	****	0.288	0.026	0	****
Min Own Temptation	0.038	0.022	0.084	*	0.039	0.02	0.055	*	0.037	0.02	0.063	*
Max Other's Temptation	0.142	0.025	0	****	0.143	0.025	0	****	0.145	0.025	0	****
Min Other's Temptation	0.002	0.024	0.946									
Game Harmony * S.D.	1.266	1.565	0.418		0.528	0.817	0.518					
Constant	0.404	0.286	0.158		0.386	0.085	0	****	0.356	0.082	0	****

Log-Likelihood (Model 1): 662.536. Log-Likelihood (Model 2): 661.275. Log-Likelihood (Model 3): 660.263. LR Test (Model 1→Model 2): $\chi^2(7)=2.52$, $P=0.925$; LR Test (Model 1→Model 3): $\chi^2(9)=4.55$, $P=0.872$; LR Test (Model 2→Model 3): $\chi^2(2)=2.02$, $P=0.364$. Numbers are approximated to the third decimal value. ****, ***, ** and * stand for significance at the 0.001, 0.01, 0.05 and 0.1 levels, respectively.

FIGURE 32: Tobit Regression on the Average RMS Error (n=2000)

Rationalizability, OSD and 1SD outperform Nash for the games they can solve in a unique way. OSD, 1SD and rationalizability predict the neural network's behaviour in 80.98%, 76.25% and 74.36% of their answerable games, respectively: this is 15-20% above Nash, and even more striking considering that these algorithms were not taught to the network! The network appears to have endogenously chosen a local error-minimizing algorithm (or rule of thumb) based on iterated deletion of strictly dominated strategies in order to better perform in the Nash solving decision problem. It does appear as if C^* were able to do some strategic thinking (it still gets about three quarters of the all rationalizable games exactly right), but the more the level of iterations (one relative to zero, full rationalizability relative to one) the more difficult it is for the neural network to do this properly.

C^* 's behaviour is best *described* by algorithms based on iterated deletion of strictly dominated strategies. However, it is clearly not the case that these alternatives to Nash fully describe the network's decision-making process. This is true not only because the network still fails to apply these methods perfectly, but also and more importantly because the network can still play reasonably well in games without a unique, or any, rationalizable strategy: this explains why, if one considers the overall set of 2000 games rather than just the answerable games, Nash is still the single best predictor.⁴³ We might conjecture that the successful alternatives to Nash are successful because they characterize game features that are also considered by the LMA achieved by C^* , but, also, that the LMA achieved by C^* does not simply coincide with these successful alternatives. We now need to make this conjecture more precise.

⁴³An objection to this result is that the right half of figure 31 is biased against the comparative success of the algorithms vs. Nash: it does not consider the fact that C^* might be playing randomly on the non-answerable games and, in so doing, it would still get an average 1/3 of these games correct by chance. However, even if we augment rationalizability, OSD and 1SD with random play on their non-answerable games, and so we credit 1/3 of their non-answerable games as correct, it is still the case that Nash outperforms. Augmented rationalizability gets closest, with 57.68% of the 2000 games correct (with $\gamma = 0.02$): this is still 9 points below Nash.

8.3 GAME FEATURES

A second strategy that we may use to gather information about the network's LMA is to analyse the game features that it has learnt to detect. If the LMA exploits certain game features to perform well on the game-solving task, then C^* will perform better on games that have those game features to a high degree. Learning about these game features will therefore give valuable information on what C^* has actually learnt.

In a prototype-based view of categorization, we could say that games in which those features are present to a high degree would be prototypical games, i.e. games that the network i would be able to classify (perfectly or almost perfectly) as best examples of games associated with playing an action from A_i .

In an exemplar-based view of categorization, if the network is sensitive to the specific examples encountered in the training set, and if the Nash equilibrium were also the nearest neighbour, we would expect the network to perform significantly better than otherwise. This would be true regardless of whether weight is given only to the nearest neighbour game in the exemplar-based algorithm, as we would expect the most weight to be given to the nearest neighbour (neighbour 1) relative to neighbour 2, 3, ..., n .

We ran Tobit regressions on the average root mean square error of the thirty neural networks trained with a learning rate of 0.5 and momentum rate of 0, and achieving a convergence level of 0.02. Tobit regressions were used because the distribution is truncated at 0, the lowest possible root mean square; 30 observations presented 0 values, implying a perfect performance by the neural network whatever the random seed. The game features that were used are listed in figure 33 together with the results; they can be classified into three groups:

(1) *Algorithm related features*. MPD and Minmax Existence look at whether a unique MPD or minmax solution exists for the game; the "Same As" variables look at whether the algorithm (for example, PSPD) has the same prediction as the NE algorithm for the game. Existence variables are not defined for algorithms that always exist (for example, NNG). In the case of the strict dominance algorithms, we chose instead to use three

dummy variables for the cases in which zero and exactly zero, one and exactly one, two and exactly two iteration levels are required to achieve a unique solution: these dummies are represented by “Strict Dominance: Level 0 Sufficient”, “Strict Dominance: Need for Level 1” and “Strict Dominance: Need for Level 2”, respectively. NE Action 1 and 2 are simply dummies equal to 1 when the NE action is actually 1 (Top) or 2 (Middle), respectively.

(2) *Payoff and Temptation variables.* These variables relate to the size of the NE payoff for the network and the other player, and to the size of the deviation from this equilibrium. Own Generic Temptation is a crude average of the payoff from deviating from Nash, assuming that the other player plays randomly. Max and Min Own Temptation are the maximum and minimum payoff, respectively, from deviating from Nash, taking the behaviour of the other player as given. Clearly, while the Generic Temptation variable assumes no strategic understanding, the Max and Min Own Temptation variables do assume some understanding of where the equilibrium of the game lies. Ratio variables reflect the ratio between the network’s own and the other’s payoff in the NE, minus 1: if the result is positive, then the Positive NE Ratio takes on this value, while the Negative NE Ratio takes a value of 0; if the ratio is negative, then the Positive NE Ratio takes a value of 0, while the Negative NE Ratio takes on the negative value, in absolute terms.

(3) *General game features.* These variables are mostly based on the moments of the game payoff distribution. We consider the mean, variance (standard deviation), skewness and kurtosis of the game payoffs for each game. We also consider the difference between the values observed for each game and the average value across the 2000 games. The *Game Harmony Index* is detailed in the appendix. It is a measure of how harmonious or disharmonious the players’ interests are in the game: it is equal to 0 if the game is perfectly harmonious, such as in the case of a pure co-ordination game; it will be equal to 1 if a zero-sum game, where the gain of a player is the loss of another; while it will take intermediate values for games with partial conflict of interests (for example, a Prisoner’s

Dilemma). The index is derived from the Gini coefficient of income distribution and is bounded between 0 and 1: a higher index entails a more disharmonious game.

Figure 33 on the next page presents the results of three Tobit regression models. Likelihood-ratio tests accept the reduction to the simplest model, Model 3 and the results yield a wealth of information on what the network is actually doing.

(1) *Go for high numbers, especially if they are yours.* The network gives better answers when the NE is associated with a high payoff - particularly for itself. The Same As MPD, the Same As PSPD and the Strict Dominance variables all work in the same direction. The coefficients on these variables are relatively small: this suggests that, although the network's behaviour can be best described by the strict dominance algorithms relative to the others in the context of games with a unique pure Nash equilibria, the network may actually be picking game features associated with strict dominance, rather than simply following strict dominance as a rule of thumb.

(2) *Feel and fear trembling hands.* The greater the temptation, the greater the chance of deviating from the right answer. The fear of the other player's temptation may perhaps be related to the Harsanyi and Selten (1988) concept of *risk dominance*, though risk dominance is only applicable to games with multiple equilibria. Again, more weight is given to one's own temptation than to the other player's, but the neural network does appear to place some strategic weight on the temptation of the other. More weight is also given to deviations from Nash playing, taking the action of the other player as given, another sign of some strategic reasoning in feeling or fearing the temptation of random trembles.

Algorithm	Correct Games, As a Percentage of Answerable Games			Correct Games, As a Percentage of All Games		
	Cvg Level=0.5	Cvg Level=0.25	Cvg Level=0.05	Cvg Level=0.5	Cvg Level=0.25	Cvg Level=0.05
Nash	80.38	75.70	66.53	80.38	75.70	66.53
<i>Overperformers over answerable games</i>						
0 Level Strict Dominance	90.83	87.79	80.98	16.62	16.06	14.82
1 Level Strict Dominance	87.63	84.00	76.25	34.83	33.39	30.31
Rationalisability	86.44	82.57	74.36	51.3	49.01	44.13
<i>Underperformers</i>						
Pure Sum of Payoff Dominance	67.71	63.48	55.97	67.71	63.48	55.97
Maximum Payoff Dominance	61.08	57.02	49.95	60.84	56.79	49.75
Minimax	57.75	53.83	46.89	57.60	53.70	46.77
Nearest Neighbour	62.97	58.78	51.48	62.97	58.78	51.48

Based on a network trained with learning rate=0.5 and momentum=0.

FIGURE 33: Tobit Regression on the Average RMS Error (n=2000)

(3) *The greater the strategic complexity, the greater the difficulty.* The coefficients on the network's own temptation to deviate from Nash, taking the action of the other player as given, is higher than the coefficient on the other's temptation taking one's own action as given; however, the coefficient on the other's generic temptation is higher than that on one's own generic temptation. The own Nash deviation temptation requires just a model of the other player as taking one's own action as given; whereas a consideration of the other player's Nash deviation temptation requires one more level of strategic understanding in the sense of Stahl and Wilson (1994), namely a model of the other player having a model of oneself as taking the other player's action as given. Faced with a more difficult task, the neural network puts more weight on the other player's generic temptation. There are two other signs that, the greater the strategic complexity, the greater the difficulty for the neural network. One is that the network placed more weight on its own than on the other's payoff; the other is that the coefficient on Strict Dominance: Level 0 Sufficient is significantly higher (in absolute terms) than the one on Need for Level 1, which is higher (again, in absolute terms), albeit insignificantly so, than the one on Need for Level 2: strict row dominance is simpler than iterated deletion of strictly dominated strategies.

(4) *High stakes provide motivation for the network.* The network finds it difficult to process payoff values distant from the mean payoff value (of 0.50154), but finds it more difficult to process games for low stakes (because of the negatively signed Game Harmony \times Mean term; the negative Skewness term is also indicative). In any laboratory experiment setting subjects have to be motivated enough by the stakes at play: similarly, C^* makes smaller mistakes when the stakes are higher, as if motivation were required.

(5) *Keeping game harmony constant, an increase in payoff variance induces less correct answers.* Insofar as it is not correlated to greater game disharmony, a greater standard deviation is likely to proxy an increase in the variance in the network's own, or the other player's, possible payoffs. The prediction is that the agent may be less keen on NE play than otherwise, if there is an increase in the variance in its own possible payoffs.

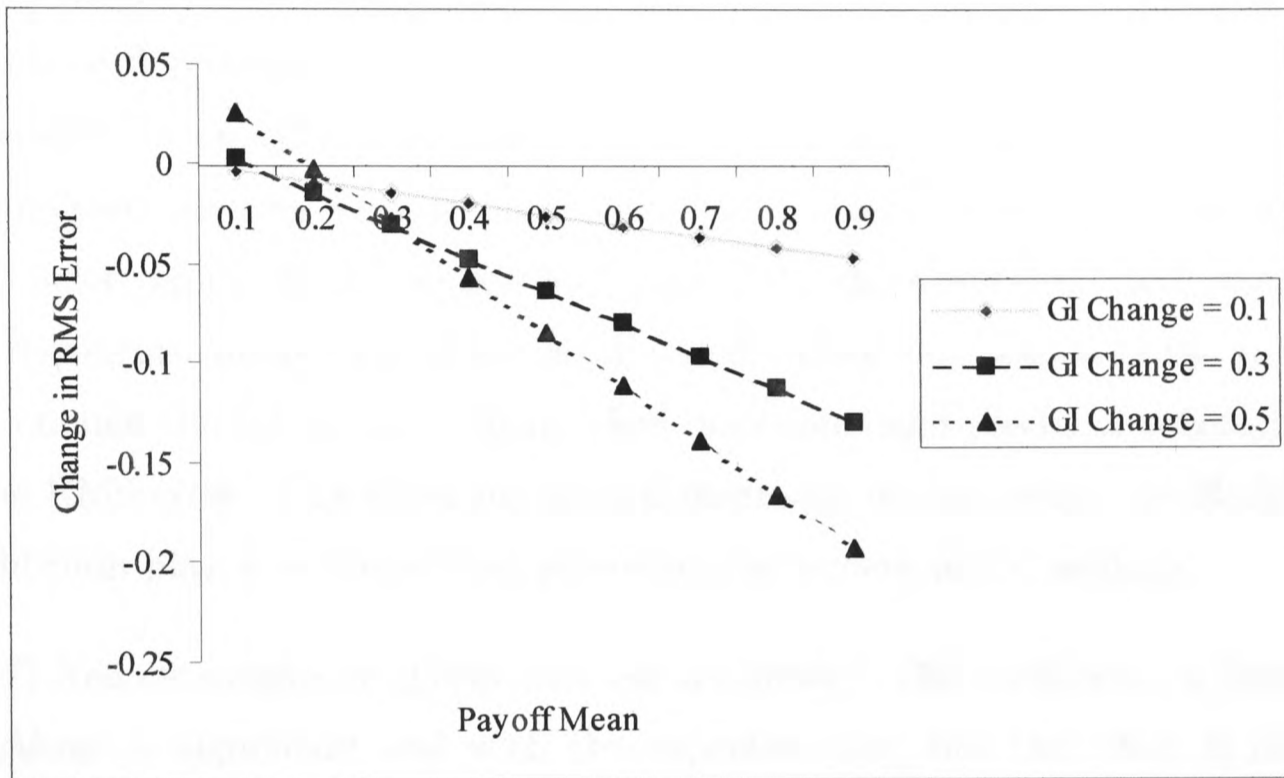


FIGURE 34: The Effect of a Change in Game Harmony on the RMS Error for Different Payoff Mean Levels

(6) *When the game is for high enough stakes, higher game disharmony induces more correct answers.* An increase in the *Game Harmony Index* (GI) implies greater game disharmony. In Model 3, this feeds into the RMS error through two channels: one, positively signed, is the Standard Deviation, because an increase in GI is also likely to increase standard deviation (Pearson $r=0.526$); the other, negatively signed, is the $\text{GI Index} \times \text{Mean}$ term. The two effects operate in opposite directions and, also, the second channel will be stronger the higher the mean. In order to analyse which effect is likely to dominate, standard deviation was regressed on GI by ordinary least squares; the coefficient on GI is 0.19 (S.E.=0.007), and used this as a proxy for $(\text{DSD})/(\text{DGI})$. We analysed how, for various mean levels, GI increases between 0.05 and 0.5 affects the RMS error; figure 34 on the previous page plots the results for $\text{DGI}=0.1, 0.3$ and 0.5 . The results are ambiguous for low mean payoff values, but as soon as the game has stakes high enough the $\text{GI Index} \times \text{Mean}$ effect dominates and produces substantial decreases in the RMS error. This effect on game disharmony on improving the likelihood of Nash equilibrium play is an interesting prediction for experimental settings.

(7) *Nearest-neighbour effects exist but are limited.* The coefficient on *Same As Nearest Neighbour* is significant and with the expected sign, but the effect is not large. On the basis of this and the previously considered evidence, we conclude that, while more sensitive to examples than a pure prototype model would be, C^* follows more a prototype-based approach to the categorization of the new games it encounters.

(8) *Other processing features.* Kurtosis has a small but significant positive effect on the RMS effect: similar payoff values may be harder to differentiate, and more subject to random trembles. The network also has a slightly higher root mean square error with actions 1 and 2, quite possibly because of its greater familiarity in producing zeros than ones as outputs.

In conclusion, the neural network appears to have found ways to get around the problem of attempting to find a Nash equilibrium strategy in never before seen games. They rely on a plausible mix of payoff dominance and a fear of trembles, together with

some strategic awareness (that decreases for more strategically complexity games). While not being taught to the network, they are game features that, being faced with too complex a problem, the network has extracted to categorize games, in a prototype-based, satisficing fashion: they correspond to emergent rules of thumb endogenously chosen by the bounded-rational agent and are in some sense to be expected given the role of local error-minimization in the learning process.⁴⁴

8.4 MULTIPLE EQUILIBRIA

In this section we still consider the network trained purely on games with unique PNE, but ask ourselves what its behaviour will be when faced not just with new games, but with *a new class* of games, namely games with multiple PNE.

If C^* has learned to categorize games according to some game features, we would expect the network to apply a similar set of tools as far as possible when faced with games with multiple equilibria. This, of course, may be impossible if C^* 's LMA is inapplicable to this context. For example, if C just followed iterated deletion of strictly dominated strategies, our best describing algorithm for the single PNE case, then the network should be unable to choose among the plurality of PNE, as these all correspond to rationalizable solutions. On the basis of the last section, however, we hypothesize that this will not be the case, even if the network has never faced games with multiple PNE in the training stage.

8.4.1 Focal Points

A second, stronger hypothesis is that the network will be displaying focal points: this means that *different* networks should tend to converge to the same pure Nash equilibria. Why this should be the case varies according to whether we view networks as working

⁴⁴We also ran Probit regressions on the best performing (prototype) games and the worst performing games. Results do not differ much qualitatively, and are omitted. Interestingly, a higher game disharmony appears particularly conducive in defining prototypical games, while it is insignificant in defining particularly badly performing games.

mainly as an exemplar or as a prototype-based model of categorization. If the former, different networks, trained under different random seeds but with the same training set (and learning rate and momentum), will tend to choose the action corresponding to the solution of the nearest neighbour to the game with multiple pure Nash equilibria: hence, there will be a focal solution. However, one might wonder whether the differences in random seeds are really so important, particularly given the fact that training is random but with replacement: perhaps, we are dealing with basically the same neural networks in each case, and so the finding of focal points may be considered uninteresting as a model of what might be focal in the real world. We shall talk in this case about focal points “in a weak sense”, or *w-focal points*.

If the neural network works mainly as a prototype-based model of categorization, we would expect focal points even if the network has been trained with different training sets, as long as they are drawn from the same distribution. This is because the latter condition is sufficient to ensure that the different neural networks will extract approximately the same game features. We shall talk in this case about focal points “in a strong sense”, or *s-focal points*.

We considered two sets of neural networks. The first set of thirty networks (*Set 1*) is the standard group considered in the previous sections, trained with the same training set, $\eta = 0.5$, $\mu = 0$, but with different random seeds. The second set of thirty networks (*Set 2*) was trained again with $\eta = 0.5$ and $\mu = 0$, but varying not only the random seed but also the training set in each case; in addition, training was random *without* replacement. Thirty training sets of $M = 2000$ games were used, with each payoff drawn from a uniform distribution with support $[0,1]$.

On the basis of the results from our previous sections, we hypothesize that the network works mainly as a prototype-based model of categorization, and that, therefore, it will display not only *w-focal points* but also *s-focal points*. Since it retains some (intuitively plausible) sensitivity to examples, however, we might expect the percentage of neural networks converging to *w-focal points* to be slightly higher than the one converging to *s-focal points*.

The testing set was made of 2000 games again, namely 100 games with three PNE and 1900 games with two PNE. Let us call a choice “decided” if both outputs are within 0.25 of a pure strategy value. Let us then consider the number of decided choices (between 0 and 30) corresponding to each action (1, 2, 3), for each game. We can formulate two null hypotheses for the absence of focal points in terms of the distribution of decided choices across actions. According to the first null hypothesis, the player would simply choose randomly which action to take, i.e. the player would be entirely naive in facing games with multiple PNE: in this case, we would expect the number of decided choices to be the same across actions, and we shall take them to be equal to the average number of decided choices. According to the second null hypothesis, the agent would be able to detect the pure Nash equilibria, but would only be able to choose among them randomly. On average, in this case we would expect the same number of decided choices for each pure Nash equilibrium. Clearly, this second null hypothesis can be meaningfully distinguished from the first only in the case of games with two, rather than three, PNE.

For the three pure Nash equilibria dataset ($n = 100$) under both nulls, $\chi^2 = 2531.256$, 198 d.f., for Set 1, and $\chi^2 = 1853.324$, 198 d.f., for Set 2. For the two pure Nash equilibria dataset ($n = 1899$) under the first null, $\chi^2 = 67653.74$, 3798 d.f., for Set 1, and $\chi^2 = 56174.93$, 3798 d.f.. For the two pure Nash equilibria dataset ($n = 1900$) under the second null, $\chi^2 = 30785.17$, 1898 d.f., for Set 1, and $\chi^2 = 23985.49$, 1899 d.f. for Set 2.⁴⁵

Hence, the network is displaying not only w-focal points but also s-focal points. Interestingly, the χ^2 is lower with Set 2 than with Set 1 in a comparable sample of two pure Nash equilibria or three pure Nash equilibria games, suggesting that the importance of focal points is somehow lower with s-focal points, as might be expected by the limited exemplar effect. Nevertheless, the strong evidence for s-focal points suggests once again that the network is mainly reasoning as a prototype-based model of categorization. Different neural networks, trained on the same game distribution although on different games, must be displaying focal points because they have learnt to detect the same game

⁴⁵One observation was excluded from Set 1 in this case because its expected value was equal to zero. In all cases, using χ^2 tests the null hypotheses are strongly rejected (at $p < 0.001$) for both Set 1 and 2.

features and so they tend to choose the same solution.

A criticism of this conclusion might be that, although we have shown that the number of decided choices tends to be focal on specific choices, we have not shown that the number of decided choices is high in the first place in games with multiple pure Nash equilibria. However, the number of decided choices only drops from an average of 8.97 per game action in the unique pure Nash equilibrium dataset to 8.52 with Set 1 and 8.49 with Set 2: taking into account that, if all choices were “decided”, the value should be equal to 10, it is apparent that the neural network is quite decided in general in its choices, and is only slightly more indecisive in games with multiple pure Nash equilibria.

8.4.2 Features of Focal Games

What are the game features that make a game focal? In order to explore this, let us define three data-points for each game, in correspondence to each action: one data-point corresponding to the number of decided choices from playing action 1, one from playing action 2 and one from playing action 3. This allows to obtain, in principle, a dataset of 6000 observations, in correspondence to the 2000 games: let us label these observations as NDecided1 if they are based on the number of decided choices with Set 1, and NDecided2 if they are based on the number of decided choices with Set 2. We can now regress NDecided1 and NDecided2 on a variety of game and action features, in order to determine what makes the network choose an action rather than another one.

Many of the features considered are identical or similar to the ones previously considered; a few are new ones:

(1) *Algorithm related features.* Same As OSD and 1SD refer to the case in which the action is dominated according to 0 or 1 iteration level strict dominance.⁴⁶ Same As NE When 2NE yields a value of one when the action is the same as a NE, but only for the games with two pure Nash equilibria - for the games with three pure Nash equilibria, each action is a pure Nash equilibrium, so it would not be a useful marker. Conversely,

⁴⁶More than one iteration is never needed in games with multiple PNE.

Presence of 3NE marks the games with three pure Nash equilibria. Same As Utilitarian Best is equal to one when the action corresponds to the best pure Nash equilibrium from a utilitarian perspective (i.e. players i and j choose the actions a_i^* and a_j^* such that $u_i(a_i^*, a_j^*) + u_j(a_i^*, a_j^*) \geq u_i(a_i, a_j) + u_j(a_i, a_j)$ for all $a_i \in A_i$ and for all $a_j \in A_j$).

(2) *Payoff and Temptation variables.* We need to modify these variables because we are considering the desirability of each action, not just of pure Nash equilibria actions. We reformulate the variables in terms of Best Response (BR): given an action of the neural network, what is the strictly best response of the other player? BR outcomes will be defined for all actions in which the top payoff for the column player in correspondence to the action being considered is strictly higher than the others. This is the case for all but three observations in our sample; in the regression analysis, we drop these three observations and restrict ourselves to a sample of $n = 5997$ observations. We also add two sets of new variables. First, we introduce temptation variables (Own/Other's BR Max/Min Temptation) on the basis of the (min/max, own/other's) BR payoffs from playing the other action. Second, we introduce two interaction terms: Game Harmony \times Own Temptation and Game Harmony \times Other's Temptation.

(3) *General game features.* These are similar to those previously considered.

Figures 35 and 36 overleaf present the result of ordered Probit regressions on NDecided1 and NDecided2; the restrictions entailed by the simpler models in both figures are, once again, accepted using likelihood-ratio tests. There are some differences between the two figures. In particular, game harmony appears a somehow better predictor of NDecided2 than NDecided1. However, in general the picture is quite similar.

Explanatory Variables	Model 1				Model 2				Model 3			
	Coef	SE	Prob	Sig	Coef	SE	Prob	Sig	Coef	SE	Prob	Sig
Same As PSPD	0.549	0.041	0	****	0.548	0.041	0	****	0.559	0.04	0	****
Same As MFD	0.054	0.039	0.169		0.055	0.039	0.159		0.064	0.038	0.095	*
Same As Minmax	0.066	0.038	0.078	*	0.067	0.038	0.076	*	0.073	0.037	0.052	*
Same As OSD	-0.046	0.103	0.652									
Same As ISD	0.151	0.103	0.145		0.117	0.072	0.104		0.115	0.071	0.108	
Same As Utilitarian Best	0.217	0.045	0	****	0.217	0.045	0	****	0.219	0.045	0	****
Positive Payoff Ratio	-0.042	0.029	0.151		-0.043	0.028	0.126					
Negative Payoff Ratio	0.23	0.192	0.231		0.228	0.191	0.233					
Own BR Max Temptation	-1.071	0.116	0	****	-1.078	0.113	0	****	-1.061	0.11	0	****
Other's BR Max Temptation	-1.044	0.166	0	****	-1.058	0.157	0	****	-1.099	0.147	0	****
Own BR Min Temptation	0.01	0.082	0.903		0.01	0.081	0.904					
Other's BR Min Temptation	-0.791	0.103	0	****	-0.797	0.102	0	****	-0.8	0.099	0	****
Own BR Payoff	2.072	0.21	0	****	2.063	0.205	0	****	1.842	0.096	0	****
Other's BR Payoff	0.41	0.157	0.009	***	0.4	0.153	0.009	***	0.565	0.099	0	****
Own Max Temptation	-1.499	0.298	0	****	-1.488	0.09	0	****	-1.465	0.09	0	****
Other's Max Temptation	-0.464	0.289	0.109		-0.403	0.085	0	****	-0.403	0.085	0	****
Own Min Temptation	-0.478	0.099	0	****	-0.475	0.097	0	****	-0.49	0.096	0	****
Other's Min Temptation	-0.705	0.089	0	****	-0.704	0.088	0	****	-0.732	0.088	0	****
Own Generic Temptation	-2.143	0.336	0	****	-2.158	0.332	0	****	-1.943	0.277	0	****
Other's Generic Temptation	-0.325	0.23	0.158		-0.328	0.23	0.153					
Presence of 3NE	0.315	0.083	0	****	0.316	0.082	0	****	0.309	0.082	0	****
Game Harmony Index	0.295	1.348	0.827									
Same As NE When 2NE	0.557	0.067	0	****	0.559	0.067	0	****	0.555	0.064	0	****
Game Harmony * Own BR Temptation	0.036	0.789	0.963									
Game Harmony * Other's BR Temptation	0.172	0.763	0.822									
Mean	6.227	1.258	0	****	5.997	0.669	0	****	5.822	0.481	0	****
Game Harmony * Mean	-0.722	3.156	0.819		0.03	0.439	0.946					
Deviation from Avg Mean	-0.032	0.511	0.95									
Standard Deviation	-0.216	0.889	0.808									
Deviation from Avg SD	-0.083	0.795	0.916									
Skewness	-0.076	0.065	0.247		-0.076	0.065	0.239					
Kurtosis	-0.041	0.054	0.449		-0.033	0.043	0.443					
Deviation from Avg Skewness	0.032	0.096	0.738		0.025	0.08	0.75					
Deviation from Avg Kurtosis	0.055	0.055	0.318		0.05	0.05	0.311					
NE Action 1	-0.087	0.037	0.018	**	-0.087	0.037	0.018	**	-0.087	0.037	0.019	**
NE Action 2	-0.229	0.035	0	****	-0.229	0.035	0	****	-0.23	0.035	0	****

Log-Likelihood (Model 1): -13880.315. Log-Likelihood (Model 2): -13880.504. Log-Likelihood (Model 3): -13883.739. LR Test (Model 1→Model 2): $\chi^2(7)=0.38$, $P=1$. LR Test (Model 1→Model 3): $\chi^2(16)=6.85$, $P=0.976$. LR Test (Model 2→Model 3): $\chi^2(9)=6.47$, $P=0.692$. White robust estimators of the variance were used to compute standard errors. Numbers are approximated to the third decimal value. ****, ***, **, * and * stand for significance at the 0.001, 0.01, 0.05 and 0.1 levels, respectively.

FIGURE 35: Ordered Probit Regressions on NDecided1, Games with Multiple Pure Strategy Nash Equilibria (n=5997)

Explanatory Variables	Model 1				Model 2				Model 3			
	Coef	SE	Prob	Sig.	Coef	SE	Prob	Sig.	Coef	SE	Prob	Sig.
Same As PSPD	0.664	0.041	0	****	0.668	0.04	0	****	0.664	0.039	0	****
Same As MPPD	0.094	0.038	0.014	**	0.099	0.038	0.009	***	0.096	0.037	0.01	**
Same As Maximin	0.069	0.036	0.06	*	0.078	0.036	0.029	**	0.078	0.036	0.029	**
Same As OSD	-0.074	0.105	0.484									
Same As 1SD	-0.02	0.104	0.847									
Same As Pareto	0.264	0.044	0	****	0.264	0.044	0	****	0.263	0.044	0	****
Positive Payoff Ratio	0.001	0.03	0.966									
Negative Payoff Ratio	0.03	0.191	0.876									
Own BR Max Temptation	-1.287	0.116	0	****	-1.257	0.112	0	****	-1.278	0.111	0	****
Other's BR Max Temptation	-1.136	0.158	0	****	-1.122	0.147	0	****	-1.147	0.14	0	****
Own BR Min Temptation	-0.092	0.079	0.247		-0.079	0.078	0.313					
Other's BR Min Temptation	-0.968	0.099	0	****	-0.97	0.096	0	****	-0.958	0.095	0	****
Own BR Payoff	2.221	0.207	0	****	2.22	0.096	0	****	2.216	0.096	0	****
Other's BR Payoff	0.713	0.158	0	****	0.733	0.096	0	****	0.725	0.095	0	****
Own Max Temptation	-1.762	0.286	0	****	-1.629	0.088	0	****	-1.616	0.088	0	****
Other's Max Temptation	-0.646	0.27	0.017	**	-0.718	0.083	0	****	-0.707	0.082	0	****
Own Min Temptation	-0.532	0.094	0	****	-0.545	0.092	0	****	-0.555	0.091	0	****
Other's Min Temptation	-0.647	0.089	0	****	-0.66	0.088	0	****	-0.668	0.087	0	****
Own Generic Temptation	-2.307	0.319	0	****	-2.267	0.282	0	****	-2.341	0.266	0	****
Other's Generic Temptation	-0.148	0.223	0.508									
Presence of 3NE	0.24	0.078	0.002	***	0.271	0.07	0	****	0.274	0.07	0	****
Game Harmony Index	0.954	1.293	0.461									
Same As NE When ZNE	0.458	0.063	0	****	0.493	0.053	0	****	0.514	0.051	0	****
Game Harmony * Own BR Temptation	0.333	0.767	0.664									
Game Harmony * Other's BR Temptation	-0.189	0.711	0.791									
Mean	6.966	1.209	0	****	6.045	0.545	0	****	6.477	0.459	0	****
Game Harmony * Mean	-1.972	3.057	0.519		0.324	0.393	0.409					
Deviation from Avg. Mean	-0.046	0.502	0.927									
Standard Deviation	0.224	0.853	0.792									
Deviation from Avg. SD	-0.289	0.782	0.712									
Skewness	-0.106	0.064	0.097	*	-0.094	0.062	0.129					
Kurtosis	-0.034	0.052	0.518									
Deviation from Avg Skewness	0.067	0.092	0.469		0.032	0.058	0.576					
Deviation from Avg Kurtosis	0.012	0.052	0.817									
NE Action 1	-0.084	0.036	0.021	**	-0.083	0.036	0.022	**	-0.085	0.036	0.019	**
NE Action 2	-0.159	0.034	0	****	-0.159	0.034	0	****	-0.159	0.034	0	****

Log-Likelihood (Model 1): -14505.105. Log-Likelihood (Model 2): -14507.422. Log-Likelihood (Model 3): -14510.319. LR Test (Model 1→Model 2): $\chi^2(13)=4.63$, $P=0.982$. LR Test (Model 1→Model 3): $\chi^2(17)=10.43$, $P=0.885$. LR Test (Model 2→Model 3): $\chi^2(4)=5.79$, $P=0.215$. White robust estimators of the variance were used to compute standard errors. Numbers are approximated to the third decimal value. ****, ***, **, * and * stand for significance at the 0.001, 0.01, 0.05 and 0.1 levels, respectively.

FIGURE 36: Ordered Probit Regressions on NDecided2, Games with Multiple Pure Strategy Nash Equilibria (n=5997)

8.4.3 Results

We can now summarize some results.

(1) *Go for high numbers, especially if they are yours.* This appears to be still true, as shown by the coefficient values on Own and Other's Payoff, on Same as PSPD, Same As MPD and Same as Utilitarian Best. On the one hand, as before, more weight is put on one's own payoff than on the other player's. On the other hand, consideration of potential trembles may make the network weigh the other player's welfare, both directly (Other's Payoff) and indirectly; utilitarianism appears a salient feature to choose among PNE.

(2) *Feel and fear trembling hands.* This is still true, as shown by the generally negative significant temptation effects. Coefficients are mostly larger for one's own temptation, but the reverse is true for the minimum temptation of the other player.

(3) *Strategic awareness.* The neural network appears capable of detecting relevant strategic features, as shown not only by the weight put on most BR temptation variables, but also by Same As Utilitarian Best (which requires the utilitarian solution to be a PNE) and, more importantly, by Same As NE When 2 NE. Considering the positive coefficient on Presence of 3NE, this means that, plausibly, the network makes more decided choices in relation to PNE actions when there are 2 PNE, then when there are 3 PNE, and last when it is not a PNE. Additional weight is given if the action is strictly dominant, even if just through 1-level iteration of strictly dominated strategies.

(4) *High stakes provide motivation for the agent.* This is still true, as shown by the large coefficient on Mean.

General game features appear to play less of a role in figures 35 and 36 than they did in figure 34. This is not really surprising: figure 34 was a regression in relation to the *best* (i.e. Nash equilibrium) action only, whereas here we are regressing in relation to all actions, good and bad, and these will share the same general game features. Hence, the

only way this may matter is in increasing the overall level of decided choices in the game - but not really in making the neural network more decisive towards one action relative to another.

In conclusion, C^* displays both w-focal points and s-focal points in games with multiple PNE. It does so because it tries to apply to these games, as far as possible, the same LMA that it has learnt to apply to games with unique PNE in a satisficing way. The existence of s-focal points corroborates once again the view that C^* mainly reasons on the basis of prototypical categorization; moreover, most of the game features it has learnt do not yield to undecidability problems when facing games with multiple PNE and so can be applied successfully to coordinate on focal choices. These results are the more striking because C^* has never faced a single game with multiple PNE during training.

8.5 GAMES WITH NO EQUILIBRIA

Unfortunately, the models of the previous section cannot be compared to those of section 8.3 in two important respects: (i) the usage of different endogenous variables, which has led to the usage of partially different exogenous variables (for example, based on the concept of Best Response); (ii) the application to different sets of games; games with a unique PNE in one case, and games with multiple PNE in the other case. A third limitation of the previous analysis is that, although we have analysed cases with 1, 2 or 3 PNE, for the sake of generality we should also analyse games with 0 PNE. After all, if the claims of the previous section are correct, we would expect the neural network to play in a meaningful bounded-rational way also in this set of games.

In this section we try to address these three concerns. To do this, we computed 6000 NDecided1 observations in relation to the games with unique PNE analysed in sections 8.2 and 8.3. We also computed 2000 games with 0 PNE, and used these games to test the thirty networks trained with learning rate of 0.5, momentum of 0 and thirty different random seeds; we were then able to derive 6000 NDecided1 observations in relation to these games with 0 PNE. Again we exclude the observations in relation to which there

is not a strictly dominant Best Response on the part of the column player: this leaves us with $n = 5993$ with the games with 1 PNE and $n = 5983$ with the games with 0 PNE. We then ran ordered Probit regressions using the same set of regressors as much as possible.⁴⁷ There are fewer regressors with 0 PNE games because, unavoidably, there are more strategic features which cannot be exploited, either by us, or by the network. The average number of decided choices per action in the 0 PNE dataset is 8.28, lower not only than games with a unique PNE but also than games with multiple PNE. Nevertheless, it is still a high number, given the maximum possible is 10.

Figures 37 and 38 contain the results of the ordered Probit regressions, in relation respectively to games with 1 and 0 PNE. The neural network still tends to go for high numbers, especially for itself. Same as OSD is wrongly (i.e., negatively) signed in figure 37, but this is really just a reduction of the positive effect that zero-iteration dominant action has on NDecided1. This is true, firstly, because Same as 1SD will always be equal to 1 when Same as OSD is equal to 1, and the coefficient on Same as 1st is large and positive; second, because, in the present sample, OSD is always a sufficient condition for PSPD and MPD, which implies an additional and large positive effect. Similarly, the occasional wrong signs on the Temptation variables can be explained because of a partial counterbalancing of the large interaction effects with Game Harmony. This is not the case, however, for Own and Other's Min Temptation with games with 0 PNE.

Game Harmony plays a significant role in games with unique and zero PNE. In line with the findings of section 8.4, a higher game harmony index, i.e. greater game disharmony, induces more decisive choices.

⁴⁷Some differences are made necessary by the way the datasets are constructed: for example, we clearly cannot use Same As NE as a regressor in the dataset of games with 0 PNE! Similarly, since the utilitarian criterion used in section 8.5 was made dependent on a choice among PNE, it cannot be introduced in the datasets with either 1 or 0 PNE.

Explanatory Variables	Model 1					Model 2					Model 3				
	Coef.	SE	Prob.	Sig.		Coef.	SE	Prob.	Sig.		Coef.	SE	Prob.	Sig.	
Same As NE	1.392	0.059	0	****		1.393	0.059	0	****		1.422	0.054	0	****	
Same As PSPD	0.495	0.049	0	****		0.498	0.049	0	****		0.538	0.043	0	****	
Same As MFD	0.28	0.04	0	****		0.28	0.04	0	****		0.281	0.039	0	****	
Same As Minmax	0.046	0.043	0.288			0.047	0.043	0.274							
Same As Rationalisability	0.08	0.07	0.247			0.082	0.069	0.234							
Same As OSD	-0.177	0.08	0.028	**		-0.176	0.08	0.028	**		-0.162	0.078	0.039	**	
Same As 1SD	0.265	0.075	0	****		0.261	0.074	0	****		0.319	0.06	0	****	
Positive Payoff Ratio	-0.085	0.057	0.133			-0.074	0.056	0.191			-0.02	0.038	0.592		
Negative Payoff Ratio	-0.181	0.207	0.382			-0.156	0.194	0.421			-0.371	0.135	0.006	***	
Own BR Max Temptation	-1.453	0.101	0	****		-1.398	0.097	0	****		-1.342	0.091	0	****	
Other's BR Max Temptation	-0.93	0.164	0	****		-0.845	0.158	0	****		-0.835	0.155	0	****	
Own BR Min Temptation	0.211	0.09	0.019	**		0.226	0.089	0.011	**		0.255	0.088	0.004	***	
Other's BR Min Temptation	-0.301	0.104	0.004	***		-0.285	0.103	0.006	***		-0.247	0.101	0.015	**	
Own BR Payoff	2.182	0.343	0	****		2.242	0.325	0	****		2.026	0.285	0	****	
Other's BR Payoff	0.012	0.409	0.977												
Own Max Temptation	0.088	0.101	0.384			0.083	0.101	0.413			-0.183	0.072	0.011	**	
Other's Max Temptation	-0.152	0.085	0.071	*		-0.139	0.084	0.099	*						
Own Min Temptation	0.145	0.109	0.183			0.116	0.108	0.28							
Other's Min Temptation	-0.078	0.098	0.428			-0.088	0.098	0.369							
Own Generic Temptation	-4.081	0.308	0	****		-4.041	0.303	0	****		-3.959	0.285	0	****	
Other's Generic Temptation	-1.092	0.195	0	****		-1.104	0.195	0	****		-1.161	0.193	0	****	
NE Action 1	-0.103	0.037	0.006	***		-0.102	0.038	0.006	***		-0.103	0.038	0.006	***	
NE Action 2	-0.136	0.037	0	****		-0.136	0.037	0	****		-0.136	0.037	0	****	
Game Harmony Index	1.919	1.491	0.198			2.51	0.518	0	****		2.028	0.409	0	****	
Game Harmony * Own BR Temptation	-2.057	0.672	0.002	***		-2.055	0.635	0.001	***		-2.016	0.623	0.001	***	
Game Harmony * Other's BR Temptation	-0.797	0.952	0.402			-0.638	0.423	0.131							
Mean	6.72	1.3	0	****		7.055	0.661	0	****		6.974	0.634	0	****	
Game Harmony * Mean	0.747	3.234	0.817			-0.499	0.387	0.197							
Deviation from Avg. Mean	-0.438	0.556	0.431												
Standard Deviation	1.39	0.887	0.117												
Deviation from Avg. SD	0.112	0.878	0.898												
Skewness	0.105	0.07	0.133			0.125	0.068	0.068	*		0.125	0.068	0.067	*	
Kurtosis	-0.015	0.059	0.799												
Deviation from Avg. Skewness	0.057	0.104	0.586												
Deviation from Avg. Kurtosis	0.016	0.063	0.795												

Log-Likelihood (Model 1): -10602.456. Log-Likelihood (Model 2): -10605.244. Log-Likelihood (Model 3): -10610.995. LR Test (Model 1→Model 2): $\chi^2(7)=5.58$, $P=0.590$. LR Test (Model 1→Model 3): $\chi^2(14)=17.09$, $P=0.252$. LR Test (Model 2→Model 3): $\chi^2(7)=11.51$, $P=0.118$. White robust estimators of the variance were used to compute standard errors. Numbers are approximated to the third decimal value. ****, ***, **, * and * stand for significance at the 0.001, 0.01, 0.05 and 0.1 levels, respectively.

FIGURE 37: Ordered Probit Regressions on NDecided1, Games with a Unique Pure Strategy Nash Equilibrium (n=5993)

Explanatory Variables	Model 1					Model 2					Model 3				
	Coef.	S.E.	Prob.	Sig.		Coef.	S.E.	Prob.	Sig.		Coef.	S.E.	Prob.	Sig.	
Same As PSPD	0.745	0.038	0	****		0.735	0.035	0	****		0.733	0.035	0	****	
Same As MPD	0.579	0.033	0	****		0.58	0.033	0	****		0.58	0.033	0	****	
Same As Minmax	-0.022	0.037	0.556												
Positive Payoff Ratio	-0.067	0.058	0.247			-0.081	0.056	0.146			0.663	0.157	0	****	
Negative Payoff Ratio	0.656	0.189	0.001	***		0.748	0.165	0	****		-0.56	0.079	0	****	
Own BR Max Temptation	-0.552	0.081	0	****		-0.563	0.079	0	****						
Other's BR Max Temptation	0.122	0.131	0.355												
Own BR Min Temptation	-0.097	0.099	0.325												
Other's BR Min Temptation	0.295	0.092	0.001	***		0.317	0.088	0	****		0.282	0.073	0	****	
Own BR Payoff	2.139	0.317	0	****		2.248	0.296	0	****		2.336	0.191	0	****	
Other's BR Payoff	0.354	0.373	0.343												
Own Max Temptation	-0.296	0.073	0	****		-0.294	0.072	0	****		-0.293	0.072	0	****	
Other's Max Temptation	0.05	0.076	0.509												
Own Min Temptation	0.313	0.072	0	****		0.306	0.072	0	****		0.308	0.072	0	****	
Other's Min Temptation	0.299	0.074	0	****		0.317	0.068	0	****		0.302	0.064	0	****	
Own Generic Temptation	-1.818	0.2	0	****		-1.846	0.187	0	****		-1.865	0.186	0	****	
Other's Generic Temptation	-0.195	0.183	0.288			-0.116	0.175	0.506							
NE Action 1	-0.277	0.034	0	****		-0.277	0.034	0	****		-0.277	0.034	0	****	
NE Action 2	-0.202	0.033	0	****		-0.203	0.033	0	****		-0.204	0.033	0	****	
Game Harmony Index	2.685	1.257	0.033	**		1.818	0.349	0	****		1.889	0.242	0	****	
Game Harmony * Own BR Temptation	0.512	0.586	0.383			0.518	0.586	0.377							
Game Harmony * Other's BR Temptation	-3.006	0.767	0	****		-2.309	0.298	0	****		-2.119	0.255	0	****	
Mean	0.208	0.927	0.822												
Game Harmony * Mean	-0.851	2.115	0.687												
Deviation from Avg. Mean	-0.303	0.426	0.476												
Standard Deviation	0.124	0.604	0.837												
Deviation from Avg. SD	-0.129	0.791	0.871												
Skewness	-0.041	0.057	0.475												
Kurtosis	0.027	0.051	0.593												
Deviation from Avg. Skewness	-0.02	0.088	0.821												
Deviation from Avg. Kurtosis	0.007	0.055	0.905												

Log-Likelihood (Model 1): -15596.259. Log-Likelihood (Model 2): -15598.954. Log-Likelihood (Model 3): -15600.497. LR Test (Model 1 → Model 2): $\chi^2(14)=5.39$, $P=0.980$; LR Test (Model 1 → Model 3): $\chi^2(17)=8.48$, $P=0.955$; LR Test (Model 2 → Model 3): $\chi^2(3)=3.09$, $P=0.378$. White robust estimators of the variance were used to compute standard errors. Numbers are approximated to the third decimal value. ****, ***, ** and * stand for significance at the 0.001, 0.01, 0.05 and 0.1 levels, respectively.

FIGURE 38: Ordered Probit Regressions on NDecided1, Games with Zero Pure Strategy Nash Equilibria (n=5983)

8.6 CONCLUSIONS

This chapter presented a simulated neural network designed to model the endogenous emergence of bounded-rational behaviour in a normal form game framework. Potentially any finite normal form could be modelled in this way, though we have concentrated on 3×3 games, and noted that 2×2 , 2×3 and 3×2 games count as a subclass of 3×3 . The inclusion of a neural network player in a population of Nash players does not change the behaviour of the Nash players, and the neural network player, having seen a sufficiently large sample of example games in which the Nash outcome was high-lighted also attempts to select Nash equilibria. However, based on the findings in chapter 7 we predict that a neural network will try but fail to achieve 100% success. Rather, it will instead prove satisfied with an algorithm that provides a lower success rate.

The simulation results suggest a figure of around 60% success on games never encountered before, as compared with 33% as the random success benchmark or the 59.6% experimental figure from Stahl and Wilson (1994). Such simulations also indicate that solution concepts other than Nash and based on dominance get closer to explaining the simulated network's actual behaviour. The network displays some strategic awareness, but this is not unbounded, and is decreasing in the levels of iterated deletion of dominated strategies required. The network goes for high payoff values. It takes into account the temptation of the other player to deviate from Nash. It plays better in higher stakes games, particularly if there is a greater conflict of interests between itself and the other player.

The trained network's behavioural heuristics carry over to a relevant degree when it faces not just new games, but new classes of games, namely games with multiple and zero pure Nash equilibria. Moreover, networks trained on different games, all with a unique pure Nash equilibrium, are able to coordinate on the same focal solution, when encountering games with multiple equilibria.

Our results suggest that, perhaps paradoxically, the fact that the network converges to a local error-minimizing algorithm is not a problem but a virtue: it is what makes

this modelling approach potentially interesting. Neural network simulations may be a promising tool to model and make predictions about bounded-rational behaviour in normal form games, with rules of thumb emerging endogenously as a result of the learning process rather than being exogenously super-imposed on the agent.

Chapters 7 and 8 have pointed the way towards understanding what a model of learning based on the actual functioning of the human brain can be expected to achieve, and how it might go about attempting to perform well. There is still much to do in firstly fully understanding the sort of complex rules the network finally follows, and secondly in applying our current understanding of neural networks to new problems in learning. Nevertheless it is hoped that a start has been made on an innovative and biologically plausible way of modelling bounded rationality.

APPENDIX FOR CHAPTER 8

The Game Harmony Index

Let a_{sj} be the payoff of player s for some given normal form game outcome j (out of m possible outcomes), where the game has n players.⁴⁸ The Gini-based index of Game Harmony can be computed in three steps. The first step is *ratio-normalization* of payoffs, by finding, for each payoff value, $a_{sj}^* = a_{sj} / \sum a_{sj}$, i.e., by dividing each payoff by the sum of all possible payoffs for the player. Ratio-normalization is essential because, otherwise, the measure would mirror the average payoff distribution of the game, but not the extent to which the players' interests in achieving one or another game outcome are the same or are in conflict.

The second step is to compute the payoff distribution index for each possible outcome j , using the Gini Index formula, but multiplied by $n/(n - 1)$ to ensure it is bounded between 0 and 1 with non-negative payoffs: labelling this normalized index as I_j^G for some outcome j , and ordering subjects from “poorest” (i.e., that with the lowest payoff) to “wealthiest” (i.e., that with the highest payoff), we can define I_j^G as:

$$I_j^G = \frac{n}{n-1} \frac{2}{n \sum_{s=1}^n a_{sj}^*} \sum_{s=1}^n s \left(a_{sj}^* - \frac{\sum_{s=1}^n a_{sj}^*}{n} \right) = \frac{1}{n-1} \frac{2}{\sum_{s=1}^n a_{sj}^*} \sum_{s=1}^n s \left(a_{sj}^* - \frac{\sum_{s=1}^n a_{sj}^*}{n} \right)$$

The third step is to find the normalized Gini-based game harmony index GH_G as a simple average of the I_j^G , so $GH_G = (1/m) \sum_{j=1}^m I_j^G$. In the case considered in this chapter, $n = 2$ and $m = 9$ (since games are 3×3).

⁴⁸The Game Harmony Index used in this chapter was designed by Daniel J. Zizzo.

CHAPTER 9

CONCLUSION

This thesis has focused on two new areas in learning theory: herding theory and neural network learning. Despite the significant difference between the two, both concern the difficulties faced by agents attempting to make the best possible decision, and both provide a great deal of scope for future work.

9.1 IMPERFECT LEARNING

Herding theory is firmly based on a clear notion of rationality, and follows the standard methodology of Bayesian updating. Despite this we have seen that suboptimality can result from the most ardent attempt to optimize, coming in the form of utility losses to individual economic agents or direct damage to a firm's profitability. A neural network is perhaps the best current model of biological learning, but is better described as a bounded-rational learner, and also will not produce consistently perfect results. The interesting parallel here is that Bayesian updating need not always produce outcomes that are superior to those generated by a bounded-rational learner, as in both instances the best we can hope for is success in a majority of cases. Although the arenas faced by the learners in this thesis were very different, the basic problem is always the same: making the best use of available information. For herding theory the key notion is that observing others is important, and it would appear unreasonable to leave out such considerations when optimizing. For neural networks there is a similar message: generalization from

examples is what humans are good at, so to completely ignore this function of the human brain will surely result in a failure to understand learning.

For those working in learning theory there are many choices. Theorists can continue to assume high levels of rationality, in the understanding that this does not presuppose the optimality of outcomes; and we have seen that the findings of chapters 2 to 6 certainly support this. Theorists may also conclude that it is not the structure of the problem (as in herding), but rather the nature of agents themselves (as in neural network learning) that is the difficulty at the heart of so much real-world suboptimality, as we saw in chapters 7 and 8.

We might also challenge the dichotomy stressed in chapter 7 between an assumption of full rationality and the simplification within evolutionary game theory, in which learners begin as myopic decision-makers and may one day approach rational behaviour, but should perhaps be reset back to zero when a new situation is met. The middle ground of bounded rationality has up to now been mainly populated by theories of limited memory and finite automata. Neural networks provide a clear alternative: it is the processing abilities of agents which may provide us with a new way to examine bounded rationality. Furthermore this is all based on a biologically sound set of principles. Although rules of thumb are present, they emerge endogenously and will not be far from optimal in many cases; they are not simply imposed on the learner from above.

9.2 FUTURE WORK

Herding is now a more or less well understood theory, but there are many more potential applications. Chapter 3 presented one possible change to the basic structure of a herd, but begs the question: what is the very best structure obtainable? This must be counter-balanced with the costs of providing such a structure. Some information structures may be already embedded in society, such as the side-effect of marriage, briefly explored in chapter 3. The second part of the thesis allowed agents to produce their own ordering through endogenous timing, but this option will not always be available, and similarly

leads to imperfections, as explored in chapters 4 to 6. We might wonder how a firm should best organize its decision-making: in pairs, in sequence, or through some complex pyramid structure? This provides a linkage between herding theory and the theory of the firm. We might ask how employers should treat new information about the actions of their employees, in the light of the dangers of the herd externality. Similarly governments and regulators might wish to free-ride on each others' information, or might face firms or consumers trapped in herds. Economic policy should at least consider the dangers of a very natural part of learning: observing others.

For neural network learning the scope for future work is even greater as this is a theoretical device which is not as yet fully developed. Increasing the biological plausibility of the learning process is one route, though it already far exceeds commonly accepted forms of learning used throughout economics. The main challenge is perhaps to delve within the process and ask exactly what the network is doing; indirectly, asking how people actually process information. Chapters 7 and 8 make a start with reference to a specific class of games, attempting to add a particular brand of formalism to the much used phrase, "rule of thumb". There are two issues here: how such rules are formed; and how such rules are used or changed. The neural network model used in this thesis assumed a particular (albeit popular and biologically plausible) method of learning, but we must wonder whether learning ever actually stops, or whether agents keep learning even after childhood ends, and the training period has ended. More complex neural network models, which retain the edge of biological plausibility, can certainly be envisaged which could be designed to tackle just such issues of ongoing learning. However, since the last two chapters present the very beginnings of an examination into the possible uses of neural networks, perhaps more basic applications should be attempted first. For example, to different classes of games, or to certain very particular games. Rubinstein (1993) uses a very simplified form of neural network to analyse monopolistic competition, but his methodology is very different. He assumes firms *use* neural networks to make decisions, rather than assuming that agents effectively *are* neural networks for the purposes of learning. There are many other possible arenas for this work, whether it be modelling

imperfect regulators or governments, or more complex strategic interactions.

Neural networks are founded on areas of research within mathematics and engineering that are relatively recent: machine learning and algorithm complexity theory. Furthermore they actually present a model which while far simpler than the human brain, genuinely attempts to model the process of learning, not merely attempting to capture basic notions of reinforcement. This thesis represents a start at incorporating recent thinking in such areas into economics. Herding models for all their attractive mathematical simplicity cannot claim to actually focus on more than one new aspect of learning: the desire to learn more about a problem. Nevertheless both are recent new ways of looking at new and old problems in learning theory and economics in general, and as shown in this thesis there is much that these new methods can add to the body of economic knowledge.

BIBLIOGRAPHY

- ABREU, D., AND A. RUBINSTEIN (1988): "The Structure of Nash Equilibria in Repeated Games with Finite Automata," *Econometrica*, 56, 1259–1282.
- ALLSOPP, L., AND J. HEY (1998): "Two Experiment to Test a Model of Herd Behaviour," Discussion paper, No. 98-28. The University of Birmingham.
- ANDERSON, J. A. (1998): "Learning Arithmetic with a Neural Network: Seven Times Seven is about Fifty," in *Methods, Models and Conceptual Issues: An Invitation to Cognitive Science*, vol. 4, ed. by D. Scarborough, and S. Sternberg, pp. 255–299. MIT Press, Cambridge, Mass. and London.
- ANDERSON, R., AND C. HOLT (1997): "Information Cascades in the Laboratory," *American Economic Review*, 87, 847–862.
- ANTHONY, M., AND P. L. BARTLETT (1999): *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, Cambridge.
- AUER, P., M. HERBSTER, AND M. K. WARMUTH (1996): "Exponentially Many Local Minima for Single Neurons," in *Advances in Neural Information Processing Systems 8*, ed. by D. S. Touretzky, M. C. Mozer, and M. E. Hasslemo, pp. 316–322. MIT Press, Cambridge, Mass. and London.
- BANDURA, A. (1977): *Social Learning Theory*. Prentice-Hall, Englewood Cliffs.
- BANERJEE, A. V. (1992): "A Simple Model of Herd Behaviour," *Quarterly Journal of Economics*, 107, 797–817.

- BIKCHANDANI, S., D. HIRSHLEIFER, AND I. WELCH (1992): "A Theory of Fads, Fashion, Custom and Cultural Change as Informational Cascades," *Journal of Political Economy*, 100, 992–1026.
- CHAMLEY, C., AND D. GALE (1994): "Information Revelation and Strategic Delay in a Model of Investment," *Econometrica*, 62, 60–74.
- CHO, I.-K., AND T. J. SARGENT (1996): "Neural Networks for Encoding and Adapting in Dynamic Economies," in *Handbook of Computational Economics, Volume 1*, ed. by H. M. Amman, D. A. Kendrick, and J. Rust. Elsevier, North Holland.
- CORMEN, T. H., C. E. LEISERSON, AND R. L. RIVEST (1992): *Introduction to Algorithms*. MIT Press, Cambridge, Mass. and London.
- ELMAN, J. L., E. A. BATES, M. H. JOHNSON, A. KARNILOFF-SMITH, D. PARISI, AND K. PLUNKETT (1996): *Rethinking Innateness: A Connectionist Perspective on Development*. MIT Press, Cambridge, Mass.
- FUKUMIZU, K., AND S. AMARI (2000): "Local Minima and Plateaus in Hierarchical Structures of Multilayer Perceptrons," *Neural Networks*, 13, 317–327.
- GALE, D. (1996): "What Have We Learned from Social Learning?," *European Economic Review*, 40, 617–628.
- GAREY, M. R., AND D. S. JOHNSON (1979): *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Bell Telephone Laboratories, Incorporated, United States of America.
- GILBOA, I., AND D. SCHMEIDLER (1995): "Case Based Decision Theory," *Quarterly Journal of Economics*, 109, 605–639.
- GRENADER, U. (1981): *Abstract Inference*. Wiley, New York.

- GUL, F., AND R. LUNDHOLM (1993): "On the Clustering of Agents' Decisions: Herd Behaviour versus the Endogenous Timing of Actions," Discussion paper, Ann Arbor: University of Michigan.
- (1995): "Endogenous Timing and the Clustering of Agents' Decisions," *Journal of Political Economy*, 103, 1039–1066.
- HARSANYI, J., AND R. SELTEN (1988): *A General Theory of Equilibrium Selection in Games*. MIT Press, Cambridge, Mass.
- HORNIK, K., M. B. STINCHOMBE, AND H. WHITE (1989): "Multi-Layer Feedforward Networks Are Universal Approximators," *Neural Networks*, 2, 359–366.
- HUCK, S., AND J. OECHSSLER (1997): "Informational Cascades with Continuous Action Spaces," Discussion paper, Department of Economics, Humboldt University.
- (2000): "Informational Cascades in the Laboratory: Do They Occur for the Right Reason?," Discussion paper, Department of Economics, University of Bonn.
- HUTCHINS, E., AND B. HAZELHURST (1991): "Learning in the Cultural Process," in *Artificial Life II*, ed. by C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen, pp. 689–705. Addison-Wesley, Redwood City, California.
- JACKSON, M. D., AND E. KALAI (1997): "Social Learning in Recurring Games," *Games and Economic Behavior*, 21, 102–134.
- KANDORI, M., G. J. MAILATH, AND R. ROB (1993): "Learning, Mutation and Long-Run Equilibria in Games," *Econometrica*, 61, 29–56.
- LEE, I. H. (1993): "On the Convergence of Information Cascades," *Journal of Economic Theory*, 61, 395–411.
- MACY, M. (1996): "Natural Selection and Social Learning in Prisoner's Dilemma: Co-adaptation with Genetic Algorithms and Neural Networks," in *Frontiers in Social*

- Dilemma Research*, ed. by W. B. G. Liebrand, and D. M. Messick, pp. 235–265. Verlag, Berlin.
- MAILATH, G. J. (1998): “Do People Play Nash Equilibrium? Lessons from Evolutionary Game Theory,” *Journal of Economic Literature*, 36, 1347–1374.
- MEYER, M. A. (1991): “Learning from Coarse Information: Biased Contests and Career Profiles,” *Review of Economic Studies*, 58, 15–41.
- MILGROM, P. R. (1981): “Good News and Bad News: Representation Theorems and Applications,” *Bell Journal of Economics*, 12, 380–391.
- MOSCARINI, G., M. OTTAVIANI, AND L. SMITH (1996): “Social Learning in a Changing World,” Discussion paper, Department of Economics, MIT.
- NEYMAN, A. (1985): “Bounded Rationality Justifies Cooperation in the Finitely Repeated Prisoners’ Dilemma Game,” *Economics Letters*, 19, 227–229.
- OSBORNE, M. J., AND A. RUBINSTEIN (1998): “Games with Procedurally Rational Players,” *American Economic Review*, 88, 834–847.
- PLUNKETT, K., AND C. SINHA (1992): “Connectionism and Development Theory,” *British Journal of Developmental Psychology*, 10, 209–254.
- RADNER, R. (1982): “Equilibrium Under Uncertainty,” in *Handbook of Mathematical Economics, Vol II*, ed. by K. Arrow, and M. D. Intriligator. North-Holland, Amsterdam.
- ROTH, A. E., AND I. EREV (1995): “Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term,” *Games and Economic Behavior*, 8, 164–212.
- (1998): “Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria,” *American Economic Review*, 88, 848–881.

- ROTHSCHILD, M. (1974): "A Two-Armed Bandit Theory of Market Pricing," *Journal of Economic Theory*, 9, 185–202.
- RUBINSTEIN, A. (1993): "On Price Recognition and Computational Complexity in a Monopolistic Model," *Journal of Political Economy*, 101, 473–484.
- RUBINSTEIN, A. (1998): *Modeling Bounded Rationality*. MIT Press, Cambridge, Mass. and London.
- RUMELHART, D. E., R. J. HINTON, AND R. J. WILLIAMS (1986): "Learning Representations by Back-Propogating Error," *Nature*, 323, 533–536.
- SCHMAJUK, N. A. (1997): *Animal Learning and Cognition: A Neural Network Approach*. Cambridge University Press, Cambridge.
- SGROI, D. (1998): *Endogenous Timing in Models of Social Learning*. M.Phil. Thesis, University of Oxford.
- (2000a): "Optimizing Information in the Herd: Guinea Pigs, Profits and Welfare," Discussion paper, No. W14-2000. Nuffield College, Oxford. Forthcoming in *Games and Economic Behavior*.
- (2000b): "The Right Choice at the Right Time: A Herding Experiment in Endogenous Time," Discussion paper, No. W15-2000. Nuffield College, Oxford.
- SIMON, H. (1955): "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics*, 69, 99–118.
- (1959): "Theories of Decision-Making in Economics and Behavioral Science," *American Economic Review*, 49, 253–283.
- SMITH, L., AND P. SORENSEN (1996): "Pathological Outcomes of Observational Learning," Discussion paper, Department of Economics, MIT.

- (1997a): “Informational Herding as Experimentation Deja Vu,” Discussion paper, Department of Economics, MIT, and No. W14-1997. Nuffield College, Oxford.
- (1997b): “Rational Social Learning with Random Sampling,” Discussion paper, Department of Economics, MIT, and Nuffield College, Oxford.
- SONTAG, E. D. (1995): “Critical Points for Least-Squares Problems Involving Certain Analytic Functions with Applications to Sigmoidal Nets,” *Advances in Computational Mathematics*, 5, 245–268.
- SONTAG, E. D., AND H. J. SUSSMANN (1989): “Backpropagation Can Give Rise to Spurious Local Minima even for Networks with No Hidden Layers,” *Complex Systems*, 3, 91–106.
- STAHL, D. O. (1998): “Population Rule Learning in Symmetric Normal-Form Games,” Discussion paper, University of Texas at Austin, Center for Applied Research in Economics Working Paper.
- STAHL, D. O., AND P. W. WILSON (1994): “Experimental Evidence on Players’ Models of Other Players,” *Journal of Economic Behaviour and Organization*, 25, 309–327.
- (1995): “On Players’ Models of Other Players: Theory and Experimental Evidence,” *Games and Economic Behavior*, 10, 218–254.
- TARABAN, R., AND J. M. PALACIOS (1994): “Exemplar Models and Weighted Cue Models in Category Learning,” in *Categorization by Humans and Machines*, ed. by G. V. Nakamura, R. Taraban, and D. L. Medin, pp. 91–127. Academic Press, San Diego.
- VAN GELDER, T. (1991): “What is the “D” in “PDP”? A Survey of the Concept of Distribution,” in *Philosophy and Connectionist Theory: Developments in Connectionist Theory*, ed. by W. Ramsey, and S. P. Stich. Lawrence Erlbaum, Hillsdale.

- VIVES, X. (1991): "How Fast do Rational Agents Learn?," *Review of Economic Studies*, 60, 329–348.
- (1996): "Social Learning and Rational Expectations," *European Economic Review*, 40, 589–601.
- WAY, E. C. (1997): "Connectionism and Conceptual Structure," *American Behavioral Scientist*, 40, 729–753.
- WELCH, I. (1992): "Sequential Sales, Learning, and Cascades," *Journal of Finance*, 47, 695–732.
- WHITE, H. (1992): *Artificial Neural Networks: Approximation and Learning Theory*. Blackwell, Cambridge and Oxford.
- YOUNG, H. P. (1993): "The Evolution of Conventions," *Econometrica*, 61(1), 57–84.
- ZIZZO, D. J. (2000a): "Implicit Learning of (Boundedly) Rational Behavior, commentary to K. E. Stanovich and R. F. West, Individual Differences in Reasoning: Implications for the Rationality Debate," *Behavioral and Brain Sciences*, 23.
- (2000b): *Relativity-Sensitive Behaviour in Economics*. D.Phil. Thesis, University of Oxford.
- ZIZZO, D. J., AND D. SGROI (2000): "Emergent Bounded Rational Behavior by Neural Networks in Normal Form Games," Discussion paper, No. W30-2000. Nuffield College, Oxford.

