Papers in Economic Theory and the Biological Foundations of Economics

by

Nikolaus Kasimatis

M.A., Simon Fraser University, 2008

Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

in the
Department of Economics
Faculty of Arts and Social Sciences

Approval

Name: Nikolaus Kasimatis

Degree: Doctor of Philosophy (Economics)

Title: *Papers in Economic Theory and the Biological Foundations of Economics*

Examining Committee: **Dr. Erik Kimbrough** (chair)
Assistant Professor

Dr. Arthur Robson
Senior Supervisor
Professor

Dr. Gregory Dow
Supervisor
Professor

Dr. Shih En Lu
Supervisor
Assistant Professor

Dr. Songzi Du
Supervisor
Assistant Professor

Dr. Bernard Crespi
Internal Examiner
Professor, Department of Biological Sciences

Dr. Yoram Halevy
External Examiner
Associate Professor, Vancouver School of Economics
University of British Columbia

Date Defended: 06 Nov 2015

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

c. as a co-investigator, collaborator or research assistant in a research project approved in advance,

or

d. as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

Abstract

Chapter 1 gives the introduction to this thesis, describing the three essays that are contained herein. Chapter 2 is joint work with Erik Kimbrough and Arthur Robson. The article investigates the evolutionary foundation for our capacity to attribute preferences to others. It develops a theoretical model of this ability that the authors call "Theory of Preferences" (*ToP*), and then shows that *ToP* yields a sharp, unambiguous advantage over less sophisticated approaches to strategic interaction. The chapter reports on experiments investigating *ToP* in a simpler version of the model. It finds highly significant learning of opponentsŠ preferences, providing strong evidence for the presence of *ToP* as in the model among subjects.

Chapter 3 studies the third party provision of information in a dynamic reputation model. Information is sold to consumers by a profit maximizing intermediary with monopoly access to information about a long run firm. The paper characterizes the optimal disclosure rule from the point of view of the intermediary, and shows that if consumers act as price takers in the market for information, then in every equilibrium the intermediary extracts from consumers the highest price possible for information. The resulting equilibrium is inefficient.

Chapter 4 extends a matching and bargaining model of decentralized trade first developed by Gale and Sabourian (2005). The extension considers a market in which sellers bring to market several units rather than just one. The article then studies the effect on efficiency of an aversion to complexity among the agents. It shows that complexity aversion can preclude efficient exchange. A several-unit seller must consider, not only the price at which he currently trades, but also the effect of his exchange on future market conditions. A seller with several units thus attempts to manipulate the price in the future by engaging in inefficient trades currently.

Keywords: Bargaining, Decentralized Markets, Information Disclosure, Theory of Mind

Dedication

To Donna, Lauretta, and Lucia

Acknowledgements

I would like to thank my advisors, Arthur Robson, Greg Dow, Shih En Lu, and Songzi Du, for their support and guidance. Simon Woodcock, who served as graduate chair throughout my studies, was also very helpful and supportive.

Arthur acted as my senior supervisor. I have been extremely lucky to have him as a mentor and as a friend. The opportunity to co-author with him has provided me an invaluable learning experience. He has also put a tremendous amount of time into reading the various versions of my papers and into helping me improve them.

I made some great friends during my graduate studies—Josh Boitnott, Nick Dadson, Cristoph Eder, and Ideen Riahi. The staff and faculty of the department made it a wonderful place to work and study. Among the staff, Gwen Wild, Kathleen Vieira-Ribeiro, and Kathy Godson have been especially kind.

Finally, I must thank Donna who has put up with me since time immemorial and given me two wonderful daughters.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Chapter 2 of this thesis is joint work with Erik Kimbrough, and Arthur Robson. The article considers the evolutionary foundation for our ability to attribute preferences to others, an ability that we refer to as "Theory of Preferences" *(ToP)*. The article develops a theoretical model of *ToP* and then gives an answer to the question: Why might *ToP* have evolved? Our answer is that *ToP* is an evolutionary adaptation to strategic environments with persistent novelty.

ToP is a key component of "Theory of Mind", the more general ability to attribute mental states to oneself and to others—mental states such as belief, desire, knowledge, and intent. It moreover plays a crucial role in game theory. Unless a player has a dominant choice in a game, her best response depends on the choices of her opponents, and thus indirectly their preferences.

The chapter considers a dynamic model in which players repeatedly interact. A perfect information game tree is fixed, with fixed terminal nodes, but there are various outcomes that are assigned to these terminal nodes in a flexible fashion. In particular, the outcomes are randomly drawn in each iteration of the game from a finite outcome set, where this outcome set grows over time, through the introduction of novel outcomes.

Individuals know how their own utility functions are defined on all these outcomes, but do not know the preferences of their opponents. A type of agent, the *ToP* player, conceives of the other agents as having preferences, and endeavors to learn these. These players are contrasted with a naive type of agent that is essentially a reinforcement learner, viewing each subgame they initiate as a distinct indivisible circumstance.

The ability to attribute preferences enables players to better deal with the innovation that arises from new outcomes than can the naive players that adapt to each subgame as a distinct circumstance. The edge to sophistication derives from a capacity to extrapolate to novel circumstances information that was learned about others' preferences in a previous situation.

In order to corroborate the theoretical results we study *ToP* among laboratory subjects. The experimental results provide strong evidence for the presence among subjects of *ToP* as formulated in the paper. Moreover, we find that *ToP* is significantly correlated with theory of mind as conceived by psychologists.

Chapter 3 studies the third party provision of information in a dynamic model of reputation. At each date a long run firm meets with a short run consumer in order to play a simple normal form game. The firm has private information about its own payoffs, and has, moreover, a short run incentive to exploit his opponent. The consumer arrives in ignorance of the events having transpired prior to his arrival, and therefore knows very little about his opponent.

The limited knowledge of the short run agent introduces a role for a third party, an information-providing intermediary. The intermediary has a monopoly on information about the past choices of the firm, and is a profit maximizer. At the outset of the game he commits to a disclosure policy consisting of a specific way of mapping the histories of the game to signals. Each consumer, upon arriving to the market, must then pay for an informative signal in order to learn about the firm.

The basic problem of the intermediary is the following. His choice of disclosure policy generates the reputational incentives of the firm, and hence indirectly affects the firm's behavior. The firm's behavior in equilibrium, on the other hand, determines the value of information to the consumer.

If the policy imposes too much discipline on the firm, then information is of little value to consumers. That is, if consumers are sure the firm will engage in good behavior, then they gain nothing from acquiring further information about their opponent. If the policy generates weak reputational incentives for the firm, the signal will again be of little value.

The chapter characterizes the disclosure policy adopted in equilibrium by the intermediary. The equilibrium policy results in an inefficient outcome. In particular, there are policies that deliver a first best outcome, but these are never chosen by a profit maximizing intermediary.

Chapter 4 considers a matching and bargaining model of decentralized trade among a finite number of agents. The article extends the model of Gale and Sabourian (2005) by permitting sellers to enter the market with several units for sale rather than just one.

Gale and Sabourian (2005) develop a complexity aversion refinement that selects competitive equilibria in their model. [1] In particular, their argument goes, if agents are averse to complexity, other things equal, then in equilibrium each player will choose the least complex strategy that earns the equilibrium payoff; as a result, simple behavior will be adopted, which, in turn, ensures Walrasian exchange.

This argument provides a rationale for selecting competitive equilibria within an important framework that describes exchange as resulting from strategic bargaining between pairwise matches. [2] This aspect of their work merits emphasis as it has been observed, since Rubinstein and Wolinsky (1990), that strategic behavior in such settings can result in non-competitive outcomes—even in the absence of trading frictions such as search costs and asymmetric information.

Of central importance to the above argument is the result that simple behavior gives rise to competitive exchange. The appeal of the above refinement then relies on the extent to which this result holds more generally in frictionless markets. Chapter 4 then asks if simple behavior will deliver Walrasian outcomes in more general environments.

By examining the more general case, in which sellers can sell many units rather than just one, this article delivers insights that are drastically different from those obtained by Gale and Sabourian (2005). In particular, it argues that simple behavior—resulting from complexity aversion, or for any other reason—does not suffice to establish competitive outcomes. In fact, it shows that when behavior is consistent with a perfect equilibrium simple behavior can rule out efficiency altogether.

[1] A competitive equilibrium here refers to an equilibrium of the matching and bargaining game in which all exchange occurs at market clearing prices, and in which the resulting allocation maximizes the gains from trade. This will also be referred to as a *Walrasian equilibrium*.

[2] These models are intended to give an explicit account of the inner workings of a decentralized market. See, for example, Rubinstein and Wolinsky (1985), Gale (1986a, 1986b), Rubinstein and Wolinsky (1990), and the recent contribution, Lauermann (2012).

Chapter 2

The Evolution of “Theory of Preferences”: Theory and Experiments

Joint work with Erik Kimbrough, and Arthur Robson

2.1 Introduction

Conventional game theory relies on agents correctly ascribing preferences to the other agents. Unless an agent has a dominant strategy, that is, her optimal choice depends on the choices of others and therefore indirectly on their preferences. We consider here the genesis of the strategic sophistication necessary to acquire others' preferences.

We address the questions: *Why* and *how* might this ability to impute preferences to others have evolved? In what types of environments would this ability yield a distinct advantage over alternative, less sophisticated, approaches to strategic interaction? In general terms, the answer we propose is that this ability is an evolutionary adaptation for dealing with strategic environments that have a persistent element of novelty.

Our interpretation of strategic sophistication is dynamic in that it entails *learning* other agents' preferences from their observed behavior. It also extends the theory of revealed preference in that knowing *others'* preferences has consequences for one's own actions. Throughout the paper, we refer to such strategic sophistication, for simplicity, as *ToP*, for "theory of preferences".

Our *ToP* is an aspect of "theory of mind", as in psychology. An individual with theory of mind has the ability to conceive of herself, and of others, as having agency, and so to attribute to herself and others mental states such as belief, desire, knowledge, and intent. It is generally accepted in psychology that human beings beyond infancy possess theory of mind. The classic experiment that suggests children have theory of mind is the "Sally-Ann" test described in Baron-Cohen, Leslie, and Frith (1985). According to this test, young children begin to realize that others may have beliefs they know to be false shortly after age four. This test relies on children's verbal facility. Onishi and Baillargeon (2005) push the age back to 15 months using a non-verbal technique. Infants are taken to express that their expectations have been violated by lengthening the duration of their gaze. The presence of this capacity in such young individuals increases the likelihood that it is, to some degree at least, innate.

The argument made here in favor of such strategic sophistication is a substantial generalization and reformulation of the argument in Robson (2001) concerning the advantage of having an own utility function in a non-strategic setting. In that paper, an own utility function permits an optimal response to novelty. Suppose an agent has experienced all of the possible outcomes, but has not experienced the particular gamble in question and so does not know the probabilities with which these are combined. This latter element introduces the requisite novelty. If the agent has the biologically appropriate utility function, she can learn the correct gamble to take; conversely, if she acts correctly over a sufficiently rich set of gambles, she must possess, at least implicitly, the appropriate utility function.

We consider here a dynamic model in which players repeatedly interact. Although the perfect information game tree is fixed, with fixed terminal nodes, there are various physical outcomes that are assigned to these terminal nodes in a flexible fashion. More particularly, the outcomes are randomly drawn in each iteration of the game from a finite outcome set, where this outcome set grows over time, thus introducing suitable novelty.

Individuals know how their own utility functions are defined on all these physical outcomes, but do not know the preferences of their opponents. There will be an advantage to an agent of sophistication—of effectively understanding that her opponents act optimally in the light of their preferences. Such a sophisticated agent can then learn opponents' preferences in order to exploit this information.

The sophisticated players are contrasted with naive players who are reinforcement learners, viewing each subgame they initiate as a distinct indivisible circumstance. Naive players condition in an arbitrary fashion on their own payoffs in each novel subgame. That is, their reinforcement learning is *initialized* in a general way.

Sophistication enables players to better deal with the innovation that arises from new outcomes than can such "naive" players that adapt to each subgame as a distinct circumstance.[1] The edge to sophistication derives from a capacity to extrapolate to novel circumstances information that was learned about others' preferences in a previous situation.[2]

Consider now our strategic environment in greater detail. We view the particular environment here as a convenient test-bed on which we can derive the speeds with which the various players can learn. The basic results do not seem likely to be specific to this particular environment, so these differences in relative learning speeds would be manifested in many alternative models.

We begin by fixing a game tree with perfect information, with I stages, say. There are I equally large populations, one for each of the stages or the associated "player roles." In each iteration of the game, a large number of random matches are made, with each match having one player in each role. The physical outcomes assigned to the terminal nodes are drawn randomly and uniformly in each iteration from the finite outcome set that is available then.

Players have preference orderings over the set of outcomes that are ever possible, and so preferences over the finite subset of these that is actually available in each

[1]The novelty here is circumscribed, but it is clear that evolution would be unable to deal with completely unrestricted novelty.

[2]The distinction between the *ToP* and naive players might be illustrated with reference to the following observations of vervet monkeys (Cheney and Seyfarth 1990, p. 213). If two groups are involved in a skirmish, sometimes a member of the losing side is observed to make a warning cry used by vervets to signal the approach of a leopard. All the vervets will then urgently disperse, saving the day for the losing combatants. The issue is: What is the genesis of this deceptive behavior? One possibility, corresponding to our *ToP* strategy, is that the deceptive vervet effectively appreciates what the effect of such a cry would be on the others, acts as if, that is, he understands that they are averse to a leopard attack and exploits this aversion deliberately. The other polar extreme corresponds to our naive reinforcement learners. Such a type has no model whatever of the other monkeys' preferences and beliefs. His alarm cry behavior conditions simply on the circumstance that he is losing a fight. By accident perhaps, he once made the leopard warning in such a circumstance, and it had a favorable outcome. Subsequent reapplication of this strategem continued to be met with success, reinforcing the behavior.

period. Each player is fully aware of her own utility function but does not directly know the preference ordering of his opponents.

At each date, at the start of each period, a new outcome is added to the set of potential outcomes, where each new outcome is drawn independently from a given distribution. The number of times the game is played within each period grows at a parametric rate, potentially allowing the preferences of other players to be learned.[3] The crucial aspect of this model is the introduction of novelty, rather than the growing complexity that is also generated.[4]

All players see the history of the games played—the outcomes that were chosen to attach to the terminal nodes in each iteration of the game, and the choices that were made by all player roles (but not, directly, the preferences of others). Players here differ with respect to the extent and the manner of utilization of this information.

All strategies use a dominant action in any subgame they face, if such an action is available. This is for simplicity, in the spirit of focussing on the implications of others's preferences, while presuming full utilization of one's own preferences. However, the current set up would permit such sequentially rational behavior to be obtained as a result rather than as an assumption.

Although the naive strategies can condition in an arbitrary way on their own observed payoffs in a novel subgame, it is crucial that they condition only on these payoffs. The other details of these naive strategies are not relevant to the main result. Indeed, even if the naive players apply a fully Bayesian rational strategy the *second* time a subgame is played, they will still lose the evolutionary race here to the *ToP* players. A slower and therefore more reasonable rate of learning for the naive players would only strengthen our results.

Once history has revealed the ordinal preferences of all subsequent players in any subgame to the *ToP* players, they choose a strategy that is a function of these ordinal preferences and their own. Furthermore, there is a particular *ToP* strategy, the *SPE-ToP* strategy, say, that not only observes subsequent preferences but uses the *SPE* strategy associated with these preferences and their own.

The *ToP* players know enough about the game that they can learn the preferences of other player roles, in the first place. In particular, it is common knowledge

[3] When there more outcomes already present, there is more to be learned about where a new outcome ranks.

[4] That is, in a model in which outcomes were also dropped, so the outcome set remained of constant size, similar results obtain, but in a slightly more awkward fashion.

among all the *ToP* players that there is a positive fraction of *SPE-ToP* players in every role.

It is not crucial otherwise how the *ToP* players behave—they could even *minimize* their payoffs according to a fully accurate posterior distribution over all the relevant aspects of the game, when the preferences of all subsequent players are not known.

We do not assume that the *ToP* players use the transitivity of opponents' preferences.[5] The *ToP* players build up a description of others' preferences only by observing all the pairwise choices. Generalizing this assumption could only strengthen our results by increasing *ToP* players' learning speed.

Between each iteration of the game, the fraction of each role that plays each strategy is updated to reflect the payoffs that this strategy obtains. This updating rule is subject to standard weak assumptions. In particular, the strategy that performs the best must increase at the expense of other strategies.

Theorem 2.1 is the main result here—for an intermediate range of values for a parameter governing the rate of innovation, a unique *SPE* is attained, with the *SPE-ToP* strategy ultimately taking over the population in each role, at the expense of all other strategies—naive or *ToP*.

Moreover, our results hold if the *ToP* incur a fixed per game cost. This is a key finding of the present paper since the previous literature has tended to find an advantage to (lucky and) less smart players over smarter players—see, for example, Stahl (1993). The underlying reasons for the reverse (and more plausible) result here are that, in the limit considered in Theorem 2.1, i) the naive players do not know the game they face while, at the same time, ii) the *SPE-ToP* players do know all the relevant preferences and, furthermore, have adapted to play the *SPE* strategy.

It is unambiguously better then to be "smart"—in the sense of *ToP* —than it is to be naive, no matter how lucky—even for the relatively mild form of naivete here.

After stating the theoretical results, we present experiments on theory of preferences that buttress the current approach by allowing us to observe 1) the presence and extent of our revealed preference version of theory of preferences in human subjects and 2) the degree to which this dynamic revealed preference interpretation of *ToP* corresponds to theory of mind as it is understood by psychologists. We

[5]Indeed, the results here would apply even if preferences were not transitive.

construct an environment similar to that in the model, but simpler, in which *ToP* yields a distinct strategic advantage, and observe the extent to which our subjects exploit this advantage.

In the experiments, subjects play a sequence of two-player extensive form games where each player role has two moves at each decision node. In each repetition, a game is constructed by drawing outcomes without replacement from a finite set. All players in a given role had the same (induced) preferences, but these players knew only their own payoff at each outcome and not that of their opponent, as is the crucial feature of the theoretical model. We randomly and anonymously paired subjects in each of 90 repetitions to observe the ability of players 1 to learn (and to exploit their knowledge of) the preferences of players 2. As reflects the theoretical model, many games in later periods that would appear novel to a naive reinforcement learner could be understood by an agent with *ToP* who had observed previous choices in the subgames. The rate at which subjects achieve subgame perfect equilibrium outcomes measures the extent to which individuals exhibit *ToP* by learning their opponents' preferences.

At the end of each experimental session, we collected two measures of theory of mind that are commonly used in psychology. Specifically, we asked the students to complete two short Likert scale surveys measuring the extent of autism spectrum behaviors. One was the Autism-Spectrum Quotient (AQ) survey due to Baron-Cohen et al. (2001); the other was the Broad Autism Phenotype Questionnaire (BAP), due to Hurley et al. (2007).

There were two striking results of the experiments that corroborate the present approach. First, we observed highly significant learning of player 2's preferences by players 1, but no such significant learning of specific games. That is, iron-clad support for the formal model of the paper is expressed in real-world behavior. Individuals do behave as if they ascribe preferences to opponents and endeavor to learn these, given that it is advantageous. Not surprisingly, this ability is present in real-world individuals to varying degrees. Second, there is strong evidence that this attribute is an aspect of theory of mind, as this term is understood in psychology: player 1's who report fewer autism-spectrum behaviors (i.e. have lower AQ and BAP scores) have a statistically significant tendency to learn player 2's preferences faster.

2.2 A Model

2.2.1 The Environment

Consider first the underlying games. The extensive game form is a fixed tree with perfect information and a finite number of stages, $I \geq 2$, and a fixed finite number of actions, $A \geq 2$, at each decision node.[6] Every complete history of the game then has I decision nodes and there are A^I terminal nodes.

There is one "player role" for each such stage, $i = 1, \ldots, I$, in the game. (In a reversal of the usual convention, the first player role to move is I and the last to move is 1. This simplifies the notation used in the proof. Role i therefore has a subgame of rank i in that there are i successor nodes in each path to a terminal node.) Each player role is represented by an equal-sized "large" population of agents, where these agents differ in their choice of strategy. The strategies are described precisely below, but they will be grouped into two "categories"—sophisticated (*ToP*) and naive.

Independently in each iteration of the game, all players are randomly and uniformly matched with exactly one player for each role in each of the resulting large number of games.

There is a fixed overall set of physically observable outcomes, each with consequences for the material payoffs of the I player roles. Player role $i = 1, \ldots, I$ has then a function mapping all outcomes to material payoffs. A fundamental novelty is that, although each player role knows her own payoff at each outcome, she does not know the payoffs for the other player roles.

For notational simplicity, however, we avoid the explicit construction of outcomes, with payoff functions defined on these. Given a fixed tree structure with T terminal nodes, we instead simply identify each outcome with a payoff vector and each game with a particular set of such payoff vectors assigned to the terminal nodes. We assume that all material payoffs are scalars, lying in the compact interval $[m, M]$, for $M > m > 0$, for simplicity.[7]

[6] The restriction that each node induce the same number of actions, A, can be relaxed. Indeed, it is possible to allow the game tree to be randomly chosen. This would not fundamentally change the nature of our results but would considerably add to the notation required.

[7] This abbreviated way of modeling outcomes introduces the apparent complication that the same payoff for role i might be associated with multiple possible payoffs for the remaining players. However, the set-up will be such that the probability of any role's payoff arising more than once, but with different payoffs for the other roles, is zero. Each player i can then safely assume that a

Assumption 2.1: *The set of all games is represented by* $Q = [m, M]^{TI}$*, for* $M > m > 0$*. That is, each outcome is a payoff vector in* $Z = [m, M]^I$*, with one component for each player role, and there are* T *such outcomes comprising each game.*

Let $n = 1, 2, \ldots,$ denote successive dates. Within each corresponding period, n, there is available a finite subset of outcomes $Z_n \subset Z$, determined in the following way. There is an initial finite set of outcomes $Z_0 \subset Z$, of size N, say, where each of these outcomes is drawn independently from Z according to a cumulative distribution function F as follows.

Assumption 2.2: *The cdf over outcomes* F *has a continuous probability density that is strictly positive on* Z*.*

At date $n \geq 1$, at the beginning of period n, a new outcome is added to the existing ones by drawing it independently from Z according to the same cdf F. Within each period, the set of outcomes is then fixed, and once an outcome is introduced it is available thereafter. Figure 2.1 is a schematic representation of the game.

We specify the number of games that are played within each period as follows.

Assumption 2.3: *The number of iterations of the game played in period* n *is* $\kappa(n) = \lfloor (N + n)^\alpha \rfloor$*, for some* $\alpha \geq 0$*.*[8]

If the parameter α is low, the rate of arrival of novelty is high in that there are not many games within each period before the next novel outcome arrives; if α is high, on the other hand, the rate of arrival of novelty is low.

Consider now a convenient formal description of the set of games available in each period.

Definition 2.1: *In period* n*, the* empirical *cdf based on sampling, with equal probabilities, from the outcomes that are actually available, is denoted by the random function* $F_n(z)$ *where* $z \in [m, M]^I$*. The set of games in period* n *is the* T*-times product of* Z_n*. This is denoted* Q_n*. The empirical cdf of games in period*

given own payoff is associated to a unique (but initially unknown) vector of other roles' payoffs. We then adopt this simpler set-up.

[8]Here $\lfloor \cdot \rfloor$ denotes the floor function. It seems more plausible, perhaps, that the number of games per period would be random. This makes the analysis mathematically more complex, but does not seem to fundamentally change the results. The present assumption is then in the interests of simplicity.

n derives from T-fold independent sampling of outcomes according to F_n and is denoted by $G_n(q)$, where $q \in Q = [m, M]^{IT}$.[9]

In each iteration, $t = 1, ..., \kappa(n)$, of the game in period n, outcomes are drawn *independently* from Z_n according to the cdf F_n, so the game is chosen *independently* in each iteration according to G_n.

The cdf's F_n and G_n are well-behaved in the limit. This result is elegant and informative and so is included here. First note that the distribution of games implied by the cdf on outcomes, F, is given by G, say, which is the cdf on the payoff space $[m, M]^{IT}$ generated by T independent choices of outcomes distributed according to F. Clearly, G also has a continuous pdf that is strictly positive on $[m, M]^{IT}$. These two later cdf's are then the limits of the cdf's F_n and G_n—

LEMMA 2.1: *It follows that $F_n(z) \to F(z)$ and $G_n(q) \to G(q)$ with probability one, and uniformly in $z \in [m, M]^I$, or in $q \in [m, M]^{IT}$, respectively.*

Proof: This follows directly from the Glivenko-Cantelli Theorem. (See Billingsley 1986, p. 275, and Elker, Pollard and Stute 1979, p. 825, for its extension to many dimensions.) ■

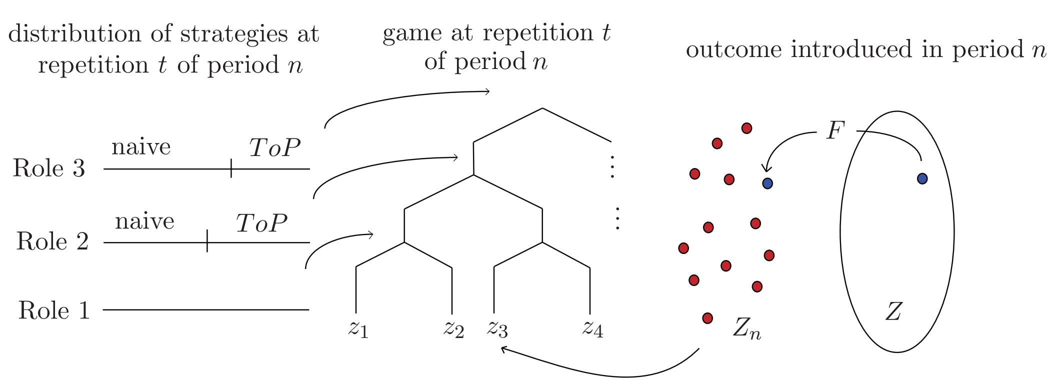

Figure 2.1: A Schematic Representation of the Key Elements of the Model.

We turn now to the specification of the strategies for each player role.

[9]Note that F_n and G_n are random variables measurable with respect to the information available in period n, in particular the set of available outcomes Z_n.

2.2.2 Strategies

When making a choice in period n and iteration t, every player, whether naive or *ToP*, knows the history so far, $H_{n,t}$, say, and the game, $q_{n,t}$, drawn for the current iteration. The history records the outcomes available in the current period, n, the randomly drawn games and the empirical distributions of choices made in all previous periods and iterations. Although each player observes the outcome assigned to each terminal node, as revealed by the payoff she is assigned at that node, it should be emphasized that she does not observe other roles' payoffs directly.

More precisely, for each player role i, given that decision-node h is reached by a positive fraction of players in period n and iteration t, let $\pi_{n,t}(h) \in \Delta(A)$ then record the aggregate behavior of i player roles at h. It follows that $H_{n,t} = \{Z_n, (q_{1,1}, \pi_{1,1}), \ldots, (q_{n,t-1}, \pi_{n,t-1})\}$.[10] Let $\boldsymbol{H}_{n,t}$ be the set of period n and iteration t histories, and let $\boldsymbol{H} = \bigcup_{n,t} \boldsymbol{H}_{n,t}$.

Strategies can be formally described as follows. Let Σ_i denote the set of choices available to the player role i's. A strategy is then a function $c : \boldsymbol{H} \times Q \longrightarrow \Sigma_i$.[11] An individual in period n at iteration t with strategy c uses $c(H_{n,t}, q_{n,t})$ in game $q_{n,t}$, $c(H_{n,t+1}, q_{n,t+1})$ in $q_{n,t+1}$, and so on.

As part of the specification of the map c, we assume that all strategies choose a strictly dominant action in any subgame they initiate, whenever such an action is available. For example, the player at the last stage of the game always chooses the outcome that she strictly prefers. This assumption is in the spirit of focussing upon the implications of other players' payoffs rather than the implications of one's own payoffs. Indeed, if players are to learn other players' preferences from observing their choices, other players cannot be completely free to act contrary to their own preferences.

More importantly, in the present model, using any such dominant choice could be made a result rather than an assumption. The key part of this assumption is sequential rationality, since such a dominant choice is optimal conditional upon having reached the node in question.

[10] If $n > 1$ but $t = 1$, then $H_{n,t} = \left\{Z_n, (q_{1,1}, \pi_{1,1}), \ldots, (q_{n-1,\kappa(n-1)}, \pi_{n-1,\kappa(n-1)})\right\}$. If $n = t = 1$, then $H_{n,t} = \emptyset$.

[11] It will not be required that *ToP* players remember the entire history. All that is needed is that they make and retain the exact inferences about other roles' binary preferences that are possible from observing the aggregate choices made in each period. It is not important whether naive players remember the entire history or not, in familiar subgames.

It is the large population in each role that is crucial in this connection. With only a single player in each role, the player in role $i < I$ might well prefer to not choose such a dominant action in order to misrepresent her preferences to some player $j > i$, so inducing j to choose in a way that is beneficial to i. However, when there is a large number of players in every role, who are randomly matched in each iteration of the game, each role i player has no effect on the distribution of i's choices that is observed by any role $j > i$ and thus no effect on j's future behavior. In these circumstances, not only is the best choice by each i myopic, in the sense of neglecting the future, but it is also sequentially rational.

Strategies that failed to use such dominant choices would eventually be pushed to an arbitrarily low level. Once this was so, we would approximate the current model. There is no reason then to be suspicious of the current assumption, but the approximation would make the proofs more complicated, so we do not pursue this option.

All strategies then satisfy—

Assumption 2.4: *Consider any i player role, and any i player subgame q. The action a at q is* dominant *for i if for every action $a' \neq a$, for every outcome z available in the continuation game after i's choice of a in q, and every outcome z' available in the continuation game after i's choice of a' in q, $z_i > z'_i$. For each $i = 1, \ldots, I$, every strategy always chooses any such dominant action.*

Naive Players

We adopt a definition of naivete that binds only if the subgame is new. This serves to make the ultimate results stronger, since the naive players can be otherwise rather smart. When the subgame is new, and there is no dominant choice, naive players condition in an arbitrary fashion on their own payoffs, but act in ignorance of other players' preferences.

Definition 2.2: *All naive strategies satisfy Assumption 2.4 in all subgames. There is a finite number of naive strategies that map their own observed payoffs to an arbitrary pure choice, whenever any of the subgames faced has never arisen previously, and a dominant strategy is lacking.*

If any subgame faced is *not* new, and there is no dominant strategy, there is no constraint imposed on any naive strategy. Although it makes an implausible combination, the naive players could then be fully Bayesian rational with respect to all of the relevant characteristics of the game—updating the distribution of opponents' payoffs, for example.

The following example illuminates the strengths and weaknesses of naive strategies, describing the opportunity that exists for more sophisticated strategies—

Example 2.1: *Consider Figure 2.1. In view of Assumption 2.4, the $P1$'s always make the SPE choice. The problem for the $P2$'s is to make the appropriate choice for each of the games they face, but where the outcome for each choice depends on the unknown preferences of the $P1$'s.*

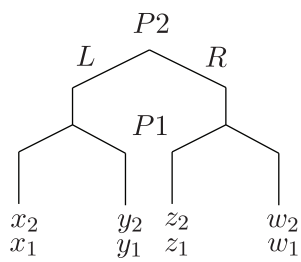

Figure 2.2: Example $I = A = 2$.

The key consideration in the long-run concerns how the various strategies perform when payoffs are chosen independently according to the cdf F. Suppose, for simplicity, that the cdf F represents independent choice of the two payoffs from the uniform distribution on [1,2].

A salient naive strategy for $P2$ is to choose L, for example, if and only if the 50-50 average of the own payoffs after choosing L exceeds the 50-50 average of the own payoffs after choosing R, in any novel game. That is, choose L if and only if $x_2 + y_2 > z_2 + w_2$. If either choice is dominant, this simple rule makes that dominant choice. Moreover, given risk neutrality in the payoffs, this naive strategy is the Bayesian rational procedure initially when there is no additional information about $P1$'s preferences, since each of $P1$'s choices are then equally likely given either choice for $P2$.

Whenever there is not a dominant choice for $P2$, however, any *naive strategy must make the wrong choice with strictly positive probability, under any F with* *full support. To show this for this F, it is enough to note the following. There is a clearly a positive probability that neither L nor R is dominant. Further, when there is no dominance, one of the following typical patterns of $P2$'s payoffs must arise i) $x_2 > z_2 > w_2 > y_2$ or ii) $x_2 > z_2 > y_2 > w_2$.*[12] *In case i), L is optimal if and only if $x_1 > y_1$, which has probability $1/2$. In case ii), L is optimal if and only if $x_1 > y_1$* or *both $y_1 > x_1$* and *$w_1 > z_1$. This has overall probability $3/4$. That is,* any *naive rule makes the wrong choice in a nontrivial subset of novel games, given the actual pattern of $P1$'s payoffs.*

Furthermore, we subsequently show that, whenever $\alpha < 3$, the naive $P2$s see only *novel games, in the limit.*[13] *Hence any naive strategy makes a suboptimal choice in a positive fraction of games, in the long run. This creates an opportunity for sophisticated strategies for $P2$ that observe the choices made by the $P1$'s and thereby build up a picture of $P1$'s preferences.*

Sophisticated Players

There are two aspects to the *ToP* strategies. The first of these, given as part i) of Definition 2.3 below, concerns the utilization of the knowledge of others' preferences. Picking the *SPE* choice at each node when the preferences of subsequent players are known characterizes the *SPE-ToP* strategy that will eventually dominate the population in every role. The second aspect, given as ii) of Definition 2.3, concerns how such knowledge of the preferences of others could be acquired from observing their behavior.

Definition 2.3: *All ToP strategies always satisfy Assumption 2.4 in all subgames. It is convenient to describe the remaining requirements on the ToP strategies in the reverse order to the temporal order in which they apply. i) If a ToP player in role i knows the ordinal preferences of all subsequent players over the set Z_n, each such ToP player maps the array of own preferences plus those of subsequent players to a pure action at each decision node (still subject to Assumption 2.4).*

[12] That is, it is without loss of generality to assume that the highest payoff is x_2. If there is no dominance, the next highest cannot be y_2 and so can be taken as z_2 without loss of generality. The only issue is how the two outcomes left rank relative to one another.

[13] To see this, observe the following. Assumption 2.3 implies that the total number of iterations in any period n history is bounded above by $n \cdot (N+n)^{\alpha} < (N+n)^{\alpha+1}$ where N is the initial number of outcomes. Since only one game is played at each iteration, this provides also an upper bound on the number of distinct games occurring along any such history. Further, in period n, there are $|Z_n|^4 = (N+n)^4$ possible games. It follows that if $\alpha + 1 < 4$, then the fraction of games that are familiar tends to zero, surely. This result is stated as Lemma 2.12 in Section 2.6.

A particular ToP strategy, the SPE-ToP strategy, maps all of these preferences to the SPE choice at each node, if this SPE choice is unique. Other ToP strategies make a non-SPE choice in at least one subgame defined by the ordinal preferences of others and of the player in question.[14] *ii) It is common knowledge among all ToP players that there exists a positive fraction of SPE-ToP players in every role.*

What is meant in Definition 2.3 i) by hypothesizing that the *ToP* strategies "know" the preferences of subsequent players? That is, what patterns of play reveal these preferences under Definition 2.3 ii)? We use Example 2.1 to illustrate these issues and then indicate how the argument can be generalized by considering the case where $I = 3$.

Example 1 Revisited: *In this example with $I = A = 2$, all of the ToP $P2$'s learn one of $P1$'s binary preferences, whenever the $P1$ are forced to make a choice between two outcomes that has not arisen before. This follows since Assumption 2.4 implies that the $P1$'s always make the SPE choice. Indeed, whenever $\alpha > 1$, so that the rate of introduction of novelty is not too fast, such learning by the ToP $P2$'s will be essentially complete in the limit.*[15] *If $\alpha \in (1, 3)$, the ToP strategies then have a clear knowledge edge over the naive strategies.*

Each ToP strategy maps the preferences of $P1$, once these are known, as well as own preferences, to an action. The SPE-ToP strategy fully exploits the knowledge edge of the ToP strategies over the naive strategies, by mapping these two preference profiles to the SPE choice. It is obvious in Example 1 that this SPE-ToP strategy will then eventually outdo all other strategies, naive or sophisticated, since only the SPE-ToP comes to correctly anticipate all of the choices of the $P1$'s.

It will not matter that in the interim—when the sophisticated strategies do not know $P1$ preferences—that they make inappropriate choices, as these instances

[14] This requirement is merely to avoid triviality. It has the following implication. Since the preferences involved are ordinal, the probability of such a subgame is positive under F. Indeed, the probability of a game that repeats this subgame for every decision node of the role in question is also positive. Such games will then give the *SPE-ToP* strategy a strict advantage over any other *ToP* strategy.

[15] This is a key theoretical result of the paper, given in the Section 2.6 as Lemma 2.7. Although the proof there is complicated by the need to allow more than two stages, it is nontrivial even for $I = 2$. It is not hard to see, however, that $\alpha > 1$ means that complete learning is not ruled out, as follows. The introduction of the n-th novel outcome results in $N + n - 1$ new pairwise choices. There are $\kappa(n) > (N + n - 1)^\alpha$ iterations of the game before the introduction of further novelty. Therefore, whenever $\alpha > 1$ the number of iterations between outcomes outstrips the number of new pairwise choices introduced. As n grows to infinity, this shows it is at least *possible* that the $P2$ *ToP*s will see nearly all of the $P1$ choices before the next outcome arrives. The more difficult task is to show that this possibility is realized, for the *ToP* strategies given in Definition 2.3.

occur with vanishing probability. Similarly, neither will it matter how sophisticated the naive strategies are on familiar games, since these also arise with vanishing probability.

We now further illustrate this mechanism, by extending the argument in Example 2.1 to the case $I = 3$. Learning about $i = 1$'s preferences remains straightforward, whenever $\alpha > 1$, and proceeds as before. Indeed, role 1's preferences then become common knowledge among all *ToP* players in role 2 and 3. The new interesting case then concerns how the *ToP* players in role 3 can learn role 2's preferences. Suppose then a game is drawn in which some subgame has a dominant choice a, say, for role 2, as Assumption 2.4, and this subgame is reached.[16] It follows from Assumption 2.4 that all players in role 2 take this dominant action. The *ToP* players in role 3 do not know that such a dominant choice exists for 2. They do know, however, that the 2's also know 1's preferences. Hence, whether such a dominant choice exists or not, the *SPE-ToP*'s in role 2 have unequivocally demonstrated to the *ToP* in role 3 that they prefer the outcome induced by a to any outcome they might have induced instead. Still under the assumption that $\alpha > 1$, the *ToP* players in role 3 can then build up a complete picture of the preferences of role 2.[17]

The common knowledge assumption for the *ToP* players, as described in Definition 2.3 ii), can be stripped to its bare revealed preference essentials. It is unimportant, that is, what or whether the *ToP* players think, in any literal sense. All that matters, in the case that $I = 3$, for example, is that it is *as if* the *ToP*s in roles 3 add to their knowledge of role 2's preferences as described above. Once a *ToP* player in role 3 has seen histories in which all of role 2's binary choices have been put to the test like this, given that this is already true for role 1, the role 3 *ToP* players effectively know all that is relevant about the ordinal preferences of subsequent players and can act on this basis. This is essentially purely a mechanical property of the map, c, used by the *ToP* players. That is, not merely can the naive players be "zombies", in the philosophical sense, but so too can the *ToP* players.[18]

[16]That this subgame is reached could be forced by assuming that this subgame arises for all of 3's choices.

[17]Lemma 2.7 in Section 2.6 proves the key result that $\alpha > 1$ ensures complete learning.

[18]That is, the revealed preference approach adopted here is agnostic about internal mental processes.

2.2.3 Evolutionary Adaptation

The population structure and associated payoffs are as follows—

Definition 2.4: *The total population of all strategies is normalized to 1 for every role i. The sophisticated (ToP) strategies are labeled as $r = 1, \ldots, R$, for $R \geq 1$, say where $r = 1$ is the SPE-ToP strategy. The naive strategies are labeled as $r = R + 1, \ldots, \bar{R}$, where $\bar{R} > R$. The fraction of the total population in role $i = 1, \ldots, I$ that uses strategy $r = 1, \ldots, \bar{R}$ in period $n = 1, 2, \ldots$ and iteration $t = 1, \ldots, \kappa(n)$ is then denoted $f^i_{n,t}(r)$, where $f^i_{n,t} = (f^i_{n,t}(1), \ldots, f^i_{n,t}(\bar{R}))$. The average material payoff obtained by such a strategy r in role i in period n and iteration t is then denoted $\bar{z}^i_{n,t}(r)$, where $\bar{z}^i_{n,t} = (\bar{z}^i_{n,t}(1), \ldots, \bar{z}^i_{n,t}(\bar{R}))$.*

The population evolves in a standard adaptive fashion between each iteration of the game. This has the property, in particular, that the fraction of individuals who use a strategy that is best increases, given only that there is some suboptimal strategy—

Assumption 2.5: *Consider role $i = 1, \ldots, I$ in period $n = 1, 2, \ldots$ and at iteration $t = 1, \ldots, \kappa(n)$. If the population structure is $f^i_{n,t}$ with average payoffs $\bar{z}^i_{n,t}$, the population structure in the next iteration is given by $f^i_{n,t+1} = \Psi(f^i_{n,t}, \bar{z}^i_{n,t})$.*[19] *This function $\Psi : \Delta^{\bar{R}-1} \times [m, M]^{\bar{R}} \to \Delta^{\bar{R}-1}$, where $\Delta^{\bar{R}-1}$ is the unit simplex in $\mathbb{R}^{\bar{R}}$, has the properties i) Ψ is continuous, ii) $\Psi_r(f^i_{n,t}, \bar{z}^i_{n,t})/f^i_{n,t}(r) > \eta$ for some $\eta > 0$, and for $r = 1, \ldots, \bar{R}$, iii) if $\bar{z}^i_{n,t}(r^*) = \max_{r=1,\ldots,\bar{R}} \bar{z}^i_{n,t}(r) > \bar{z}^i_{n,t}(r')$, for some $r' \in \{1, \ldots, \bar{R}\}$, then $\Psi_{r^*}(f^i_{n,t}, \bar{z}^i_{n,t}) > f^i_{n,t}(r^*)$ and iv) if $\bar{z}^i_{n,t}(r) = \bar{z}^i_{n,t}(r')$, for all $r, r' \in \{1, \ldots, \bar{R}\}$, then $\Psi(f^i_{n,t}, \bar{z}^i_{n,t}) = f^i_{n,t}$.*[20]

Recall that Figure 2.1 gives a schematic representation of the model.

2.2.4 The Main Result

The main result is that, in the limit, the *SPE-ToP* strategy fully learns the preferences of others, applies this knowledge to choose the optimal action, and dominates the population.

Theorem 2.1: *Suppose Assumptions 2.1-2.5 all hold. Suppose that there are a finite number of ToP strategies, including SPE-ToP in particular, as in Definition 2.3, and a finite number of naive strategies, as in Definition 2.2. If $\alpha \in (1, A^2 - 1)$,*

[20] Ψ_r denotes the rth component of the vector Ψ, $r = 1, \ldots, \bar{R}$.

then the proportion of SPE-ToP players in role i, $f^i_{n,t}(1)$, tends to 1 in probability, as $n \to \infty$, for all $t = 1, \ldots, \kappa(n)$, and for all $i = 2, \ldots, I$. The observed pattern of play in each realized game converges to an SPE, in probability.

The proof of this is given Section 2.6.

The following specific remarks apply—

i) The result for $i = 1$ holds trivially by Assumption 2.4.

ii) The proof shows that all of the *ToP* strategies learn others' preferences essentially always if $\alpha > 1$, but all naive strategies see essentially only new subgames if $\alpha < A^2 - 1$, in the long run. If both inequalities hold, as above, there is an opportunity for the *ToP* strategies to outdo the naive strategies, one that the *SPE-ToP* fully exploits.

iii) We focus here on the case that $\alpha \in (1, A^2 - 1)$. These bounds are tight in the sense that, if $\alpha < 1$ then it is mechanically impossible for the *ToP* players to learn the preferences of opponents from their binary choices.[21] On the other hand, if $\alpha > A^2 - 1$, then naive players in role 2 see only familiar subgames in the limit.[22]

iv) If the role i is earlier in the game, so i is larger, it is harder for naive strategies to learn all the subgames they initiate. In the role i, that is, the cutoff value for a naive strategy is $\alpha = A^i - 1$, below which learning is impossible in the long run, and this increases with i. However, the task faced by the *ToP* strategies does not become more complex in the same way, in that the cutoff value of $\alpha = 1$ is unaffected by the role i involved.

v) If it were assumed that naive players need to have experienced the *entire* game, and not just a subgame they initiate, before they can learn it, the upper bound for α would be $A^I - 1$, uniformly in $i = 2, \ldots, I$.

vi) If $\alpha < 1$, so that all the *ToP* players are overwhelmed with novelty, as are the naive players, the outcome of the evolutionary contest hinges on the default behavior of the naive and *ToP* strategies when these face their respective novel circumstances. As long as the naive players are not given a more sophisticated default strategy than the *SPE-ToP*s players, the naive players will, at best, match the *SPE-ToP*s players.

[21] The proof is analogous to that of Lemma 2.12 in the Section 2.6 which establishes the corresponding property for naive players.

[22] The proof is analogous to that of Lemma 2.7 in Section 2.6 which establishes the corresponding property for the *ToP* players.

vii) If $\alpha > A^2 - 1$, naive players in at least role 2 have seen most subgames previously, in the long run. The relative performance of the *SPE-ToP* and the naive players then depends on the detailed long run behavior of the naive players. If the naive players play a Bayesian rational strategy the second time they encounter a given subgame, they might tie the *SPE-ToP* players. It is, in any case, not intuitively surprising that a clear advantage to *SPE-ToP*s relies upon there being at least a minimum rate of introduction of novelty.

The eventual predominance of the *SPE-ToP*s over all the naive strategies resolves the issue raised by Stahl (1993) in this context. Consider any particular naive strategy that maps own payoffs to an action, where this choice cannot, of course, condition on the future realization of the sequence of games. If there is a dominant strategy in any subgame, this naive strategy chooses that by assumption. Otherwise, although there may be a set of subgames, with positive probability under F conditional on the observed own payoffs, in which the naive strategy makes the *SPE* choice, there must also be a set of subgames, also with positive conditional probability under F, for which this is not true. Since any particular naive strategy must therefore, with probability one, choose suboptimally in a positive fraction of games, in the limit, it is outdone, with probability one, by the *SPE-ToP* that is not preprogrammed but rather adapts to the outcomes and games that are drawn, and ultimately chooses optimally essentially always.[23]

That is—

COROLLARY 2.1: *Under the hypotheses of Theorem 2.1, any particular naive strategy will, with probability one, choose suboptimally in a positive fraction of new subgames in the limit.*

Further, *ToP* strategies could be extended to deal with occasional shifts in preferences over outcomes. Such a generalized model would be noisier than the current model, and therefore harder to analyze, but this potential flexibility of the *ToP* strategies would constitute a telling additional argument in their favor.

It follows, significantly, that the evolutionary dominance of the *SPE-ToP* is robust to the introduction of sufficiently small cost, completing the resolution of

[23]This argument has the following subtlety. Consider a particular *realized* sequence of games. With probability one, each observed own payoff is associated with a unique vector of payoffs for the other roles. It follows that, with probability one, there exists a naive strategy that maps own payoffs to an action that is the *SPE* choice in every such realized subgame. To choose this naive strategy in advance is to condition on the future, however, given that there are uncountably many possible naive strategies.

the issue raised by Stahl (1993). Suppose that all *ToP* strategies entail a per game cost of $\omega > 0$, to reflect the cognitive cost associated with deriving the preferences of others from observation. Then we have

COROLLARY 2.2: *Theorem 2.1 remains valid when all ToP strategies entail a per game cost ω (where the naive players have zero cost), if ω is small enough.*

If $\alpha > A^2 - 1$, however, then naive players in at least role 2 are usually familiar with the subgame they initiate, in the long run. The presence of a fixed cost might then tip the balance in favor of the naive players. If $\alpha < 1$, so all players, naive or sophisticated, are overwhelmed with novelty, this might also be true, when the default play of the naive and sophisticated players is comparable.

The presence of such a per game cost, that is independent of the number of outcomes, is not unreasonable since the *ToP* strategies would require the maintenance of a brain capable of sophisticated analysis. However, the *memory* demands of the naive players here are likely to be greater than the memory demands of *ToP*. The naive players need to remember each game; the *ToP*s need only remember preferences over each pairwise choice for opponents, and if memory is costly then these costs would be lower for the *ToP*s whenever there are a large number of outcomes. In this sense, consideration of all costs might well reinforce the advantage of the *ToP* players over the naive players.

The attainment of an *SPE* in Theorem 2.1 relies on the assumption that there is a large population in each role, with random matching for each iteration of the game. Even though a non-*SPE* choice by all role i players might benefit all role i players since it could advantageously influence the choice of a role $j > i$, this benefit is analogous to a public good. The choice by just one role i player has no effect on j's information bearing on i's preferences. Thus, the optimal choice by any particular role i player is sequentially rational. (The large population in each role, together with random matching, also ensures choices are myopic, ignoring, that is, future iterations of the game.) This argument that an *SPE* is attained once the preferences of others are known is analogous to Hart (2002).[24]

We close this subsection with several additional general remarks.

1) The key issue here is how *ToP*s deal with *novelty*—the arrival of new outcomes—rather than with *complexity*—the unbounded growth of the outcome set. Indeed,

[24] Hart considers a finite population in each role, with mutation ensuring all subgames are reached. His result is that the *SPE* is attained for a large enough common population size and small enough mutation rate.

the model could be recast to display the role of novelty as follows. Suppose that a randomly chosen outcome is dropped whenever a new outcome is added, at each date n, so the size of the outcome set is fixed, despite such updating events. There will then be a critical value such that, if the number of games played between successive dates is less than this critical value, the naive players will be mechanically unable to keep up with the flow of new games. There will also be an analogous but lower critical value for the *ToP*s. If the fixed interval between updating events is chosen to lie between these two critical values, the naive players will usually be faced with novel subgames; the *ToP*s will face a stochastic but usually positive fraction of subgames in which the preferences of subsequent player roles are known. This provides a version of the current results, although one that is noisier and therefore more awkward than the current approach.[25]

2) Suppose, hypothetically, that the naive types have all been eliminated. The eventual ascendancy of each *SPE-ToP* type over the other *ToP* types is not a matter of strategic dominance but relies on the previous ascendancy of *SPE-ToP* types at all subsequent stages. That is, given a particular pattern of subsequent *ToP* roles, there may be a *ToP* that outdoes the *SPE-ToP*. It is only once *SPE* behavior has been established for subsequent players, by backwards induction, that the *SPE* choices become optimal.[26]

3) The sophisticated players here do not use the transitivity of others' preferences. If they were to do so, this could only extend the range of α over which complete learning of opponents' preferences would arise, and therefore the range over which the sophisticated strategies would outcompete the naive strategies.[27]

4) Consideration of a long run equilibrium, as in Theorem 2.1, is simpler analytically than direct consideration of the speed of out-of-equilibrium learning of the various strategies. More importantly, it also permits the use of minimal restrictions on the naive and *ToP* strategies, as is desirable in this evolutionary context.

[25] The need in the current model for the number of games played between updating events to grow with time is a reflection of the fact that each new outcome produces a larger number of novel games when there is already a larger number of outcomes.

[26] This is perhaps analogous to the difficulty that the Connecticut Yankee has at King Arthur's Court, according to Mark Twain. That is, to his consternation, the choice made by Twain's hero often fails to be optimal because the choice by his opponents is non-optimal.

[27] Although they do not apply directly, the results of Kalai (2003) concerning PAC-learning and P-dimension, Theorem 2.1 and Theorem 3.1, in particular, suggest that the use of transitivity might lower the critical value of α as far as 0.

5) Our results show how an increase in the rate of introduction of novelty might precipitate a transition from a regime in which there is no advantage to strategic sophistication to one in which a clear advantage is evident. This is consistent with theory and evidence from other disciplines concerning the evolution of intelligence. For example, it is argued that the increase in human intelligence was in part due to the increasing novelty of the savannah environment into which we were thrust after we exited our previous arboreal niche. (For a discussion of the intense demands of a terrestrial hunter-gatherer lifestyle, see, for example, Robson and Kaplan, 2003.)

2.2.5 Related Literature

We outline here a few related theoretical papers in economics. The most abstract and general perspective on strategic sophistication involves a hierarchy of preferences, beliefs about others' preferences, beliefs about others' beliefs about beliefs about preferences, and so on. (Robalino and Robson, 2012, provide a summary of this approach.) Harsanyi (1967/68) provides the classic solution that short circuits the full generality of the hierarchical description.

A strand of literature is concerned to model individuals' beliefs in a more realistic fashion than does the general abstract approach. An early paper in this strand is Stahl (1993) who considers a hierarchy of more and more sophisticated strategies analogous to iterated rationalizability. A smart$_n$ player understands that no smart$_{n-1}$ player would use a strategy that is not $(n-1)$-level rationalizable. A key aim of Stahl is to examine the evolution of intelligence in this framework. He obtains negative results—the smart$_0$ players who are right in their choice of strategy cannot be driven out by smarter players in a wide variety of plausible circumstances. Our positive results, in Corollary 2.2, in particular, stand in sharp contrast to these previous results.

Mohlin (2012) provides a recent substantial generalization of the closely related level-k approach that allows for multiple games, learning, and partial observability of type. Nevertheless, it remains true that lower types coexist with higher types in the long-run. This is not to deny that the level-k approach might work well in fitting observations. For example, Crawford and Iriberri (2007) provide an explanation for anomalies in private-value auctions based on this approach.

There is by now a fairly large literature that examines varieties of, and alternatives to, adaptive learning. Camerer, Ho and Chong (2002), for example, extend a model of adaptive, experience-weighted learning (EWA) to allow for best-responding to predictions of others' behavior, and even for farsighted behavior that involves teaching other players. They show this generalized model outperforms the basic EWA model empirically. Bhatt and Camerer (2005) find neural correlates of choices, beliefs, and 2nd-order beliefs (what you think that others think that you will do). These correlates are suggestive of the need to transcend simple adaptive learning. Finally, Knoepfle, Camerer and Wang (2009) apply eye-tracking technology to infer what individuals pay attention to before choosing. Since individuals actually examine others' payoffs carefully, this too casts doubt on any simple model of adaptive learning.

2.3 Experiments on Theory of Preferences

2.3.1 Experimental Design

We report here the results of experiments that are simplified versions of the theoretical model. These test the ability of individuals to learn the preferences of others through repeated interaction and to use that information strategically to their advantage. The game tree is a two-stage extensive form where each player has two choices at each decision node.

There are then two player roles, 1 and 2.[28] Player roles differ in their position in the game tree and their (induced) preferences, but all players of a given role have identical preferences. In each period, each role 1 participant is randomly and anonymously matched with a single role 2 participant to play a two-stage extensive form game, as depicted in Figure A.21, in Appendix A.2. We employ this matching scheme to at least diminish the likelihood of supergame effects. In each game, role 1 players always move first, choosing one of two intermediate nodes (displayed in the figure as blue circles), and then based on that decision, the role 2 player chooses a terminal node that determines payoffs for each participant (displayed in the figure as a pair of boxes).

When making their decisions, participants observe only their own payoff at each outcome and are originally uninformed of the payoff for the other participant.[29] Instead, they know only that payoff *pairs* are consistent over time. That is, whenever

[28]Here we revert to the usual convention that role 1 moves before role 2.

[29]Note that payoff privacy has the added benefit of mitigating the effects of non-standard preferences on individual choice; since individuals are unaware of exactly how their choices impact others' payoffs, altruistic and reciprocal actions, which may depend on the relative effect on own and other's payoffs (as in Charness and Rabin, 2002, for example), will be controlled. Indeed, it

the payoff for role 1 is X, the payoff to role 2 will always be the same number Y. In Figure A.21, which is shown from the perspective of a role 1 participant, his own payoff at each terminal node is shown in the orange box, while his counterpart's payoff is displayed as a "?" in the blue box. Similarly, when role 2 players make their decisions, they only observe their own payoffs and see a "?" for their counterpart (see figure A.22).

In each period, the payoffs at each terminal node are drawn *without replacement* randomly from a finite set of V payoff pairs.[30] Each element in each pair of payoffs is unique, guaranteeing a strict preference ordering over outcomes. This set is fixed in the experiments in contrast to its growth in the theoretical model. We do not then attempt to study the theoretical long run in the experiments, but content ourselves with observing the rate of learning of opponents' preferences. Allowing for the strategic equivalence of games in which the two payoff pairs at a given terminal node are presented in reverse order, there are $\binom{V}{2}\binom{V-2}{2}/2$ strategically distinct games that can be generated from V payoff pairs, each of which has a unique subgame perfect equilibrium.

Thus, as in the theoretical model, despite their initial ignorance of their counterpart's preferences, role 1 players can learn about these preferences over time, by observing how role 2 players respond to various choices presented to them. If role 1 players correctly learn role 2 players' preferences, they can increase their own payoff by choosing the *SPE* action. On the face of it, role 1 players will have then developed a theory of a role 2 player's mind.

This suggests investigating whether role 1 players choose in a manner that is increasingly consistent with the *SPE*. Initial pilot sessions revealed two issues with this strategy: 1) many of the randomly generated games include dominant strategies for player 1, which are not informative for inferring capacity to learn the preferences of others, as indeed reflected in the theoretical model, and 2) more subtly, there is a simple "highest mean" rule of thumb that also often generates *SPE* play. Consider a player 1 who is initially uncertain about player 2's preferences. From the point of view of player 1, given independence of player 2's preferences, player 2 is equally likely to choose each terminal node, given player 1's choice. The expected payoff

has long been known that payoff privacy encourages the achievement of equilibrium outcomes in market settings (Smith 1982).

[30]We sample with replacement in the theoretical model, although this assumption is merely a minor convenience. We do not allow replacement here to make the most of our experimental resources of time and money.

maximizing strategy is to choose the intermediate node at which the average of potential terminal payoffs is highest. Indeed, our pilot sessions suggested that many participants followed this strategy, which was relatively successful.

For these reasons, we used a 3x1 within-subjects experimental design that, over the course of an experimental session, pares down the game set to exclude the games in which choice is too simple to be informative. Specifically, each session included games drawn from 7 payoff pairs (so there are 105 possible games). In eighteen of our sessions, payoff possibilities for each participant consisted of integers between 1 and 7, and in two sessions the set was $\{1,2,3,4,8,9,10\}$. This variation was intended to reduce noise by more strongly discouraging player 2 from choosing a dominated option, but observed player 2 choices in these sessions are comparable to those in other sessions, so we pool the data for analysis below. Each session lasted for 90 periods in which, in the first 15 periods, the game set included 15 randomly chosen games from the set of possible games, $\bar{Q}$, say. Finally, starting in the 16th period, we eliminate all games in which player 1 has a dominant strategy, and the next 15 periods consist of games randomly drawn from this subset of $\bar{Q}$. Finally, starting in the 31st period, we also eliminate all games in which the optimal strategy under the "highest mean" rule of thumb corresponds to the *SPE* of the game, and our final 60 periods consist of randomly drawn games from this smaller subset. Thus, our final 60 periods make it harder for player 1 to achieve high payoffs, since the only effective strategy is to learn the preferences of the role 2 players.

Learning by role 1 players here would be disrupted by the presence of any role 2 player who fails to choose his dominant action. For this reason, we considered automating the role 2 player. However, on reflection, this design choice seems untenable. In the instructions, we would need to explain that algorithmic players 2 maximize their payoffs in each stage, which would finesse much of the inference problem faced by player 1—in essence the instructions would be providing a key part of the theory of mind. It is also conceivable that individuals would behave differently towards a computer program than they would towards a human agent.

A second potential issue is that foregone payoffs (due to role 1 player's choice) may lead to non-myopic behavior by some player 2s. Such behavior involves role 2 players solving a difficult inference problem. A spiteful (or altruistic) player 2, who wanted to punish (or reward) player 1 on the basis of player 2's foregone payoffs, first must infer that player 1 has learned player 2's preferences and then infer player 1's own preferences on the basis of this assumption. Player 2 could then, given his options, choose the higher or lower of the two payoffs for player 1 as either punishment or reward. However, players 2 chose their dominant action roughly 90% of the time, which suggests that these sources of error were not a prominent feature of our experiment.

We relate our results directly to theory of mind in psychology, as measured by two short survey instruments. At the conclusion of the experiment, participants completed the Autism-Spectrum Quotient (AQ) survey designed by Baron-Cohen et al. (2001), since autism spectrum reflects varying degrees of inability to "read" others' minds. This short survey has been shown to correlate with clinical diagnoses of autism spectrum disorders, but it is not used for clinical purposes. The instrument was designed for use on adults of normal intelligence to identify the extent of autism spectrum behaviors in that population. Participants also completed the Broad Autism Phenotype Questionnaire (BAP) due to Hurley et al. (2007), which provides a similar measure of autism spectrum behavior and is highly correlated with the AQ. With this additional data we will be able to evaluate how each participant's ability to perform as player 1 in our experiments correlates with two other well-known *ToP* metrics.[31] A copy of the AQ questionnaire is available in Appendix A.3.[32]

We report data from 20 experimental sessions with a total of 174 participants (87 in each role). Each experimental session consisted of 6, 8 or 10 participants, recruited from the students of Simon Fraser University between April and October 2013. Participants entered the lab and were seated at visually isolated computer terminals where they privately read self-paced instructions. A researcher was available to privately answer any questions about the instructions. After reading the instructions, if there were no additional questions, the experiment began. Instructions are available in Appendix A.1.

[31] One might be concerned that any differences we observe in behavior that are correlated with AQ are actually driven by differences in intelligence. Indeed, it is well-known that extreme autistics tend to have low IQs. Crucially, however, within the normal range of AQ scores (those surveyed who had not been diagnosed with an autism spectrum disorder), the survey measure is uncorrelated with intelligence (Baron-Cohen et al., 2001). Our sample consists of undergraduates none of whom (to our knowledge) are diagnosed with any autism spectrum disorder. Thus any relationship we observe between AQ and performance is unlikely to be due to differences in intelligence.

[32] In the first wave of these experiments performed in April and June 2013 (76 subjects total), we conducted the AQ questionnaire with a 5-point Likert scale that allowed for indifference rather than the standard 4-point scale which requires participants to either agree or disagree with each statement. The AQ questionnaire is scored by assigning 1 or 0 to each response and summing. In our data analysis below, we assign indifferent responses a score of 0.5.

Each experimental session took between 90 and 120 minutes. At the conclusion of each session, participants were paid privately in cash equal to their payoffs from two randomly chosen periods. We use this protocol to increase the salience of each individual decision, thereby inducing participants to treat each game as payoff-relevant. For each chosen period, the payoff from that period was multiplied by 2 or 3 (depending on the session) and converted to CAD. Average salient experimental earnings were $25.00, with a maximum of $42.00 and a minimum of $6.00. In addition to their earnings from the two randomly chosen periods, participants also received $7 for arriving to the experiment on time. Upon receiving payment, participants were dismissed.

2.3.2 Experimental Results

Since the decision problem is trivial for player 2, our analysis focuses entirely on decisions by player 1. We focus on the probability with which player 1 chooses an action consistent with the *SPE* of the game. For a fixed game, and with repeated play with fixed matching and private information about individual payoffs, pairs frequently converge to non-cooperative equilibrium outcomes over time (McCabe et al., 1998).[33] This is not surprising since, in their environment an individual merely need learn her counterpart's preferences over two pairwise comparisons. However, since these experiments employ static repetition of the same game, the data do not clearly distinguish theory of mind from reinforcement learning. Our experiment is the first (that we know of) to test theory of mind capacity in a dynamic setting in which inferences drawn from the play of one game may be employed to predict play in novel, future games. In this sense our setting is more strategically complex than those previously studied, and hence we are able to both distinguish *ToP* from reinforcement learning *and* observe heterogeneity in *ToP* capabilities, which we can exploit in our data analysis.

First, we describe overall learning trends, and we show that individuals' performance as players 1 depends on how much information they have acquired about the preferences of players 2. This suggests that our players 1 exhibit *ToP* in the sense of the theoretical model. Finally, we compare our measure of *ToP* to measures from psychology and show that the learning speed of players 1 is significantly correlated

[33] See also Fouraker and Siegel (1963) and Oechssler and Schipper (2003).

with survey responses, suggesting that our theoretical concept of *ToP* corresponds, at least to some extent, with theory of mind as understood by psychologists.

Learning Others' Preferences

Figure 2.3 displays a time series of the probability that player 1 chose an action consistent with knowledge of player 2's preferences (i.e. consistent with *SPE*) over the 90 periods of the experiment. After 15 periods, the game set no longer included instances where player 1 had a dominant strategy. After 30 periods, the game set no longer included instances where player 1 would choose correctly by following the "highest mean" rule of thumb. At period 31, when subjects enter the NoDominant/NoHeuristic treatment, there is a significant downtick in player 1's performance, but afterwards there is a notable upward trend in the probability of player 1 choosing optimally.Despite the fact that individuals tend to learn player 2's preferences over time on average, we observe substantial heterogeneity in rates of learning, which we exploit in the next section.

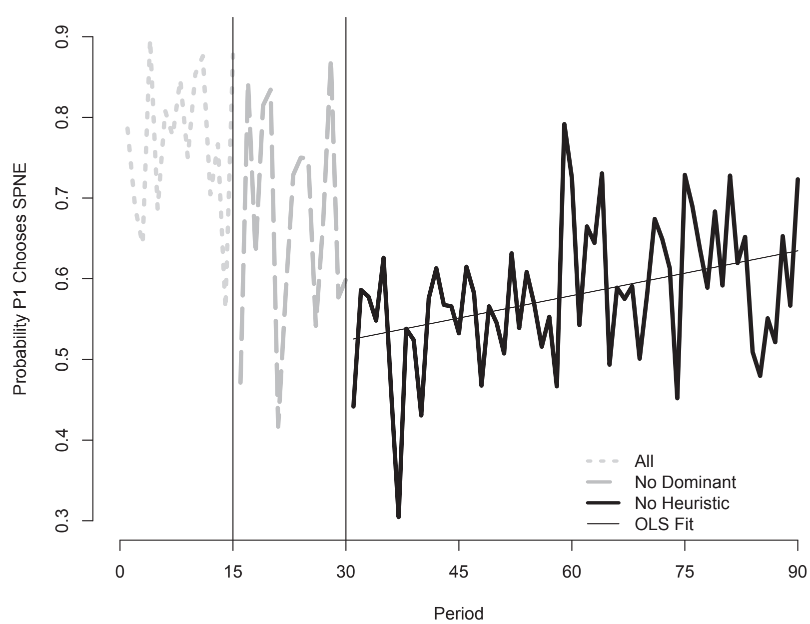

Figure 2.3: Time Series of Learning Opponent's Preferences.

To provide statistical support for these observations, Table 2.1 reports logistic regressions where the dependent variable takes a value of 1 if player 1 chooses an action consistent with the *SPE* of the game and 0 otherwise. We include treatment dummies for periods 1-15 and periods 16-30 to control for the game set. To identify the impact of feedback quality from player 2 choices on the likelihood of *SPE* choices, column (2) also includes two variables that control for the proportion of dominant choices made by players 2 in previous periods. Specifically, let $W_{i,t}$ be an indicator variable that takes a value of 1 when player i's partner chose the dominant action in the randomly chosen game q_t. Then we compute the lagged proportion of dominant choices observed by player i as $\frac{\sum_{s=1}^{t-1} W_{is}}{t-1}$. Observing dominant choices by *all* players 2 is also informative, so we compute a second measure for each period of each session that measures the lagged proportion of dominant choices made by all players 2. To test for naive reinforcement learning, as in the theoretical model, column (3) also includes a variable that counts the number of times participants have played the randomly chosen game at time t in the past. Finally to test for *ToP* learning, as in the model, column (4) includes two additional variables that measure the amount of information player 1 has at a given time about player 2's preferences. Specifically, let q_t be a feasible subgame in period t and $I(q_t)$ be an indicator function that takes a value of 1 if any player 2 is observed making a choice in that subgame in period t (or in the mirror image subgame).[34] In a given period, there are two feasible subgames q_t^1 and q_t^2, say. We then measure the previous exposure to player 2's preferences in game q_t by computing: $\min\{\sum_{s=1}^{t-1} I(q_s^1), \sum_{s=1}^{t-1} I(q_s^2)\}$. This provides a rough measure of what player 1's should know about player 2's preferences. It is a function of the total number of times that player 2 has chosen between each of the two relevant outcome pairs. These two totals are then aggregated using the function min for simplicity. As with the variables we introduced in column (2), we also construct an analogous measure that includes only those choices made by the person with whom player 1 was paired. We also include both session and individual fixed effects.

The positive and significant estimated coefficient on Period in column (1) indicates that participants are increasingly likely to choose optimally over time. This is consistent with the evidence in Figure 2.3. In column (2), when we include two variables measuring the fraction of previous dominant choices by players 2, we find

[34]Recall that players 1 receive aggregated information about the choices of all players 1 and 2 in their session at the end of each period. See Figure A.23 in the appendix.

P1 Chose SPE	(1)	(2)	(3)	(4)
Period	0.009***	0.009***	0.009***	0.002
	(0.002)	(0.002)	(0.002)	(0.003)
No Dominant Options	0.811***	0.808***	0.805***	0.842***
	(0.096)	(0.096)	(0.098)	(0.098)
All Treatments	1.415***	1.430***	1.425***	1.444***
	(0.119)	(0.120)	(0.124)	(0.124)
Cumulative Fraction My Partner Chose Dominant$_{t-1}$		1.226***	1.226***	1.237***
		(0.440)	(0.440)	(0.441)
Cumulative Fraction All P2s Chose Dominant$_{t-1}$		-0.265	-0.264	-0.187
		(0.819)	(0.819)	(0.822)
# of Times Played Previously			0.009	0.013
			(0.051)	(0.051)
# of Previous Choices Observed My Partner				-0.010
				(0.022)
# of Previous Choices Observed All P2s				0.080***
				(0.019)
Constant	-0.497**	-1.293*	-1.286*	-1.329**
	(0.241)	(0.660)	(0.661)	(0.663)
Observations	7830	7743	7743	7743
Session Fixed Effects	Y	Y	Y	Y
Individual Fixed Effects	Y	Y	Y	Y

Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.1: Logistic Regression Analysis of Learning.

a significant effect only of dominant choices made by partnered player 2s, but not by all players 2. Player 1s who have observed a greater share of dominant actions by their partners are more likely to choose optimally in later periods.[35] Column (3) tests for naive reinforcement learning as in the theoretical model. We find no evidence of a significant effect of repetition of the same game on the probability of choosing correctly. This is driven in part by the fact that our games were drawn from a relatively large set, which reduces the potential for repetition. Finally, column (4) includes our measures of the amount of information players 1 had about

[35] If player 1s were Bayesian rational, they would treat the observations on all player 2's as equally informative. It is plausible psychologically, however, that they pay particular attention to their partnered player 2, especially since this partner's behavior affects the current payoff for player 1. Indeed, this finding is consistent with evidence that individuals overweight private information; see, for example, Goeree et al., (2007).

the preferences of player 2. A highly significant and positive coefficient on the variable measuring the information that could be gleaned from all previous choices by other players 2 implies that players 1 improve their performance by applying what they have learned about the preferences of players 2 in the past. That is, they exhibit *ToP* in the sense of our model. In contrast to our findings from column (2), we find that the total previous number of choices made by *all* players 2 is a better determinant of learning than those made by their partner, as would be Bayesian rational. Importantly, when we include these variables, the coefficient on Period is no longer statistically significant, suggesting that the significant estimated trend in columns (1) - (3) was actually capturing the effects of *ToP*. Thus, even in this complex setting, individuals are able to learn the preferences of others. We summarize these observations below:

Finding 1: On average, there is a significant increase in understanding of others' preferences over time, despite individual variation.

Finding 2: The increase is driven by observation of player 2's preferences (*ToP*) rather than naive reinforcement learning, as is consistent with the theoretical results.

Comparing Measures of ToP

Table 2.1 provides evidence that increases in the rate of *SPE* choices result from *ToP*. However, our data reveal clear heterogeneity across individuals. Thus, we exploit this heterogeneity to ask whether our measure of *ToP* correlates with previous survey measures of theory of mind from psychology. Specifically, we examine correlations between subjects' AQ and BAP scores and the rate at which players *learn* the preferences of others.

We estimate learning rates separately for each player 1 with logistic regressions where the dependent variable takes a value of 1 when the player chose a node consistent with *SPE* and 0 otherwise, and the independent variables are our measure of the information available about player 2's preferences from previous choices (as described above), the lagged proportion of dominant choices made by their partners, and a constant term. The β coefficient on the first independent variable provides an estimate of each individual's rate of learning.[36] In both computations, we restrict attention only to choices that are informative for inferences about *ToP* by excluding games with dominant strategies for players 1 and games in which the "highest mean" heuristic corresponds to the *SPE*.[37]

We then compute simple correlation coefficients between estimated learning rates and measures of theory of mind from the AQ and BAP questionnaires. Recall that on both instruments, a higher score indicates increased presence of autism spectrum behaviors. Thus, negative correlations will indicate that our concept of *ToP* is analogous to the information in the AQ and BAP surveys, while the absence of correlation or positive correlations will indicate otherwise.

	Learning Rate
BAP	-0.22**
BAP_Rigid	-0.02
BAP_Aloof	-0.27***
BAP_Prag	-0.17*
AQ	-0.28***
AQ_Social	-0.28***
AQ_Switch	-0.14*
AQ_Detail	0.02
AQ_Commun	-0.23**
AQ_Imagin	-0.14*

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.2: Correlations between Autism Spectrum Measures and Individual Learning Rates.

BAP and AQ are overall scores from each instrument. Other variables are individual scores on subscales of each instrument. BAP_Rigid = Rigidity, BAP_Aloof = Aloofness, BAP_Prag = Pragmatic Language Deficit, AQ_Social = Social Skills, AQ_Switch = Attention Switching, AQ_Detail = Attention to Detail, AQ_Commun = Communication Skills, and AQ_Imagin = Imagination.

Table 2.2 report these simple correlations between measures from our experiment and survey measures of autism spectrum intensity.[38] From the table, we can

[36] The data reported here exclude one extreme outlier from our 20th session who chose the *SPE*-consistent action in 87/90 periods and whose estimated β was more than 12 times greater (18.64) than the next fastest-learning subject (1.46).

[37] Note that this regression equation is derived from the findings from column 4 of Table 2.1 in that we include only those independent variables that were statistically significant.

[38] Figure A.42 displays these correlations for the AQ and BAP scores.

see that learning rates are significantly negatively correlated with both the AQ and BAP scores as well as most of the subscales.[39] Taken together this provides solid evidence that our games measure theory of mind as it is conceived by psychologists. Reading through the questionnaires, this correlation agrees with intuition. For example, consider the finding that our measure of *ToP* is highly significantly correlated with the two subscales that emphasize *social skills*: AQ_Social and BAP_Aloof. We highlight these subscales because they are explicitly designed to measure capacity for and enjoyment of social interaction, which is particularly reliant on theory of mind. One particularly telling item on the AQ_Social subscale asks individuals how strongly they agree with the statement:

"I find it difficult to work out people's intentions."

This is consistent with our notion of *ToP* in a strategic setting. We also observe that learning is correlated with the AQ_Commun and AQ_Imagin subscales. The latter measures "imagination" by asking respondents to what degree they enjoy/understand fiction and fictional characters. One question asks about the ability to impute motives to fictional characters, which suggests some overlap with the AQ_Social subscale.

Interestingly, there is one AQ subscale, AQ_Detail, that exhibits a non-negative correlation and it emphasizes precision in individual habits and attention to detail. In a strategic setting such as ours, these traits might be expected to partly counteract the negative effects of other typical theory of mind deficits, perhaps accounting for the lack of significant correlation.

Importantly, our survey data exhibit scores in the normal range. Thus, differences in the strategic aspects of theory of mind vary significantly across individuals in the normal range of social intelligence.[40]

Finding 3: Our dynamic measure of *ToP* based on observed learning is significantly correlated with survey measures of theory of mind.

[39] Following convention, the BAP score is the mean of the three BAP subscale scores, and the AQ score is the sum of the five AQ subscale scores.

[40] Figure A.43 in the appendix displays histograms of our participants' AQ and BAP scores over the range of feasible scores.

2.4 Conclusions

This paper presents a model of the evolution of strategic sophistication. The model investigates the advantages to learning opponents' preferences in simple games of perfect information. An unusual feature is that the outcomes used in the game are randomly selected from a growing outcome set. We show how sophisticated individuals who recognize agency in others can build up a picture of others' preferences while naive players, who react only to their own observed payoffs in novel situations, remain in the dark. We impose plausible conditions under which some sophisticated individuals, who choose the *SPE* action, dominate all other strategies—naive or sophisticated—in the long run. That is, we establish a clear sense in which it is best to be smart, in contrast to previous results.

We then perform experiments measuring the ability of real-world individuals to learn the preferences of others in a strategic setting. The experiments implement a simplified version of the theoretical model, using a two-stage game where each decision node involves two choices. We find 1) evidence of highly significant learning of opponents' preferences over time, but not of complete games, and 2) significant correlations between behavior in these experiments and responses to two well-known survey instruments measuring theory of mind from psychology. Indeed, the experiments here raise the interesting possibility of developing a test for autism that is behavioral rather than purely verbal.

We show that the essential capacity to attribute preferences to others is theoretically evolutionarily plausible and actually present in the population to a varying degree. Other social phenomena that assume the presence of theory of mind then gain firmer footing, and so an indirect contribution of our work is to set the stage for future research on such phenomena.

2.5 Proof of the Main Result: The Simplest Case

We establish here our main result for the simple version of the model in which there are only two player roles, two choices at each node, and only two strategies adopted by the 1's—an *SPE-ToP* strategy and a naive strategy (see Definitions 2.2, and 2.3, where these are introduced). Role 2 makes the *SPE* choice in every game (Assumption 2.4), and thus the results here concern the player 1's.

The notation is modified slightly in the proof. In particular, a single subscript is used to denote the total number of accumulated iterations, in lieu of subscripting the period n, and iteration t. For example, H_s is written in place of the history $H_{n,t}$, where s is now the number of accumulated iterations along $H_{n,t}$. [41] For each period $n = 1, 2, \ldots$, the notation $s(n)$ is used to denote the iteration $s = \sum_{m=1}^{n-1} \kappa(m) + 1$. Notice, in particular, that the n-th novel outcome arrives at the beginning of iteration $s(n)$.

Assume throughout the section that Assumptions 2.1-2.5 hold.

2.5.1 The ***SPE-ToP*** Strategy is Optimal if $\alpha > 1$.

The first step in the proof of Theorem 2.1 in this simple environment, with two players and two actions at each stage, is to show that if $\alpha > 1$, then the *SPE-ToP*s learn their opponents' preferences completely in the limit, and therefore choose optimally against their opponents with probability tending to one.

In this simple environment with two player roles and two actions, each choice by the 2's directly reveals a pairwise preference. Specifically, the 2's make the dominant choice, as in Assumption 2.4, and moreover, every choice by the 2's eliminates all ambiguity about their preferred option, since there are no remaining players. One measure of how much has been revealed about 2's preferences is therefore the number of distinct 2 role subgames reached along the history. Consider in particular the following.

Definition 2.5: *Let K_s denote the number of distinct role 2 subgames that have been reached along H_s. Recall that there are $|Z_n|^2$ role 2 subgames throughout period n. For each $s = s(n), \ldots, s(n+1) - 1$ write $L_s = K_s/|Z_n|^2$ as a measure of how much can be learned about 2's preferences from H_s.*

L_s is a conservative measure of how much information is conveyed by the history about 2's preferences, but it suffices for the present purpose. [42] Specifically, we have the key result that L_s converges in probability to one whenever $\alpha > 1$. The proof is immediate in the light of the next two results (Lemma 2.2, and Lemma 2.3).

[41] Recall that $H_{n,t}$ records the randomly drawn games, and the empirical distributions of choices in all prior iterations, and note that there are $s = \sum_{m=1}^{n-1} \kappa(m) + t - 1$ iterations of the game along $H_{n,t}$.

[42] This ignores the transitivity of 2's preference ordering. Moreover, the denominator accounts for all of the available 2 subgames, including trivial ones in which the 2's face the same two outcomes.

LEMMA 2.2: *Suppose that there are two player roles, and two actions available for each role. Suppose further that L_s converges in probability to some random variable L. If $\alpha > 1$, then $L = 1$ a.e.*

Proof: With probability $(1 - L_s)^2$ the game at iteration s is such that neither of its role 2 subgames have occurred along the history. Such a game at s ensures K_s increases by one, at least—that is, no matter how the 1's choose they observe 2's choice in a novel subgame. Hence, for every $s = 1, 2, \dots,$

$$E(K_{s+1} \mid H_s) - K_s \geq (1 - L_s)^2. \tag{2.1}$$

That is, the smaller is the proportion of 2 role subgames that have been seen along the history, the more likely it is that an unfamiliar one will arise.

Summing (2.1) over $s = 1, 2, \dots, \tau$, and taking the unconditional expectation of the result yields

$$E(K_\tau) - E(K_1) \geq \sum_{s=1}^{\tau-1} E((1 - L_s)^2). \tag{2.2}$$

Now, for each iteration, $\tau = 1, 2, \dots$, let $n(\tau)$ denote the period prevailing during the iteration. Notice, in particular, that $L_\tau = K_\tau / |Z_{n(\tau)}|^2$ for each $\tau = 1, 2, \dots$. Next, observe that $K_1 = 0$ by definition. Dividing both sides of equation (2.2) by $|Z_{n(\tau)}|^2$ thus gives

$$E(L_\tau) \geq \frac{\tau - 1}{|Z_{n(\tau)}|^2} \cdot \left[\frac{1}{\tau - 1} \cdot \sum_{s=1}^{\tau-1} E((1 - L_s)^2) \right]. \tag{2.3}$$

Now suppose $\alpha > 1$, and consider (2.3) as τ tends to infinity. The first thing to note is that the $(\tau - 1)/|Z_{n(\tau)}|^2$ term in the expression diverges to infinity. To see this observe the following. The iteration corresponding to the arrival of the n-th novel outcome, $s(n) = \sum_{m=1}^{n-1} \kappa(m) + 1$, is non-decreasing in n, and has order of $n^{1+\alpha}$. Since each iteration $\tau = 1, 2, \dots,$ satisfies $s(n(\tau)) \leq \tau \leq s(n(\tau) + 1) - 1$, it follows that $n(\tau)$ has order of $\tau^{\frac{1}{1+\alpha}}$, and hence that $|Z_{n(\tau)}|^2 = (N + n(\tau))^2$ has order of $\tau^{\frac{2}{1+\alpha}}$. Clearly if $\alpha > 1$, then $\tau - 1$ grows at a faster rate than $|Z_{n(\tau)}|^2$.

Next, notice that the quantity on the right hand side of (2.3) is surely bounded above by one, uniformly in τ (since surely $L_s \leq 1$). The limit inferior of the bracketed means in the expression must then be zero, since otherwise the quantity on the right hand side would diverge to infinity (because $(\tau - 1)/|Z_{n(\tau)}|^2 \longrightarrow \infty$ as argued already).

Now suppose L_s converges in probability to the random variable L as hypothesized in the statement of the claim. The bracketed means in (2.3) will then converge to $E((1 - L)^2)$. Since the limit inferior of these means is zero, it follows that $E((1 - L)^2) = 0$, and hence that $L = 1$ a.e. ■

The next result states that if $\alpha > 1$, then L_s converges in probability to some random variable L. Taken together with Lemma 2.2, this implies that L_s converges in probability to one if $\alpha > 1$. The proof of convergence is rather technical, and involved even for the simple version of the model that we focus on here. A complete proof is given in this section but intuitive arguments are relied upon whenever these are thoroughly convincing. (An entirely rigorous proof is given in the proof of the general case in the subsequent section.) In anticipation of the convergence result consider first some key observations.

The crucial factor regarding the putative convergence of L_s is the behavior of the process along the subsequence, $\{s(n)\}$, of iterations corresponding to the arrivals of novel outcomes. In particular, if the process along this subsequence converges to some limit, then the overall sequence must converge, and moreover, it must possess the same limit. In order to see this, the first thing to notice is that L_s is non-decreasing in between the arrivals of novel outcomes. Specifically, the numerator, K_s, never decreases, and the denominator, $|Z_n|^2$, is constant until the next outcome is introduced. Notice, however, that although the process has the sub-martingale property in between arrivals of novelty, it is not properly a sub-martingale. The introduction of the $n + 1$-th new outcome causes the denominator to increase by a factor of n (i.e., the denominator changes from $|Z_n|^2$ to $|Z_{n+1}|^2$), inducing a sudden decrease in L_s. [43] It is important, however, that as the number of outcomes increases, the drop in L_s due to the arrival of yet another outcome becomes smaller, tending to zero eventually. That is, as n becomes larger, the increase in the denominator caused by the $n+1$-th new outcome ($|Z_{n+1}|^2 - |Z_n|^2$, in particular) occupies a smaller fraction of the number of pairs of outcomes that were already there (i.e., $|Z_n|^2$). To be more precise about this we note that

[43] If L_s were a sub-martingale, the almost sure convergence of the sequence would follow immediately from the martingale convergence theorem.

$$\begin{aligned} L_{s(n)} &\geq K_{s(n)-1}/|Z_n|^2 \\ &= (K_{s(n)-1}/|Z_{n-1}|^2) \cdot (|Z_{n-1}|^2/|Z_n|^2) \\ &= L_{s(n)-1} \cdot (|Z_{n-1}|^2/|Z_n|^2), \end{aligned}$$

and hence that $\liminf\{L_{s(n)} - L_{s(n)-1}\} = 0$, surely, since $|Z_{n-1}|^2/|Z_n|^2$ surely converges to one. Suppose now that $L_{s(n)}$ converges in probability to the random variable L. The above discussed facts (displayed graphically in Figure 2.4) imply that, for sufficiently large n, if $L_{s(n-1)}$ and $L_{s(n)}$ are close to their limiting value, then L_s must be close to this limit as well, for each $s = s(n-1), \ldots, s(n) - 1$.

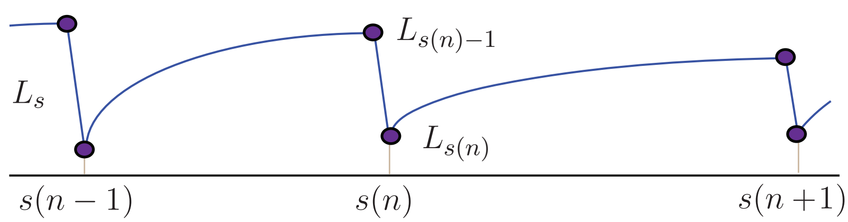

Figure 2.4: A Representative Sample Path of L_s

Clearly if $L_{s(n)}$ converges in probability to L, then so must the overall sequence, $\{L_s\}$, a fact that is used in proving the next result—

LEMMA 2.3: *Suppose that there are two player roles, and two actions available for each role. If $\alpha > 1$, then there is a random variable L such that L_s converges in probability to L.*

Proof: Given the discussion prior to the statement of the lemma it suffices to show that if $\alpha > 1$, then the subsequence $\{L_{s(n)}\}$ converges in probability to some random variable L. Since we work exclusively with this subsequence in the proof, we elect from now on to simplify notation by writing $\bar{L}_n$, $\bar{K}_n$, and $\bar{H}_n$ in place of $L_{s(n)}$, $K_{s(n)}$, and $H_{s(n)}$, respectively, for each $n = 1, 2, \ldots$. In order to establish the convergence in probability of $\bar{L}_n$ we rely on the following definition and result (found in Egghe (1984) [Definition VIII.1.2, and Theorem VIII.1.22]).

SUBMIL CONVERGENCE: *The process $\{\bar{L}_n\}$ is a sub-martingale in the limit (submil) if for each $\eta > 0$ there is almost surely an integer M such that $n > m \geq M$ implies*

$E(\bar{L}_n \mid \bar{H}_m) - \bar{L}_m \geq -\eta$. If $\bar{L}_n$ is a submil, then there exists a random variable L such that $\bar{L}_n$ converges in probability to L. [44]

Considering that submils converge in probability, we proceed to prove the lemma by showing that if $\alpha > 1$, then $\bar{L}_n$ is a submil. Toward this end, consider two periods, m, and n, such that $n > m$. Given that $\bar{L}_n = \bar{K}_n/|Z_n|^2$, by definition, it is straightforward to show that

$$\begin{aligned} E(\bar{L}_n - \bar{L}_{n-1} \mid \bar{H}_m) &< 0 \quad \Longrightarrow \\ E(\bar{K}_n - \bar{K}_{n-1} \mid \bar{H}_m) &< |Z_n|^2 - |Z_{n-1}|^2. \end{aligned} \tag{2.4}$$

That is, $E(\bar{L}_n \mid \bar{H}_m)$ is less than $\bar{L}_{n-1}$ only if the expected number of new subgames reached during period $n-1$ (the expected increase in the numerator of $\bar{L}_{n-1}$) is less than the number of new subgames introduced by the n-th novel outcome.

With (2.4) in mind, consider that $|Z_n|^2 - |Z_{n-1}|^2$ has order of $n-1$, while the number of iterations during period $n-1$ is $\kappa(n-1)$, which has order of $(n-1)^\alpha$. We see that when $\alpha > 1$ the number of iterations between arrivals outstrips the number of new pairwise choices introduced by each outcome. It is thus intuitively plausible that the $\bar{L}_n$ process will keep up with the arrival of novelty, and that the process is therefore indeed a submil. In order to confirm this intuition we revisit equation (2.1), summing this time over $s = s(n), \ldots, s(n+1)-1$, to obtain

$$\begin{aligned} E(\bar{K}_n - \bar{K}_{n-1} \mid \bar{H}_m) &\geq \sum_{s=s(n-1)}^{s(n)-1} E((1-L_s)^2) \mid \bar{H}_m) \\ &\geq \kappa(n-1) \cdot E((1-L_{s(n)-1})^2 \mid \bar{H}_m). \end{aligned} \tag{2.5}$$

To get the second line here we used the fact that there are $\kappa(n-1)$ terms in the summation, and that L_s is non-decreasing as s ranges from $s(n-1)$ to $s(n)-1$. Combining (2.5) with (2.4) we see that

$$\begin{aligned} &E(\bar{L}_n - \bar{L}_{n-1} \mid \bar{H}_m) < 0 \Longrightarrow \\ &E((1-L_{s(n)-1})^2 \mid \bar{H}_m) < \left(|Z_n|^2 - |Z_{n-1}|^2\right)/\kappa(n-1). \end{aligned} \tag{2.6}$$

Now suppose $\alpha > 1$. In this case the $(|Z_n|^2 - |Z_{n-1}|^2)/\kappa(n-1)$ term in equation (2.6) surely converges to zero. The same equation then implies that for sufficiently large n, $E(\bar{L}_n - \bar{L}_{n-1} \mid \bar{H}_m)$ is negative only if $E(L_{s(n)-1} \mid \bar{H}_m)$ is suffi-

[44] The convergence of a submil requires that the process in question be uniformly integrable. L_s satisfies this requirement since, by definition, $|L_s| \leq 1$ surely.

ciently close to one. [45] But as we argued before the statement of the lemma, $\liminf\{\bar{L}_n - L_{s(n)-1}\} = 0$ surely, and thus it follows that for sufficiently large n, $E(\bar{L}_n - \bar{L}_{n-1} \mid \bar{H}_m)$ is negative only if $E(\bar{L}_n \mid \bar{H}_m)$ is close to one. More formally, we have the following. Whenever $\alpha > 1$, for each $\eta > 0$ there is a finite integer, $M(\eta)$, such that:

$$\begin{aligned}&\text{If } n > m \geq M(\eta), \text{ then}\\ &E(\bar{L}_n - \bar{L}_{n-1} \mid \bar{H}_m) \ < 0 \quad \Longrightarrow \quad E(\bar{L}_n \mid \bar{H}_m) \ > \ 1 - \eta.\end{aligned} \tag{2.7}$$

It is this feature of $\bar{L}_n$ that imparts to the process the submil property. This is what is shown next.

To see that $\bar{L}_n$ is a submil fix η and choose $M(\eta)$ as in (2.7). Consider any m, and n such that $n > m \geq M(\eta)$. Suppose for the sake of the argument that $E(\bar{L}_n \mid \bar{H}_m) \leq 1 - \eta$. Then (2.7) implies $E(\bar{L}_n - \bar{L}_{n-1} \mid \bar{H}_m) \geq 0$, and therefore that $E(\bar{L}_{n-1} \mid \bar{H}_m) \leq 1 - \eta$. This in turn implies (using (2.7) again) that $E(\bar{L}_{n-1} - \bar{L}_{n-2} \mid \bar{H}_m) \geq 0$, and therefore that $E(\bar{L}_{n-2} \mid \bar{H}_m) \leq 1 - \eta$. Proceeding recursively in this manner we see that $E(\bar{L}_n \mid \bar{H}_m) \leq 1 - \eta$ implies $E(\bar{L}_k - \bar{L}_{k-1} \mid \bar{H}_m) \geq 0$, for each $k = m + 1, \ldots, n$, and therefore that $E(\bar{L}_n \mid \bar{H}_m) - \bar{L}_m \geq 0$. Clearly $E(\bar{L}_n \mid \bar{H}_m) - \bar{L}_m < 0$ only if $E(\bar{L}_n \mid \bar{H}_m) > 1 - \eta$. Since $\bar{L}_m$ is surely no greater than one this means that $E(\bar{L}_n \mid \bar{H}_m) - \bar{L}_m \geq -\eta$, and hence that $\bar{L}_n$ is a submil.■
Lemmas 2.2, and 2.3 in combination give—

LEMMA 2.4: *Suppose that there are two player roles, and two actions available for each role. If $\alpha > 1$, then the history reveals role 2 preferences completely in the limit, that is, L_s converges in probability to one.*

The *SPE-ToP* strategy of role 1 makes the *SPE* choice whenever the 2's choices in the game have been observed previously along H_s (See Definition 2.3 and the discussion in the example after it). Lemma 2.4 then sets the stage for the ultimate dominance of the *SPE-ToP* strategy.

[45] For further detail, notice that Jensen's inequality implies

$$E((1 - L_{s(n)-1})^2 \mid \bar{H}_m) \geq E(1 - L_{s(n)-1} \mid \bar{H}_m)^2.$$

Applying this in (2.6) gives, after a straightforward rearrangement, that $E(\bar{L}_n - \bar{L}_{n-1} \mid \bar{H}_m) \ < 0$ only if

$$E(L_{s(n)-1} \mid \bar{H}_m) > 1 - \sqrt{\frac{|Z_n|^2 - |Z_{n-1}|^2}{\kappa(n-1)}}.$$

2.5.2 The Naive Strategy is Suboptimal if $\alpha < 3$.

The next result is that the naive strategy makes a suboptimal choice with positive probability in the long run. Taken together with Lemma 2.4, the result implies that the *SPE-ToP* strategy outdoes the naive strategy eventually.

LEMMA 2.5: *Suppose $\alpha < 3$, and that the fixed game tree has four terminal nodes. Then the payoff to the 1's from the naive strategy is dominated by the SPE payoff with probability that is bounded away from zero in the limit.*

Proof: Let G denote the distribution of games implied by F, where F is the cdf according to which new outcomes are introduced in every period.

Recall that the naive strategy has a preprogrammed initial reaction in every game that is unfamiliar, that lacks moreover a dominant choice (cf. Definition 2.2). The proof proceeds along the following lines: 1) There is a set of games with positive measure under G, for which the initial reaction of the naive strategy differs from the *SPE* choice, 2) these games come up with positive probability in the limit, and 3) the game is new with probability tending to one.

We begin with the first claim. Notice that there is a positive measure subset of games, say Q', that lack a dominant action for 1, and in which role 2's payoffs are all distinct. For every game in this subset, if the initial response of the naive strategy in the game is the *SPE* choice, then this choice can be rendered suboptimal by some rearrangement of 2's payoffs. Therefore, if there is a positive measure subset of Q' such that the initial reaction by the naive type is optimal, then there must also be a positive measure subset within Q' where the initial reaction is sub-optimal. (Important here is that the initial reaction of the naive strategy conditions only on own payoffs [cf. Definition 2.2]).

The second claim is implied by the first in the light of the Glivenko-Cantelli Lemma (stated as Lemma 2.1 in the main body of the paper). (In particular, Glivenko-Cantelli implies that the long run empirical distribution over games converges to G almost surely.)

The last claim, that all the games are new in the long run, is proved as follows. The number of distinct subgames occurring along a history H_s is at most s, since only one game is played at each iteration. The number of games available during iteration s has order of $s^{\frac{4}{1+\alpha}}$. If $\alpha < 3$, then $s/s^{\frac{4}{1+\alpha}}$ tends to zero, and thus the fraction of familiar games tends to zero. ■

2.5.3 The *SPE-ToP* Strategy Dominates if $\alpha \in (1, 3)$.

Taking stock of what has been established thus far, we see that if $\alpha \in (1, 3)$, then the *SPE-ToP*s make the *SPE* choice with probability tending to one (Lemma 2.4), while at the same time the payoff to the naive strategy is dominated by the payoff from the *SPE* choice, with probability bounded away from zero in the limit (Lemma 2.5). Since the player 2's make the *SPE* choice always, the *SPE* choice is the optimal for the 1's. The *SPE-ToP* strategy therefore outdoes the naive strategy in terms of payoffs in the long run. It follows naturally that the *SPE-ToP* strategy eventually dominates.

We therefore end with—

LEMMA 2.6: *Suppose the fraction of role* 1*'s that use the SPE-ToP strategy in iteration* s *is* $f_s \in [0, 1]$. *If* $\alpha \in (1, 3)$, *then* f_s *converges in probability to one.*

Although the intuition for the result is compelling, a fully rigorous proof involves rather long calculations. We therefore defer the formal proof to the fully general case in the next section.

2.6 Proof of the Main Result: The General Case

The proof of the theorem is given in two parts.

2.6.1 The *SPE-ToP* Strategy is Optimal if $\alpha > 1$.

The result proved here is that the *ToP*s learn their opponents' preferences completely in the limit whenever $\alpha > 1$. (Recall that α determines the number of iterations within in each period as in Assumption 2.3.) This sets the stage for the ultimate dominance of the *SPE-ToP*s. [46] We first introduce some notation.

DEFINITION 2.6: $\Delta_n^i(z, z')$ *is the set of* i *role subgames, available in period* n, *that satisfy the following. The subgame* q *is in* $\Delta_n^i(z, z')$ *if and only if, for two actions,*

[46] In particular, the *SPE-ToP*s will then eventually choose an *SPE* in each game. Although in general this *SPE* choice is sub-optimal initially, it is the appropriate strategy in the long run (as will be established later).

say a, $a' \in A$, z is the unique SPE outcome of the subgame following i's choice of a, and z' is the unique SPE outcome of the subgame following i's choice of a', and moreover one of the actions a, a' is strictly dominant for the i players themselves.

The subgames of Definition 2.6 play a special role in how *ToP*s learn preferences. Recall that player roles are enumerated in reverse order of play.

Consider a situation in which player 1's reach a subgame $q \in \Delta_n^1(z, z')$. Suppose that z is strictly preferred by player 1's to z'. Assumption 2.4 implies all the player 1's reaching q will there choose the action resulting in z. Any *ToP* observing this choice will then know that player 1's prefer z to z'.

In order for a *ToP* player to learn then player 2's preferences over, say, z, z' it suffices for player 2's to be observed making a choice in a subgame $q \in \Delta_n^2(z, z')$ when all the player 1 pairwise choices in q had already been learned as above described. With this in mind consider the following.

Suppose $z, z' \in Z$ is such that the 1 players prefer z to z'. Say that the history $H_{n,t}$ *reveals* players in role 1 prefer z to z' if along $H_{n,t}$ a subgame $q \in \Delta_n^1(z, z')$ was reached and all the 1 role players there chose the action delivering z. Proceeding inductively, suppose $z, z' \in Z$ is such that player i's prefer z to z'. Say that $H_{n,t}$ *reveals* players in role $i > 1$ prefer z to z' if along $H_{n,t}$ a subgame $q \in \Delta_n^i(z, z')$ was reached, after all the pairwise preferences of the $i-1, \dots, 1$ role players in q had been revealed, and there all of the i role players chose into the $i-1$ subgame delivering z in the *SPE*.

As in the previous section, from now on we use a single subscript to denote the number of accumulated iterations of the game, rather than subscripting the period n, and iteration t of period n. For example, instead of $H_{n,t}$, we write H_s, where s is now the total number of iterations having occurred along $H_{n,t}$. Recall that the number of iterations in period n is $\kappa(n)$ (Assumption 2.3), and thus the number of iterations along $H_{n,t}$ is $\sum_{m<n} \kappa(m) + t - 1$.

We again make use of the notation $s(n)$ to denote $\sum_{m<n} \kappa(m) + 1$. Notice that for each $n = 1, 2, \dots$, iteration $s = s(n)$ corresponds to the arrival of the n-th novel outcome.

To keep an account of how much information has been revealed along a given history we define—

Definition 2.7: *For each iteration $s = 1, 2, \dots$, the random variable K_s^i is number of outcome pairs such that H_s reveals i role preferences on $\{z, z'\}$. For each*

$s = s(n), \dots, s(n+1) - 1$ *write* $L_s^i = K_s^i / |Z_n|^2$*—the fraction of* i *role pairwise preferences that are revealed along* H_s*.* [47]

A key step in establishing the eventual dominance of the *SPE-ToP*s is to show that if $\alpha > 1$, then L_s^i tends to one in probability, for each $i \geq 1$. The aim of the remainder of this section is to prove the following result.

LEMMA 2.7: *Suppose* $\alpha > 1$*, then* L_s^i *tends in probability to one,* $i = 1, \dots, I$*.* [48]

The proof relies on two preliminary results (Lemmas 2.8 and 2.9 below). First two definitions—

DEFINITION 2.8: *Consider* $i > 1$*. For each iteration* $s = 1, \dots,$ $I_s \in \{0, 1\}$ *is such that* $I_s = 0$ *if and only if the game drawn at iteration* s *is such that all the available pairwise choices of players* $j = 1, \dots, i-1$ *have been revealed along* H_s*.*

DEFINITION 2.9: *For each* $\varepsilon > 0$*, and* $n = 1, 2, \dots,$

$$\boldsymbol{S}_n^i(\varepsilon) = \left\{ (z, z') \in Z_n \times Z_n \,:\, |\Delta_n^i(z, z')| / |Z_n|^{A^i - 2} < \varepsilon \right\}.$$

When ε is small, the number of subgames in $\bigcup_{\boldsymbol{S}_n^i(\varepsilon)} \Delta_n^i(z, z')$ is a small fraction of the number of possible i role subgames. [49]

The roles of I_s, and $\boldsymbol{S}_n^i(\varepsilon)$ are clarified in the following lemma.

LEMMA 2.8: *Each of the following is true.*

1. *Consider* $i > 1$*. For each* $\varepsilon > 0$*, each period* $n = 1, 2, \dots,$ *and each* $s = s(n), , \dots, s(n+1) - 1$*,*

$$\begin{aligned} &E(K_{s+1}^i \,|\, H_s) - K_s^i \geq \\ &- E(I_s \,|\, H_s) + \left[\varepsilon \cdot \max\left\{ 0, 1 - L_s^i - \frac{|\boldsymbol{S}_n^i(\varepsilon)|}{|Z_n|^2} \right\} \right]^{A^{I-i}}. \end{aligned}$$

[47] Assumption 2.2 implies that with probability one, throughout period n, the number of pairs $(z, z') \in Z_n \times Z_n$ such that $z_i \neq z_i'$ is $|Z_n| \cdot [|Z_n| - 1]/2$. Thus, with probability one this is the maximal number of preference revelations possible up to period n. We opt in favor of the simpler expression $|Z_n|^2$ in the denominator of L_s^i.

[48] If $\alpha < 1$, it follows that $L_s^i \to 0$ surely. This can be proved with a simple adaptation of the proof of Lemma 2.12.

[49] Recall that the set of i role subgames in period n is $Q_n^i \equiv |Z_n|^{A^i}$, and thus

$$\frac{\left| \bigcup_{(z,z') \in \boldsymbol{S}_n^i(\varepsilon)} \Delta_n^i(z, z') \right|}{|Q_n^i|} < \varepsilon \cdot \frac{|\boldsymbol{S}_n^i(\varepsilon)|}{|Z_n|^2} \leq \varepsilon.$$

If $i = 1$, the above expression holds for i with I_s identically equal to zero.

2. *There exists an $S(\varepsilon)$ such that $S(\varepsilon)$ tends to zero as ε tends to zero, and such that for each $\varepsilon > 0$, $|\boldsymbol{S}_n^i(\varepsilon)|/|Z_n|^2$ almost surely converges to $S(\varepsilon)$.*

Proof: Consider the first enumerated claim. Fix an $i \geq 1$, and a period n. Consider in all of the following an iteration, $s \in \{s(n), \ldots, s(n+1)-1\}$.

Let N_s denote the outcome pairs $(z, z') \in Z_n \times Z_n$ such that i's preferences on $\{z, z'\}$ have not been revealed along H_s. Notice that $L_s^i = 1 - |N_s|/|Z_n|^2$.

Define $J_s \in \{0, 1\}$ such that $J_s = 1$ if and only if, at iteration s for each i player subgame, q, there is some $(z, z') \in N_s$ such that q is in $\Delta_n^i(z, z')$. For $i = 1$, since there are no players after i, set $I_s \equiv 0$ in all of the following expressions.

Note that if $[1-I_s] \cdot J_s = 1$, then i's choice at any i subgame reveals i preferences over some pair of outcomes $(z, z') \in N_s$, and therefore

$$K_{s+1}^i - K_s^i \geq [1 - I_s] \cdot J_s.$$

Since $[1 - I_s] \cdot J_s \geq J_s - I_s$, it follows that

$$E(K_{s+1}^i \mid H_s) - K_s^i \geq E(J_s \mid H_s) - E(I_s \mid H_s). \tag{2.8}$$

Next observe that $E(J_s \mid H_s) = P\{J_s = 1 \mid H_s\}$, and that

$$P\{J_s = 1 \mid H_s\} \geq \left[\sum_{(z,z') \in N_s} \frac{|\Delta_n^i(z, z')|}{|Q_n^i|}\right]^{A^{I-i}}. \tag{2.9}$$

This is because the distribution over games in period n can induced by the A^{I-i}-times independent sampling of i player subgames, uniformly from Q_n^i, while the fraction of i role subgames in $\Delta_n^i(z, z')$ at date n is $|\Delta_n^i(z, z')|/|Q_n^i|$.

Using $\boldsymbol{S}_n^i(\varepsilon)$ in equation (2.9) gives,

$$\begin{aligned} P\{J_s = 1 \mid H_s\} &\geq \left(\sum_{(z,z') \in N_s \setminus \boldsymbol{S}_n^i(\varepsilon)} \frac{|\Delta_n^i(z, z')}{|Q_n^i|}\right)^{A^{I-i}} \\ &\geq \left(\varepsilon \cdot \frac{|N_s \setminus \boldsymbol{S}_n^i(\varepsilon)|}{|Z_n|^2}\right)^{A^{I-i}} \end{aligned} \tag{2.10}$$

$$\geq \left[\varepsilon \cdot \max\left\{0, \frac{|N_s|}{|Z_n|^2} - \frac{|\boldsymbol{S}_n^i(\varepsilon)|}{|Z_n|^2}\right\}\right]^{A^{I-i}}$$

$$= \left[\varepsilon \cdot \max\left\{0, 1 - L_s^i - \frac{|\boldsymbol{S}_n^i(\varepsilon)|}{|Z_n|^2}\right\}\right]^{A^{I-i}} .$$

Equations (2.8) and (2.10) together deliver the desired result.

The second enumerated claim follows by direct application of Lemma 2.1 in the light of Assumption 2.2. ■

LEMMA 2.9: *Let x_s, $s = 1, 2, \ldots,$ be a sequence taking values in $[0, 1]$. Given $\bar{\varepsilon} > 0$, consider a family of sequences, $\{\theta_s(\varepsilon)\}$, $\varepsilon \in (0, \bar{\varepsilon}]$, that satisfy the following conditions. For each $\varepsilon \in (0, \bar{\varepsilon}]$, $\lim_{s\to\infty} \theta_s(\varepsilon) = \theta(\varepsilon) \in \mathbb{R}_+$, where $\theta(\varepsilon)$ tends to one as ε tends to zero. Suppose $\liminf[x_{s+1} - x_s] \geq 0$, and that $x_{s+1} - x_s < 0$ only if $x_s > \theta_s(\varepsilon)$. Then x_s converges to some limit $\hat{x} \in [0, 1]$.*

Proof: Fix an arbitrary $\eta > 0$. Choose a $\varepsilon \in (0, \bar{\varepsilon}]$ such that $\theta(\varepsilon) > 1 - \eta/3$, and then choose T_1 so that if $s \geq T_1$, then $\theta_s(\varepsilon) > 1 - 2 \cdot \eta/3$. Choose T_2 such that $x_{s+1} - x_s > -\eta/3$ for every $s \geq T_2$. Define $T = \max\{T_1, T_2\}$, and $\sigma = \inf\{s \geq T : x_{s+1} - x_s < 0\}$. If $\sigma = \infty$, then clearly x_s converges. Suppose then that σ is finite. Observe the following. Whenever $s \geq T$, if $x_{s+1} - x_s < 0$, then $x_s > \theta_s(\varepsilon) > 1 - 2 \cdot \eta/3$. Thus, since $x_{s+1} - x_s > -\eta/3$, for all $s \geq T$, it follows that $x_{s+1} - x_s < 0$ implies $x_{s+1} > 1 - \eta$, whenever $s \geq T$. It follows then that $x_s > 1 - \eta$ for each $s > \sigma$. Clearly x_s converges since η can be chosen arbitrarily small. ■

Lemma 2.9 plays a role in showing that the sequences $\{E(L_s^i)\}$, $i = 1, \ldots, I$, are convergent (Lemma 2.10 below). First another preliminary result—

LEMMA 2.10: *If the subsequence $\{E(L_{s(n)}^i)\}$ converges, then $\{E(L_s^i)\}$ converges and possesses the same limit. If the subsequence $\{L_{s(n)}^i\}$ converges in probability to $\bar{L}^i$, say, then so does $\{L_s^i\}$.*

Proof: It suffices to prove the claim concerning convergence in probability (i.e, convergence in probability implies the expectations converge). Suppose then that the subsequence $\{L_{s(n)}^i\}$ converges in probability to $\bar{L}^i$.

Fix a period n, and consider an iteration $s \in \{s(n), \ldots, s(n+1)-1\}$. Recall that $L_s^i = K_s^i/|Z_n|^2$. Since K_s^i is non-decreasing in s, and the denominator $|Z_n|^2$ in L_s^i is constant for $s = s(n), \ldots, s(n+1)-1$, it follows that $L_s^i \geq L_{s(n)}^i$.

One novel outcome is introduced at iteration $s(n+1)$, and therefore $L_{s(n+1)}^i \geq L_s^i \cdot |Z_n|^2/|Z_{n+1}|^2$. It thus follows that

$$L_{s(n)}^i \;\leq\; L_s^i \;\leq\; L_{s(n+1)}^i \cdot |Z_{n+1}|^2/|Z_n|^2. \tag{2.11}$$

This establishes the result since $|Z_{n+1}|^2/|Z_n|^2$ converges to one surely. ■

LEMMA 2.11: *Suppose $\alpha > 1$. Then $E(L_s^1)$ converges to some limit $\bar{L}^1 \in [0,1]$. Suppose L_s^j converges in probability to one for each $j = 1, \ldots, i-1$, then $E(L_s^i)$ converges to some limit $\bar{L}^i \in [0,1]$.*

Proof: Fix $i \geq 1$. In view of Lemma 2.10 it suffices to prove that the subsequence $\{E(L_{s(n)}^i)\}$ converges. With this in mind notice first that

$$E(K_{s(n+1)}^i) - E(K_{s(n)}^i) = \sum_{s=s(n)}^{s(n+1)-1} E\left(E(K_{s+1} \,|\, H_s) - K_s\right). \tag{2.12}$$

Write, for each $s = s(n), \ldots, s(n)-1$, $Y_s(\varepsilon) = 1 - L_{n,t}^i - |\boldsymbol{S}_n^i(\varepsilon)|/|Z_n|^2$, and apply Lemma 2.8 to equation (2.12) to obtain

$$\begin{aligned} E(K_{s(n+1)}^i) - E(K_{s(n)}^i) \geq & - \sum_{s=s(n)}^{s(n+1)-1} E\left(I_s\right) \\ & + \sum_{s=s(n)}^{s(n+1)-1} E\left(\left[\varepsilon \cdot \max\left\{0, Y_s(\varepsilon)\right\}\right]^{A^{I-i}}\right). \end{aligned} \tag{2.13}$$

Twice applying Jensen's inequality (i.e., first $E(X^N) \geq E(X)^N$, given a random variable X, and then $E(\max\{X_1, Y_2\}) \geq \max\{E(X_1), E(X_2)\}$ given X_1, X_2) yields

$$\begin{aligned} E(K_{s(n+1)}^i) - E(K_{s(n)}^i) \geq & - \sum_{s=s(n)}^{s(n+1)-1} E\left(I_s\right) \\ & + \sum_{s=s(n)}^{s(n+1)-1} \left[\varepsilon \cdot \max\left\{0, E\left(Y_s(\varepsilon)\right)\right\}\right]^{A^{I-i}}. \end{aligned} \tag{2.14}$$

L_s^i is non-decreasing as s ranges from $s(n)$ to $s(n+1)-1$, and thus $Y_s(\varepsilon) \geq Y_{s(n+1)-1}(\varepsilon)$, for each $s = s(n), \ldots, s(n+1)-1$. Using this in (2.14) yields

$$E(K^i_{s(n+1)}) - E(K^i_{s(n)}) \geq -\sum_{s=s(n)}^{s(n+1)-1} E\left(I_s\right) + [\kappa(n)-1] \cdot \left[\varepsilon \cdot \max\left\{0, E\left(Y_{s(n+1)-1}(\varepsilon)\right)\right\}\right]^{A^{I-i}}. \tag{2.15}$$

Observe next that

$$L^i_{s(n+1)} - L^i_{s(n)} < 0 \implies K^i_{s(n+1)} - K^i_{s(n)} < |Z_{n+1}|^2 - |Z_n|^2.$$

Combining this with (2.15) gives

$$\begin{aligned} &E(L^i_{s(n+1)}) - (L^i_{s(n)}) < 0 \implies \\ &|Z_{n+1}|^2 - |Z_n|^2 > -\sum_{s=s(n)}^{1} E\left(I_s\right) \\ &+ [\kappa(n)-1] \cdot \left[\varepsilon \cdot \max\left\{0, E\left(Y_{s(n+1)-1}(\varepsilon)\right)\right\}\right]^{A^{I-i}}. \end{aligned} \tag{2.16}$$

After some rearranging in (2.16) it becomes clear that $E(L^i_{s(n+1)}) - E(L^i_{s(n)}) < 0$ only if

$$E\left(Y_{s(n+1)-1}(\varepsilon)\right) < \frac{1}{\varepsilon}\left[\frac{|Z_{n+1}|^2 - |Z_n|^2}{s(n+1)-1} + \frac{1}{s(n+1)-1}\sum_{s=s(n)}^{s(n+1)-1} E\left(I_s\right)\right]^{1/A^{I-i}}. \tag{2.17}$$

Recall that $Y_s(\varepsilon) = 1 - L_s^i - |\boldsymbol{S}_n^i(\varepsilon)|/|Z_n|^2$. Since $L^i_{s(n+1)} \geq L^i_s - 2/|Z_n|$ for each $s = s(n), \ldots, s(n+1)-1$, it follows that

$$Y_s(\varepsilon) \geq 1 - \hat{L}_{n+1} + 2/|Z_n| - |\boldsymbol{S}_n^i(\varepsilon)|/|Z_n|^2,$$

for each $s = s(n), \ldots, s(n+1)-1$.

Using (2.6.1) in (2.17) and then solving for $E(L^i_{s(n+1)})$ yields: $E(\hat{L}_{n+1}) - E(\bar{L}_n) < 0$ only if $E(\hat{L}_{n+1}) > \theta_n(\varepsilon)$, where

$$\begin{aligned}\theta_n(\varepsilon) \equiv& 1 - \frac{2}{|Z_n|} - E\left(\frac{|\boldsymbol{S}_n^i(\varepsilon)|}{|Z_n|^2}\right)\\ &- \frac{1}{\varepsilon}\left[\frac{|Z_{n+1}|^2 - |Z_n|^2}{s(n+1)-1} + \frac{1}{s(n+1)-1}\sum_{s=s(n)}^{s(n+1)-1} E\left(I_s\right)\right]^{1/A^{I-i}}.\end{aligned} \tag{2.18}$$

Consider now the following. First, if $\alpha > 1$, then

$$\frac{|Z_{n+1}|^2 - |Z_n|^2}{s(n+1)-1} \longrightarrow 0.$$

Second, for $i > 1$, if L_s^j converges to one in probability, for each $j = 1, \ldots, i-1$, then I_s tends to zero in probability, and thus (recall that I_s is identically zero for $i = 1$)

$$\sum_{s=s(n)}^{s(n+1)-1} E(I_s)/[s(n+1)-1] \longrightarrow 0.$$

This means that $\theta_n(\varepsilon)$ tends to $1 - S(\varepsilon)$, where $S(\varepsilon)$ is the almost sure limiting value of $\boldsymbol{S}_n^i(\varepsilon)/|Z_n|^2$ (from Lemma 2.8). Notice then that $S(\varepsilon)$ tends to zero as ε tends to zero (Lemma 2.8 again).

Going back to equation (2.11) we see that $L_{s(n+1)}^i \geq L_{s(n)}^i \cdot |Z_{n+1}|^2/|Z_n|^2$, and thus $\liminf[E(L_{s(n+1)}^i) - E(L_{s(n)}^i)] \geq 0$. An application of Lemma 2.9 now gives the result. ■

We are now in a position to prove Lemma 2.7.

Proof of Lemma 2.7: Fix $\alpha > 1$. The proof is by induction. Consider first $i > 1$. The induction hypothesis is: If L_s^j tends to one in probability, $j = 1, \ldots, i-1$, then L_s^i converges to one in probability.

In order to prove the induction claim, it suffices to prove that the subsequence $\bar{L}_n$ converges to one in probability (Lemma 2.10).

Toward that end, first write $Y_{n,t}(\varepsilon) = 1 - L_{n,t}^i - |\boldsymbol{S}_n^i(\varepsilon)|/|Z_n|^2$ (for role i). Write (with equation (2.15) in mind) $\bar{Y}_n(\varepsilon) = \left[\varepsilon \cdot \max\left\{0, E(Y_{n\kappa(n)}(\varepsilon))\right\}\right]^{A^{I-i}}$, and

$$\bar{X}_n = \frac{2 \cdot |Z_n|}{\kappa(n)-1} + \frac{1}{\kappa(n)-1}\sum_{t=1}^{\kappa(n)-1} E(I_s).$$

Consider dates s, m, such that $s > m$. Summing the terms of equation (2.15) from m to s gives,

$$\begin{aligned} E(\hat{L}_s) - E(\hat{L}_m) &= \sum_{n=m}^{s-1} [E(\hat{L}_{n+1}) - E(\bar{L}_n)] \\ &\geq \sum_{n=m}^{s-1} \frac{\kappa(n)-1}{|Z_n|^2} \left[\bar{Y}_n(\varepsilon) - \bar{X}_n\right]. \end{aligned} \tag{2.19}$$

Lemma 2.11 gives that the sequence $\{E(\bar{L}_n)\}$ converges, and therefore implies that $\lim_{m\to\infty} \sup_{s\geq m}[E(\hat{L}_s) - E(\hat{L}_m)] = 0$. When $\alpha > 1$ the series $\sum_{n=m}^{s}[\kappa(n) - 1]/|Z_n|^2$ diverges to infinity as s tends to infinity.[50] It follows then from (2.19) that $\liminf[\bar{Y}_n(\varepsilon) - \bar{X}_n] \leq 0$ for each $\varepsilon > 0$. Observe now that if $L^j_{n,t} \longrightarrow 1$ in probability, then I_s tends to zero in probability, and thus $E(I_s)$ tends to zero in probability. The Cesaro means, $\sum_{t=1}^{\kappa(n)-1} E(I_s)/[\kappa(n) - 1]$ thus tend to zero also, and hence $\bar{X}_n$ tends to zero. It follows that $\liminf \bar{Y}_n(\varepsilon) = 0$, for all $\varepsilon > 0$. In view of the definition of $\bar{Y}_n(\varepsilon)$, this implies $\liminf Y_{n,t}(\varepsilon) = 0$ for all $\varepsilon > 0$, and thus $\liminf[1 - E(\bar{L}_n) - |\boldsymbol{S}^i_n(\varepsilon)|/|Z_n|^2] = [1 - \bar{L} - S(\varepsilon)] = 0$, where $\bar{L}$ is the limiting value of $E(\bar{L}_n)$. Since $S(\varepsilon)$ can be made arbitrarily small by choice of ε, it follows that $\bar{L} = 1$. Since $\bar{L}_n$ is surely bounded above by one, we have that $\bar{L}_n$ converges to one in probability. This completes the proof of the induction claim.

What is needed now to complete the proof is to show that $L^1_{n,t}$ tends to one in probability. This follows by applying the previous arguments in establishing the convergence of $L^i_{n,t}$ for $i > 1$. In particular, in the definition of $\bar{X}_n$ set $I_s = 0$, and then proceed as above. ■

2.6.2 The *SPE-ToP* Strategy Dominates if $\alpha \in (1, 3)$

Fix an $i > 1$ throughout. From now on a single subscript will denote the total number of iterations. For example, rather than writing $H_{n,t}$ for the history at iteration t of date n, H_s will be used where $s = \sum_{m=1}^{n-1} \kappa(m) + t$. In this section it will be proved that if the arrival rate of novelty α lies in the range $(1, A^2 - 1)$, and if the fraction of players in role $j = 2, \ldots, i-1$ that are the *SPE-ToP* tends to one in probability, then the fraction of players in role i that is the *SPE-ToP* will tend to one in probability also. Consider first some required definitions and results.

[50]That is, that $\lim_{s\to\infty} \sum_{n=1}^{s} \frac{1}{n} = \infty$, and $\sum_{n=m+N}^{s+N} \frac{1}{n} = \sum_{n=m}^{s} \frac{1}{|Z_n|} < \sum_{n=m}^{s}[\kappa(n)-1]/|Z_n|^2$.

Definition 2.10: *The game q is new to i at H_s if no i subgame of q has occurred along H_s. $N_s \in \{0, 1\}$ is such that $N_s = 1$ if and only if the game in iteration s is new to i.*

Lemma 2.12: *Suppose $\alpha < A^i - 1$, for $i = 2, \ldots, I$, then every subgame is new to i in the limit. In particular, $P\{N_s = 1 | H_s\}$ converges to one almost surely.*[51]

Proof: First observe the following. If s is the total number of iterations along $H_s = H_{n,t}$, then $s \leq \sum_{m \leq n} \kappa(m)$. Assumption 2.3, where N is the number of outcomes initially, then gives $s \leq n \cdot (N + n)^{\alpha}$, and hence $s \leq (N + n)^{\alpha+1} = |Z_n|^{\alpha+1}$. Since there are A^{I-i} i role subgames in the fixed game tree, the number of i role subgames that have been encountered along H_s is surely bounded above by $A^{I-i} \cdot |Z_n|^{\alpha+1}$, and therefore the fraction of i subgames encountered previously along H_s is no greater than $A^{I-i} \cdot |Z_n|^{\alpha+1}/|Z_n|^{A^i}$, which clearly converges to zero whenever $\alpha + 1 < A^i$. This establishes the result as the distribution over games at iteration t of date n can be induced by drawing the appropriate number of i subgames uniformly from the A^i-times product of Z_n. ■

Definition 2.11: *The measure induced by F on the full set of games Q is μ, and the measure induced by F on the full set of i role subgames, Q^i, is μ^i.*

Lemma 2.13: *For each strategy r of role i that is not the SPE-ToP strategy there exists a set of games $Q(r)$ with positive measure under μ such that if $q \in Q(r)$ and q is new to i at H_s, then for every subgame q' of q, the choice made by r in q' at H_s is not part of an SPE of q'.*

Proof: It suffices to show that for any alternative r to the *SPE-ToP* there exists a set of i subgames, $Q^i(r)$, with positive measure under μ^i, such that for all $q \in Q^i(r)$, if q is new, then r's choice in q is not part of an *SPE*. [52]

If $r > 1$ is a *ToP* alternative to the *SPE-ToP* this follows by definition. Thus, assume r is a naive strategy. Recall that each naive strategy maps own payoffs to a

[51] If $\alpha > A^i - 1$, then the fraction of games that are new to i converges to 0 in probability. The proof is analogous to that of Lemma 2.7.

[52] Any game that has each i subgame in $Q^i(r)$ will belong to $Q(r)$, and thus $Q(r)$ has measure at least $[\mu^i(Q^i(r))]^{A^{I-i}} > 0$, since there are A^{I-i} subgames of player i in each game and since μ is derived from the A^I-fold independent sampling from F, while μ^i is derived from the A^i-fold independent sampling from F.

fixed choice whenever making a choice in a new subgame. Although this mapping might correspond to an *SPE* in some i subgames, the richness of the set of possible games ensures it does not correspond to an *SPE* choice on a set of i subgames with positive measure under μ^i. To see this fix an action $a \in A$. Suppose i's choice of a is part of an *SPE* in every subgame in a subset $\tilde{Q}$ of i role subgames. Then, for almost every $q \in \tilde{Q}$, lacking a dominant choice for i, the action a can be rendered suboptimal in some q' obtained from q through a re-assignment of the remaining player $j = i-1, \ldots, 1$ payoffs. The result then follows since the set of subgames in $\tilde{Q}$ such that i has no dominant choice has positive measure under μ^i. ■

Definition 2.12: *For each i role strategy $r = 2, \ldots, \bar{R}$, $B_s(r) \in \{0,1\}$ is such that $B_s(r) = 1$ if and only if the game drawn at iteration s is new to i and belongs to $Q(r)$.*

Definition 2.13: *Let $C_s \in \{0,1\}$ be such that $C_s = 1$ if and only if at iteration s some alternative to the SPE-ToP in role i outdoes the SPE-ToP in any i role subgame reached by the i players.*

Definition 2.14: *Q_δ is the set of games such that the absolute difference between any payoffs of the game is at least δ. For each $\xi > 0$, and $\delta > 0$, $D_s(\xi, \delta) \in \{0,1\}$ is such that $D_s(\xi,\delta) = 1$ if and only if at iteration s each of the following hold: 1) the game is in Q_δ, 2) the fraction of remaining players after i, $j = 1, \ldots, i-1$, that chooses an SPE in every subgame is at least $1-\xi$, and 3) the SPE-ToPs in role i themselves make an SPE choice at each node they reach.*

Now the last of the preliminary results—

Lemma 2.14: *1) If $\alpha < A^2 - 1$, then $P\{B_s(r) = 1\}$ converges to $\mu(Q(r))$. In addition suppose $\alpha > 1$, and that the fraction of players in roles $j = 1, \ldots, i-1$ that is the SPE-ToP tends to one in probability. Then, 2) C_s tends to zero in probability, and 3) for each $\xi > 0$, $P\{D_s(\xi,\delta) = 1\}$ tends to $\mu(Q_\delta)$.*

Proof: 1) Let $I\{\cdot\}$ denote the indicator function. Recall that $B_s(r)$ is equal to one if and only if the game at iteration s is new to i and in $Q(r)$ (Definition 2.12), and that $N_s = 1$ if and only if the game at iteration s is new to i. Thus, where q_s denotes the game at iteration s, $B_s(r) = N_s \cdot I\{q_s \in Q(r)\}$, and therefore $N_s + I\{q_s \in Q(r)\} - 1 \leq B_s(r) \leq I\{q_s \in Q(r)\}$. Taking expectations gives $E(B_s(r)) \longrightarrow$

$E(I\{q_s \in Q(r)\})$ (i.e., Lemma 2.12 implies $E(N_s) \longrightarrow 1$). Notice that $P\{S\} = E(I\{S\})$ given an event S. Thus, $P\{B_s(r) = 1\} \longrightarrow P\{q_s \in Q(r)\}$. Lemma 2.1 delivers the desired result since it implies $P\{q_s \in Q(r)\}$ converges to $\mu(Q(r))$.

2) If $\alpha > 1$, then Lemma 2.7 implies the *SPE-ToP*s in role i choose an *SPE* of the underlying game with probability tending to one. By hypothesis, the remaining players also choose an *SPE* with probability tending one. Thus, in the long run, the *SPE-ToP*s at role i choose optimally.

3) Let $\hat{D}_s \in \{0,1\}$ equal one if and only if the *SPE-ToP*s in role i choose an *SPE* of the game drawn at iteration s, and at least $1-\xi$ of the remaining players behave as in an *SPE* of the game. Then, $D_s(\xi,\delta) = I\{q_s \in Q_\delta\} \cdot \hat{D}_s$, and thus $I\{q_s \in Q_\delta\} + \hat{D}_s - 1 \leq D_s(\xi,\delta) \leq I\{q_s \in Q_\delta\}$. Taking expectations gives $E(D_s(\xi,\delta)) \longrightarrow E(I\{q_s \in Q_\delta\})$ (since by hypothesis $E(\hat{D}_s) \longrightarrow 1$). Thus, $P\{D_s(\xi,\delta) = 1\} \longrightarrow P\{q_s \in Q_\delta\}$. Lemma 2.1 gives $P\{q_s \in Q_\delta\} \longrightarrow \mu(Q_\delta)$, and hence $P\{D_s(\xi,\delta) = 1\} \longrightarrow \mu(Q_\delta)$ as claimed. ■

We are now in position to prove the key result of this section.

LEMMA 2.15: *Suppose $\alpha \in (1, A^2-1)$. If the fraction of players in role $j = 2,\dots,i-1$ that is the SPE-ToP tends to one in probability, then the fraction of players in role i that is the SPE-ToM tends to one in probability.*

Proof: It will first be proved that $E(f_s^i(1))$ converges by showing that the sequence $\{E(f_s^i(1))\}$ satisfies the hypotheses imposed on $\{x_s\}$ from Lemma 2.9. With that in mind notice the following, which is implied by Assumption 2.5 (Parts i, and iii). For each $\varepsilon > 0$ and $\delta > 0$ there are positive numbers Δ and ξ such that the following is true for any strategies of i, r and r'. Suppose the fraction of i's using strategy r' exceeds ε at iteration s. Suppose at the same time that 1) strategy r' choses an *SPE* in every subgame reached by the i's, 2) the game is such that the minimal absolute payoff difference between any i payoffs is greater than δ, 3) the proportion of remaining players $j = i-1,\dots,1$ that choose an *SPE* in each subgame is at least $1-\xi$, and 4) the strategy r makes a non-*SPE* choice in every i subgame. Then, 5) the fraction of i players that use strategy r' increases by at least Δ. [53] Notice that the previous facts imply the following. For each $\varepsilon > 0$ there is

[53]That is, given a minimal payoff difference of δ, if ξ is sufficiently small, a large enough fraction of the remaining players choose the unique *SPE* so that the *SPE* choice is optimal for i. Since, by assumption, r deviates from the *SPE* in every reached subgame, the payoff to r is dominated stricty

a triple of positive numbers $(\delta(\varepsilon), \Delta(\varepsilon), \xi(\varepsilon))$ that give the above implications with $\delta = \delta(\varepsilon), \Delta = \Delta(\varepsilon)$, and $\xi = \xi(\varepsilon)$, where $\delta(\varepsilon)$ can be chosen so that $\lim_{\varepsilon\to 0} \delta(\varepsilon) = 0$. [54] In the remainder let $(\delta(\varepsilon), \Delta(\varepsilon), \xi(\varepsilon))$ be as just described.

Next, recall the definitions of $B_s(r)$, and $D_s(\xi, \delta)$ (Definitions 2.12, and 2.14, respectively). For each $\varepsilon > 0$, and strategy $r > 1$, if the fraction of role i players that use strategy r at iteration s is no less than ε, then the fraction of i players that use the *SPE-ToP* strategy increases by at least $\Delta(\varepsilon)$, whenever $D_s(\xi, \delta) \cdot B_s(r) = 1$, for any $\xi \leq \xi(\varepsilon)$, and $\delta \geq \delta(\varepsilon)$ (i.e., $B_s(r) \cdot D_s(\xi, \delta) = 1$ if and only if the 1)-4) above with $r' = 1$, the *SPE-ToP* strategy). Thus, where $I\{\}$ denotes the indicator function, if $B_s(r) \cdot D_s(\xi(\varepsilon), \delta(\varepsilon)) \cdot I\{f^i_s(r) \geq \varepsilon\} = 1$, then $f^i_{s+1}(1) - f^i_s(1) > \Delta(\varepsilon)$. Next, recall that $C_s = 1$ if and only if some alternative to the *SPE-ToP* at role i outdoes the *SPE-ToP* in some subgame at iteration s. Since $f^i_{s+1}(1) - f^i_s(1) \geq -1$, if follows that $f^i_{s+1}(1) - f^i_s(1) \geq -C_s$ (Assumption 2.5 (iii-iv) implies this since $C_s = 1$ if and only if some strategy obtains a higher payoff than does the *SPE-ToP* in some subgame at iteration s). Hence, for each $r = 2, \ldots, \bar{R}$ and $s = 1, 2, \ldots,$

$$\begin{aligned}
&f^i_{s+1}(1) - f^i_s(1)\\
\geq&\Delta(\varepsilon) \cdot B_s(r) \cdot D_s(\xi(\varepsilon), \delta(\varepsilon)) \cdot I\left\{f^i_s(r) \geq \varepsilon\right\} - C_s\\
\geq&\Delta(\varepsilon) \cdot \Big[B_s(r) \cdot I\left\{f^i_s(r) \geq \varepsilon\right\} + D_s(\xi(\varepsilon), \delta(\varepsilon)) - 1\Big] - C_s\\
=&\Delta(\varepsilon) \cdot \Big[I\left\{f^i_s(r) \geq \varepsilon\right\} \cdot [B_s(r) + \mu(Q(r)) - \mu(Q(r))] + D_s(\xi(\varepsilon), \delta(\varepsilon)) - 1\Big] - C_s\\
\geq&\Delta(\varepsilon) \cdot \Big[I\left\{f^i_s(r) \geq \varepsilon\right\} \cdot \mu(Q(r)) - \mid B_s(r) - \mu(Q(r)) \mid + D_s(\xi(\varepsilon), \delta(\varepsilon)) - 1\Big] - C_s.
\end{aligned} \tag{2.20}$$

In the third line of (2.20) we use the fact that $B \cdot D \geq B + D - 1$, for any binary variables $B, D \in \{0, 1\}$. Taking expectations in equation (2.20) gives

$$\begin{aligned}
E\left(f^i_{s+1}(1)\right) - E\left(f^i_s(1)\right) \geq \Delta(\varepsilon) \cdot \Big[P\left\{f^i_s(r) \geq \varepsilon\right\} \cdot \mu(Q(r))\\
- \left|E\left(B_s(r) - \mu(Q(r))\right)\right| + E(D_s(\xi(\varepsilon), \delta(\varepsilon))) - 1\Big] - E(C_s).
\end{aligned} \tag{2.21}$$

by the payoff to r'. Assumption 2.5 then implies an increase in the fraction of the i population that uses r'. Note, however, that even when the average payoff to r' exceeds the payoff to r, the rate at which role i's abandon r in favor of r' is limited by the fraction of i's that use r at iteration s, and hence the requirement that $f^i_s(r) > \varepsilon$.

[54] As asserted initially, Assumption 2.5 gives: For each $\varepsilon > 0$ and $\delta > 0$ there are positive numbers Δ and ξ such that, if 1)-4) above hold, then 5) holds. To obtain the desired $(\delta(\varepsilon), \Delta(\varepsilon), \xi(\varepsilon))$, choose the function $\delta(\varepsilon)$ first, so that $\lim_{\varepsilon\to 0} \epsilon = 0$. Then, choose $\Delta(\varepsilon)$ and $\xi(\varepsilon)$ as required to to make 1)-4) imply 5).

Therefore, $E\left(f_{s+1}^i(1)\right) - E\left(f_s^i(1)\right) < 0$ implies, after rearranging in (2.21), that for each $r' = 2, \ldots, \bar{R}$,

$$P\left\{f_s^i(r') \geq \varepsilon\right\} < \max_{r\geq 2}\left\{\frac{1}{\mu(Q(r))}\cdot\left[1 - E(D_s(\xi(\varepsilon),\delta(\varepsilon))) + \frac{E(C_s)}{\Delta(\varepsilon)} + \left|E\left[\mu(Q(r)) - B_s(r)\right]\right|\right]\right\}. \tag{2.22}$$

Let $\phi_s(\varepsilon)$ denote the value of the maximum in (2.22). Since $E\left(f_{s+1}^i(1)\right) - E\left(f_s^i(1)\right) < 0$ implies

$$E(f_s^i(r)) < \varepsilon\cdot(1-\phi_s(\varepsilon)) + \phi_s(\varepsilon),\;\; r = 2,\ldots,\bar{R}, \tag{2.23}$$

we have

$$\begin{aligned} &E\left(f_{s+1}^i(1)\right) - E\left(f_s^i(1)\right) < 0, \;\text{ only if} \\ &E(f_s^i(1)) \geq 1 - [\bar{R}-1]\cdot[\varepsilon\cdot(1-\phi_s(\varepsilon)) + \phi_s(\varepsilon)]. \end{aligned} \tag{2.24}$$

With (2.24) in mind set $\theta_s(\varepsilon)$ from the statement of Lemma 2.9 to $1-[\bar{R}-1]\cdot[\varepsilon\cdot(1-\phi_s(\varepsilon))+\phi_s(\varepsilon)]$. Lemma 2.14 implies $\phi_s(\varepsilon) \longrightarrow [1-\mu(Q_{\delta(\varepsilon)})]/\min_{r\geq 2}\{\mu(Q(r))\}$. Then, set $\theta(\varepsilon)$ from Lemma 2.9 to

$$1 - [\bar{R}-1]\cdot\left[\varepsilon\cdot\left(1 - \frac{1-\mu(Q_{\delta(\varepsilon)})}{\min_r\{\mu(Q(r))\}}\right) + \frac{1-\mu(Q_{\delta(\varepsilon)})}{\min_{r\geq 2}\{\mu(Q(r))\}}\right],$$

so that $\theta_s(\varepsilon) \longrightarrow \theta(\varepsilon)$. In view of our choice of $\delta(\varepsilon)$, $\theta(\varepsilon)$ thus defined tends to zero as ε approaches zero (since $\lim_{\delta\to 0}\mu(Q_\delta) = 1$, by Assumption 2.2). Next, since $f_{s+1}^i(1) - f_s^i(1) \geq -C_s$, $\liminf[E(f_{s+1}^i(1)) - E(f_s^i(1))] = 0$, and thus Lemma 2.9 gives that $E(f_s^i(1))$ is a convergent sequence.

To see that $f_s^i(1)$ must converge in probability to one, note that Lemma 2.14 implies that the right-hand-side of (2.21) converges to $\Delta(\varepsilon)\cdot[P\left\{f_s^i(r)\geq\varepsilon\right\}\cdot\mu(Q(r)) + \mu(Q_{\delta(\varepsilon)}) - 1]$. Since $E(f_{s+1}^i(1)) - E(f_s^i(1)) \longrightarrow 0$, in view of equation (2.21), it follows that $\limsup[P\left\{f_s^i(r)\geq\varepsilon\right\} + \mu(Q_{\delta(\varepsilon)}) - 1] \leq 0$, for all $\varepsilon > 0$, and $r = 2,\ldots,\bar{R}$.

This gives $\limsup P\left\{f_s^i(r) \geq \varepsilon\right\} \leq 0$, for all $\varepsilon > 0$, and $r = 2, \ldots, \bar{R}$, which establishes the result. ■

Theorem 2.1 now follows by induction.

Chapter 3

Reputation and Third Party Information Provision

3.1 Introduction

This paper is concerned with the third party provision of information in a dynamic model of reputation. As in the standard reputation literature, the focus is on a situation with a long run firm and a sequence of short run consumers. The basic problem faced by consumers is that they have only imperfect information regarding the payoffs of the firm, while the firm has a short run incentive to engage in opportunistic behavior. In addition, consumers are limited in their knowledge of the previous choices of the long run agent, further exacerbating the information asymmetry. [1]

The novel aspect of the paper is the introduction of an intermediary, whose role it is to gather and disclose information about the long run agent. The third party commits to a disclosure policy for conveying information about the past behavior of the firm, and then sells informative signals to buyers as they arrive on the market.

There are several real world scenarios that fit the description here, with participants purchasing information from an outside party, and in which reputation occupies an important role. A salient example is found in the market for personal credit, where credit agencies act as the information providing intermediaries. In

[1] The present model is intended to describe a general type of interaction between a long run agent, and a sequence of short run agents. The terms "consumers" and "firm" provide a concise way of referring to the short run agents, and the long run agent, respectively, but the framework encompasses scenarios besides exchange between a firm and consumers.

the example creditors are analogous to the short run agents in the model, and a borrower is comparable to the long run firm. [2]

As in the model of this paper, lenders know very little about a potential borrower when initially approached for a loan. Credit agencies address this informational asymmetry by providing information to lenders about individuals' borrowing and bill-paying behavior. [3] The disclosure policy adopted by the intermediary will typically consist in this case of a particular way of mapping borrower credit histories to credit ratings.

The accessibility of borrower credit data affects behavior on both sides of the market. Lenders pay for reports in order to screen borrowers, thereby avoiding loans to high risk candidates. On the borrower side, a bad rating restricts future access to credit, increasing the cost of default. The intermediaries thus alleviate the problems of adverse selection and moral hazard inherent in the market. They are, however, more concerned about their own profits than about the welfare of the individuals in their service market.

This paper puts forth a specific model along the lines of the above described situation. The focus of the analysis is then on the information disclosure policy chosen in equilibrium by the intermediary, and on the resulting reputation dynamics. The information provider here is a self-interested monopolist, with profits derived from selling information. The value of information to consumers, however, is endogenous, determined in large part by the responses of market participants to the intermediary's chosen disclosure policy. Consider again the personal credit market for an illumination of the issues that will arise.

In order to profit from selling information, a credit agency must reliably sort borrowers that will repay a loan from those that will default. The greatest value to lenders is yielded by policies that sort perfectly the good borrowers from the bad ones. The problem of the intermediary, however, cannot be reduced to simply adopting an accurate sorting policy. Loan repayment is here endogenous, determined indirectly by the disclosure rule itself.

[2] Although real life creditors operate over a long time horizon, a candidate borrower has private information about his own credit history, and thus one may think of the creditor in a particular borrower-creditor interaction as the short run agent.

[3] Other rating agencies evaluate the creditworthiness of debt issuing firms. However, in those markets the borrower pays to be rated, while in the personal credit market lenders pay for reports on potential borrowers. The paper focuses on the latter case, where the uninformed party pays the intermediary for information.

By way of illustration, suppose the private information of a candidate borrower consists of own opportunity cost of loan repayment. The threshold value of this cost that separates borrowers who will honor a loan from those that will default is determined by the disclosure policy as it generates the reputational incentives of borrowers. Stronger incentives result in a higher threshold, for instance. The strength of these incentives are determined by how credit ratings are meted out, e.g., whether or not default is punished with a persistent bad rating, and whether or not a history of consistent debt repayment is rewarded with a good rating, with high probability. [4] With an eye toward his profits, the intermediary must therefore carefully consider this.

If the reputational incentives induced by credit reports are very strong, imposing a high degree of discipline on borrowers, then information has little value to creditors in a Bayesian equilibrium—even when the signal sorts borrowers very accurately (by opportunity cost of repayment, for example). That is, if both low and high cost candidates engage in good behavior to avoid unfavourable credit ratings, then creditors will prefer to eschew any costly information in favor of simply extending loans to all candidates.

On the other hand, a rating policy that generates weak reputational incentives is also of little value to creditors. This again holds in cases where ratings sort candidates with a high degree of precision. If reputational incentives are weak, then both high and low income borrowers will fail to overcome their short run incentive to act opportunistically. A lender will then reason that there are few opportunities to make profitable loans, and thus that there is little to be gained from updating his prior belief about a borrower's propensity for default.

Disclosure policies that yield the highest revenues to the credit agency are then ones that sort borrowers with a high degree of accuracy, while at the same time generating moderate reputation incentives for borrowers. The first feature delivers value to lenders by reducing uncertainty about the loan candidate (interim uncertainty here refers to the uncertainty remaining about the borrower after observing his credit rating). The latter feature determines the threshold type of borrower separating those that will default from those that will repay. Its role is to introduce a sufficient amount of uncertainty about candidates at the ex-ante stage (before

[4] In the paper, the interpreted meaning of a signal, i.e., a "bad credit rating", or a "good credit rating", arises endogenously in equilibrium.

observing a credit rating) to induce lenders to pay for credit reports in the first place.

This paper characterizes the optimal disclosure policy from the point of view of the information provider. It shows, under reasonable restrictions on the pricing strategy of the intermediary, that the intermediary can guarantee himself the payoff from this disclosure rule. That is, in every equilibrium the intermediary obtains from consumers the highest price a consumer would ever pay for information. In addition, the paper shows that resulting equilibrium outcome is inefficient, in the sense that there are disclosure policies that generate higher social welfare, but these are never adopted by the intermediary.

Consider the setup in greater detail. In each period the long run firm and one of the short run consumers play a two-by-two normal form stage game. One way to interpret the situation is as follows. The consumer chooses whether or not to make an unrecoverable investment. The firm decides whether or not to exert costly effort. The payoffs of the stage game are such that the firm and buyer can together achieve a mutually beneficial surplus, but this requires investment by the consumer and effort by the firm. If the buyer invests but the firm withholds effort, the latter gains at the former's expense. In fact, the one shot stage game has a unique equilibrium in which the consumer does not invest and firm withholds effort. The equilibrium outcome is Pareto dominated by the one in which the consumer invests and the firm exerts effort.

With some positive prior probability the firm is a commitment type that exerts effort in every period regardless of the history. With the residual probability the firm is a strategic type that chooses a best response to its opponents, given its cost of effort.

Consumers are imperfectly informed—the cost of effort to the firm, and whether or not it is the commitment type, is private information. Moreover, a consumer cannot observe the events that transpired prior to his arrival, or even his place in the sequence of buyers. Each short run agent is thus ex-ante identical upon arriving to the market, with ex-ante beliefs about the firm's choice determined solely by inferring from the long run agent's strategy, and his prior belief about the cost of effort to the firm. The limited knowledge of the consumer creates a role for the information providing intermediary.

At the outset of the game, the intermediary commits to an information disclosure policy consisting of a set of signals, and a mapping from histories to probability distributions on the signals. (Histories here are sequences describing the choices of the firm and of the short run agents.) Once thus chosen, the disclosure policy becomes common knowledge among all the agents, and determines a dynamic game of incomplete information between the firm, the consumers, and the intermediary. Notice that the third party can observe only the choices made by the agents. Incomplete information therefore is due to the partial knowledge, of consumers, and of the intermediary, regarding the cost of effort to the firm.

In each period of the game, before the consumer and firm play their stage game, the intermediary makes a price offer to the consumer for information about the previous interactions of the firm. By paying the price, the buyer can avail himself of the informative signal, using it to update his prior belief about the underlying history of the game, and to update as well his belief about the cost of effort to the firm. When electing instead to reject the offer for information, the buyer's choice conditions only on his prior belief about the firm's cost, and about the prevailing history.

The lifetime payoff to the intermediary is the discounted sum of the prices paid by consumers for information. [5] The optimal disclosure policy from his perspective is therefore one that extracts from consumers the highest possible price, given that they are Bayesian rational.

The results given in the paper are the following. First, if firm is the commitment type with a small, but positive, probability, then the intermediary can guarantee himself close to the highest possible price for selling information. [6] Second, the disclosure policy chosen by the intermediary results in inefficiency. That is, there are alternative disclosure rules that deliver higher social welfare.

It is worth clarifying how the results are related to the possibility of the commitment type of firm. A sufficiently small, but positive, probability of such a type delivers uniqueness of the intermediary's equilibrium payoff. In the case where the firm is strategic with probability one, intermediary payoffs are no longer uniquely determined. In particular, there will always be equilibria in which the buyers do

[5] It is assumed that there is no cost incurred by the intermediary by putting in place and executing a disclosure rule.

[6] The payoff of the intermediary will depend on the prior probability the firm is the commitment type. In particular, the smaller is this probability, the closer will be the price obtained by the intermediary to the highest possible price.

not avail themselves of the signal, and thus the intermediary obtains a payoff of zero. [7]

3.1.1 Literature

There is by now a substantial literature on information disclosure in cases where the sender desires to influence the decision of a receiver, e.g., costly signalling (Spence (1973)), cheap talk (Crawford and Sobel (1982)), and auctions (Milgrom and Weber (1982), and Ottaviani and Prat (2001)). There is also a growing literature characterizing optimal disclosure policies when the sender can credibly commit to the policy, and wishes to influence the behavior of the receiver (Rayo and Segal (2010), and Kamenica and Gentzkow (2011)).

The current paper differs from the ones in the above literatures. The sender here is not interested in using information for the purpose of influencing the receiver's choice, but rather derives benefit from the actual sale of information. In this paper, the sender is concerned with the effects of a disclosure policy on the behavior of the long run player, but only because this behavior determines the value of the signal to receivers.

A second important point of departure from the cheap talk and costly signalling literature is that here the sender commits to a disclosure policy at the outset of the game. This is a reasonable assumption in our setting. The sender here is a long run player and can thus credibly reject, from a short run agent, requests for specially tailored disclosure rules.

A related paper (Lizzeri (1999)) shows that an intermediary with monopoly power over information can capture all the surplus by manipulating information disclosures in an appropriate way. In that paper the intermediary contracts with the privately informed agent, rather than with the uninformed party as in the current setup. Moreover, the model there is of a static game, where information disclosure concerns the quality of an item for sale, which is exogenous. In the current setting information disclosure is focused on the past behavior of a long run player, which is endogenous, and itself affected by the disclosure policy through its effects on the firm's long run incentives.

In the reputation literature, the most closely related work is Liu (2011). Liu studies reputation formation in a model where the short run agents are ex-ante

[7] The interpreted meaning of a signal, whether it is a "good" signal, a "bad" signal, or otherwise, arises endogenously in the equilibrium, as in the cheap talk literature.

identical, as they are here, arriving without any knowledge of prior events. The short run agents are boundedly rational, incurring a cost to process the history of play. The cost of information increases as the buyer looks further back in time. In this paper the short run agents pay a third party for information. The central focus here is on the disclosure policy chosen by the intermediary in order to optimally exploit his monopoly over information.

3.2 Model

Consider first a long run firm, $P1$, and a sequence of short run (one period) buyers $P2(n)$, $n = 1, 2, \ldots$. In time period, $n = 1, 2, \ldots$, the firm meets with $P2(n)$. The matched pair play a two-by-two simultaneous move game. The buyer chooses either a trusting action, or a non-trusting action. The firm decides whether to take a cooperative action, or a defecting action. The period payoffs obtained by the agents from the interaction are as in given the following matrix.

***P*2(*n*)**

		Trust	Not
Firm	Coop.	$1-\theta$; u	$-\theta$; 0
	Defect	1 ; -1	0 ; 0

Here, $u > 0$ is the benefit to the consumer from trusting a cooperative firm. A loss of 1 is incurred by the buyer when he trusts but his opponent defects. The value of the buyer's outside option, secured by taking the non-trusting action, is normalized to zero.

With probability $\varepsilon \geq 0$ the firm is a commitment type that chooses to cooperate in every period. With the residual probability the firm is a strategic type that chooses a best response to the strategies of its opponents. Whether the firm is the strategic type or the commitment type is private information. The value of $\theta > 0$, the cost of cooperation to the firm, is also private information. The type of the firm is determined by Nature at date $n = 0$.

In particular, Nature first determines whether the firm is the commitment type, or the strategic type. Then, if the firm is strategic, its cost of cooperation is drawn from the uniform distribution on $[\theta_1, \theta_2]$, where $0 < \theta_1 < \theta_2 < 1$. The set of firm types is thus $T = [\theta_1, \theta_2] \cup \{B\}$, where B is the commitment type.[8] Once determined by Nature, the type of the firm is fixed throughout its tenure.

Notice that in the stage game, taken as a one-time interaction, the choice to defect by the firm strictly dominates cooperation for every strategic type, and thus the outcome $Defect/Not\text{-}trust$ is the unique equilibrium outcome of the one-shot game, whenever the firm is strategic. When the situation is repeated many times other outcomes are possible, but what can be achieved will depend on the long run incentives of the firm, in particular, on how its choices affect payoffs in future interactions with buyers.

If consumers could observe the history of interactions, cooperation by some types of strategic firms could be sustained with the threat of punishment for defection. Recall, however, that each buyer arrives in complete ignorance of the events having transpired, and moreover, is unaware of his place in the sequence of short run agents. These information constraints create a role for the second long run agent, an information providing *intermediary*, to be denoted $P3$. The intermediary collects and stores data on the interactions between the long run agent, and the $P2$'s. In order to acquire information about the past behavior of the firm, a buyer must purchase from $P3$ a signal conveying information about the history of play. How information is generated, and sold, is clarified next.

3.2.1 Information

Let x_n denote the profile of actions chosen by the buyer and the firm in period n, i.e., $x_n \in \{Coop., Defect\} \times \{Trust, Not\text{-}trust\}$. Denote by R_k the set of partial histories of length k. That is, for example, at date $n = 1, 2, \ldots$, the k-length partial history is $(x_{n-k}, \ldots, x_{n-1})$. A *disclosure policy* consists of an integer k, a finite set of signals S, and a mapping $\sigma : R_k \longrightarrow \Delta(S)$. Denote by Σ the set of all disclosure policies.

Notice that the focus above is on policies that generate a finite number of states of the game. That is, given a bound k, and a finite set of signals, S, the product $R_k \times S$ is finite. $R_k \times S$ can be thought of as the set of states of the game, provided strategies are stationary (i.e., the firm's strategy conditions on elements of $R_k \times S$, and buyer strategies condition only on signals).

[8] The uniform distribution is chosen because of its simplicity, but our results do not depend in any special way on this choice.

By restricting attention to finite disclosure policies it is ensured that for every stationary strategy profile there is at least one implied stationary distribution on the states of the game. [9] This is desired in the current context. In particular, it will be assumed, since buyers are ex-ante symmetric, that buyer beliefs in equilibrium correspond to a steady state distribution over the states implied by the stationary strategies. [10]

The game proceeds as follows. In period $n = -1$ the intermediary publicly commits to a disclosure policy, (k, S, σ). His choice determines a dynamic game of incomplete information, $G(k, S, \sigma)$, between $P1$, the $P2$'s, and $P3$. We refer to $G(k, S, \sigma)$ as the subgame induced by the policy (k, S, σ).[11] Imperfect information in $G(k, S, \sigma)$ arises due to the partial knowledge of the buyers, and the intermediary, about the cost of cooperation to the firm.

The extensive form of the overall game itself consists of the initial node at which $P3$ commits to a disclosure policy, and then the continuation subgames, $G(k, S, \sigma)$, for each $(k, S, \sigma) \in \Sigma$.

Within $G(k, S, \sigma)$ each period is divided into three sub-periods, $m = 1, 2, 3$. The subgame unfolds as described next. Nature first determines the firm's type (e.g., commitment type or strategic type, and cost type) and the result is revealed to the firm. Then, in each period $n = 1, 2, \ldots,$:

At $m = 1$: The period n signal, $s_n \in S$, is realized by drawing from the distribution $\sigma(r)$ on S, where r is the k-length partial history in period n. The firm observes the signal. The n-th buyer arrives and the intermediary makes him a price offer τ_n for the opportunity to observe s_n.

At $m = 2$: The buyer decides whether or not to accept the offer from $P3$. If the offer is accepted, a transfer of τ_n is made by $P2(n)$ to $P3$, and the buyer observes s_n. The choice by the buyer to acquire information is not observed by the firm.

At $m = 3$: The firm, and $P2(n)$ play the normal form stage game. The buyer chooses trust or not-trust, and the firm decides whether or not to cooperate.

[9] In any case, the intermediary can obtain the optimal payoff with a disclosure policy that conditions on finite histories.

[10] Other possibilities can arise. In particular, each agent might assign positive prior probability to some finite set of entry dates. The assumption that buyers believe their entry occurs in the steady state results in a more parsimonious model and cleaner results.

[11] Although this is not a subgame in the strict sense, it is useful for the present purpose.

The assumption that the firm can condition its choice on the signal is important. The assumption yields a simpler model, and permits a sharp characterization of the equilibrium payoff to the intermediary. It allows, in particular, the firm and buyer to coordinate their choices on the signal. The key effect of this is to bolster the value of the signal to the consumer.

The above assumption is natural in some cases, however, such as in a model of the credit market, where borrowers have access to their own ratings. In any case, relaxing this would not change our results provided the intermediary had the option of revealing the prevailing signal to the firm. The payoff to the intermediary in the resulting model could not be lower than it is in the current one. This is because the intermediary would have the option of revealing the signal to the firm for free, and then negotiating with the buyer under the same terms as in the present context.

The payoff to the n-th short run agent is as given in the above payoff matrix, less τ_n if and only if he accepts $P3$'s offer. The period payoff to the type θ firm is also as given in the payoff matrix. The intermediary's n period payoff is τ_n, if the n-th buyer accepts his offer, and zero otherwise. The firm and the intermediary evaluate future payoffs according to the discount factor $\delta \in (0,1)$.

3.2.2 Strategies and Equilibrium

An equilibrium of the game consists of an assignment of a perfect Bayesian equilibrium to each subgame, $G(k, S, \sigma)$, and then, by backward induction, the optimal choice of disclosure policy by the intermediary at the initial date.

In general, the price offer by the intermediary can itself convey information to a consumer about the unobserved sequence of interactions between the firm and previous buyers. That is, the information provider is privy to the actions of the players, and it is therefore possible for his choice of strategy to condition on this knowledge. Attention, however, is restricted to pricing strategies of the intermediary that do not condition on the history of the game.

The above restriction simplifies the analysis, and greatly reduces the number of equilibria that arise in the model. [12] It is, moreover, an appropriate assumption to make in the current environment. To illustrate this, a useful parallel will be drawn again with the personal credit market.

[12] Unrestricted pricing strategies allow beliefs of consumers about the underlying history to be mapped to price offers in arbitrary ways, yielding a vast multiplicity of equilibria.

A credit agency might service a large market consisting of many creditors and many potential borrowers. The market is decentralized, and one can imagine that borrowers and lenders are matched in every period according to some random process. A uniform fee for an individual credit report is established by the intermediary before the market participants are thrown together. It is unlikely then for a creditor to infer from the price any knowledge about his particular candidate borrower. [13]

Of course, if there is aggregate uncertainty about the characteristics of borrowers, the possibility remains for the intermediary to condition his fee on the aggregate state of all the individual credit histories. Prices in such an event could be informative, which is what is hoped to be avoided in the current analysis. However, this latter scenario too is implausible. In particular, creditors will likely already be aware of the relevant aggregate statistics pertaining to borrowers, e.g., the aggregate default rate, average income, average indebtedness, etc. It is thus natural to assume that the information provider in the credit market sets prices in response only to the demand for information (although this demand might itself be determined in part by the aggregate state), and this is the assumption made here.

Our focus is further restricted to stationary equilibria. The stationarity assumption yields greater tractability, and is also reasonable in the current environment with ex-ante symmetric buyers.

Consider a disclosure policy σ and its induced subgame, $G(k, S, \sigma)$. Buyers adopt a common strategy in $G(k, S, \sigma)$ consisting of two functions. The first, $\beta_1(\tau)$, gives the probability $P2$ accepts the offer of τ from the intermediary. The second, $\beta_2 : S \cup \{\emptyset\} \longrightarrow [0, 1]$, gives the probability $P2$ takes the trusting action upon observing $s \in S$—in an abuse of notation, $\beta_2(\emptyset)$ is the probability with which the buyer makes the trusting choice after rejecting the offer to observe the signal. [14] A stationary strategy for the type θ firm in $G(k, S, \sigma)$ is a function $\alpha_\theta : R_k \times S \longrightarrow [0, 1]$, where $\alpha_\theta(r, s)$ gives the probability the firm makes the cooperative choice at r, after the observing the signal $s \in S$. Since the $P2$'s use a common strategy, and

[13] Although the model here postulates one long run agent (the firm, which is analogous to a long run borrower in the credit market example) the setup can easily be extended to one with many firms (many long run borrowers). Under reasonable assumptions such an extended model gives results similar to the present ones. In particular, the required assumptions are the following. Firm strategies should be independent of one another's private histories, and symmetric for each firm given cost of effort. The disclosure policy gives signals that are independent across firms conditional on each firm's private history, and that depend only on the histories, and not, for example, on the identity of the firm.

[14] Notice that the β_2 component of the buyers' strategy does not condition on the price offer by $P3$.

also given our stationarity assumptions, it is without loss of generality to restrict attention to $P3$ strategies consisting of a fixed price offer $\tau \in \mathbb{R}_+$ in every period $n = 1, 2, \ldots$.

In order to complete the description of an equilibrium buyer beliefs are next specified.

Recall that buyers are ex-ante symmetric, arriving in ignorance of the number of periods elapsed, and of any events transpired previously. We assume that equilibrium beliefs coincide with the steady state beliefs induced by the strategies in the game. That is, each buyer updates his beliefs, given new information, according to Bayes' rule as if assigning probability one to having arrived in the steady state (relevant discussions are contained in Liu (2011), and Rosenthal (1979)).

Such steady state beliefs are well defined in the model. For each firm type $t \in [\theta_1, \theta_2] \cup \{B\}$, the stationary profile of strategies $(\alpha_t, \beta = (\beta_1, \beta_2), \tau)$ induces a Markov chain on the states of the game, $R_k \times S$, with transition matrix, $M(t)$, say. An invariant distribution $\hat{\mu}(\cdot|t) \in \Delta(R_k \times S)$, satisfying $M(t) \cdot \hat{\mu}(\cdot|t) = \hat{\mu}(\cdot|t)$ is sure to exist for each type t. More is needed besides $\hat{\mu}$, however, in order to completely determine an equilibrium. In particular, buyer behavior must be specified following a signal that does not occur along the equilibrium path (i.e., a signal such that $\hat{\mu}(r, s|t) = 0$ for all firm types, t, and all histories, $r \in R_k$).

Buyer beliefs in an equilibrium will then consist of a family of conditional pdf's $\mu(r, t|s)$, $s \in S$, and $(r, t) \in R_k \times T$, that are consistent with the strategies, in the sense that they can be derived from the steady state distributions, $\hat{\mu}$, according to Bayes' rule. (Recall that T denotes the set of possible firm types, $[\theta_1, \theta_2] \cup \{B\}$.) Consider now the following.

A price-taking perfect Bayesian equilibrium (*PT-PBE*) of $G(k, S, \sigma)$ consists of a profile of stationary strategies $(\{\alpha_\theta\}, \beta, \tau)$ for $G(k, S, \sigma)$, as above described (here $\beta = (\beta_1, \beta_2)$), and a family of conditional pdf's, $\mu(r, t|s)$, $(r, t) \in R_k \times T$ (one for each $s \in S$) such that:

— The profile of strategies are best responses to each other in $G(k, S, \sigma)$, when buyers maintain the beliefs μ.

— For each $s \in S$, the pdf $\mu(r, t|s)$ can be derived using Bayes' rule from the distributions $\hat{\mu}(\cdot|t)$, $t \in T$, where for each $t \in T$, $\hat{\mu}(\cdot|t)$ is an invariant distribution on $R_k \times S$ consistent with the stationary profile (α_t, β, τ).

These are referred to as price taking equilibria for the following reason. Beliefs, μ, do not vary with price offers by the firm. Consequently, a buyer evaluates each price offer by the intermediary taking into account only his ex-ante expected value for the informative signal.

The letter ξ will throughout be used to denote strategy and belief profiles. Our equilibrium notion is then—

DEFINITION 3.1: *An equilibrium for the game consists of an assignment of a PT-PBE to each $G(k, S, \sigma)$, say $\xi(k, S, \sigma)$, and then the initial choice of disclosure policy, by the intermediary, that is a best response to the continuation PT-PBE strategies.*

Without loss of generality behavior in the initial $k-1$ periods can be left unspecified. That is, the steady state occurs long after the initial periods of the game. Therefore, the behavior in these early periods can affect the equilibrium in the long run only via any restrictions it might impose on the steady state beliefs that can emerge. Notice, however, that given an equilibrium as above described, it is possible to specify the disclosure policy for the initial histories of length $j = 0, 1, \ldots, k-1$, and then equilibrium behavior for the firm at these early histories, so that the disclosure policy in these early periods does not have an effect on the long run. Hence given the strategy profile, and the disclosure policy, the long run distribution will converge to the steady state equilibrium beliefs specified in the putative equilibrium.[15]

3.3 Main Result

This section presents the main results of the paper. In anticipation, we first make some key observations.

Consider $P\{\cdot\}$, an arbitrary joint distribution over a set of signals S, and the actions of the firm. Suppose for the moment that $P\{\cdot\}$ corresponds to a buyer's ex-ante belief about the joint distribution over realized signals and firm choices (ex-ante refers to the buyer's situation prior to acquiring a signal from the intermediary). These need not arise in an equilibrium, but are merely hypothesized beliefs for the purpose of deriving an upper bound on the value of information to buyers.

[15] What is important here is that in these early periods it should not be possible, along the equilibrium path, to reach an absorbing state, i.e., a situation in which some particular signal occurs every period along the equilibrium path for the remainder of the game.

We abuse notation slightly in the remainder: $P\{a\}$ refers to the probability the firm will take action $a \in \{(C)oop., (D)efect\}$. Conditional probabilities derived from $P\{\cdot\}$ are described analogously.

Taking the above beliefs as given, the expected payoff to the buyer from choosing trust, conditional on observing $s \in S$, is $u \cdot P\{C|s\} - P\{D|s\}$. With this in mind, write

$$S^* = \{s \in S : u \cdot P\{C|s\} - P\{D|s\} > 0\}.$$

Notice that the rational course of action for the buyer is to choose trust after observing a signal in S^*. Therefore, the ex-ante expected value to the buyer of a signal from $P\{\cdot\}$ is

$$(u \cdot P\{C|S^*\} - P\{D|S^*\}) \cdot P\{S^*\} \tag{3.1}$$

(S^* might be empty, and in that case the value is zero to the buyer). This expression is bounded above by $u \cdot P\{C\}$. Hence, the buyer is willing to pay τ to observe a signal from $P\{\cdot\}$ only if

$$u \cdot P\{C\} - \tau \geq 0, \tag{3.2}$$

since otherwise paying τ for the signal is ex-ante strictly worse than simply making the non-trusting choice.

Notice also that the buyer is willing to pay τ only if

$$u \cdot P\{C\} - \tau \quad \geq \quad u \cdot P\{C\} - P\{D\}. \tag{3.3}$$

To see this, notice that the right hand side is the ex-ante expected payoff to the buyer from taking the trusting action. If the inequality does not hold, then making the uninformed choice to trust is strictly better than paying τ for the signal.

Combining equations (3.2) and (3.3) gives that if a Bayesian rational buyer is willing to pay τ for a signal from $P\{\cdot\}$, then

$$\tau \leq \min\{u \cdot P\{C\}\ ,\ P\{D\}\}. \tag{3.4}$$

The expression on the right hand side of (3.4) achieves its maximum when $P\{\cdot\}$ is such that $P\{C\} = 1/(1+u)$. It thus follows that the fee paid for information by a rational buyer cannot exceed

$$\tau^* = \frac{u}{1+u}. \tag{3.5}$$

One last assumption is needed. It imposes a sufficient degree of patience so that every strategic type of firm can be induced to cooperate in some equilibrium.

ASSUMPTION 3.1: *The discount rate, δ, and the highest cost of effort type, θ_2, are such that $\delta \geq \theta_2$.*

The first result characterizes the equilibrium payoffs for sufficiently small ε. A complete proof is given in Section 3.5.

THEOREM 3.1: *For each $\eta \in (0, \hat{\eta}]$, where $\hat{\eta}$ is some positive number, there exists a $\varepsilon > 0$ such that the following is true. Suppose the firm is the commitment type with prior probability ε. Then, claims (i) through (iv) below hold along the PT-PBE path of the subgame induced by the equilibrium choice of disclosure policy.*

(i) The period payoff to the intermediary is bounded below by $\tau^* - \eta$.

(ii) The ex-ante belief of the buyer that the firm will choose cooperation, say ϕ, satisfies

$$\phi \in \left[\frac{1-\eta}{1+u}, \frac{1+\eta}{1+u}\right].$$

(iii) Whenever the buyer observes the signal, the probability he correctly anticipates the choice the firm is at least $1 - \eta$.

(iv) The buyer always pays for the signal, his ex-ante payoff is no greater than η.

Implication (i) of the theorem is that if the prior probability of a commitment type is small (but positive) then the intermediary can guarantee himself a period payoff that is close to the highest possible price that a Bayesian buyer would ever pay for information (see the derivation of τ^* culminating in (3.5)).

It is worth mentioning that when the probability of a commitment type is exactly zero, τ^* can be attained by the intermediary in an equilibrium, but it can no longer be guaranteed. In particular, in the event $\varepsilon = 0$ there is a *PT-PBE* of $G(k, S, \sigma)$, for every (k, S, σ), in which every strategic firm chooses to defect in every period, and the buyer rejects every offer for information, and simply makes the uninformed choice not to trust the firm. When there is a small possibility that the firm is the commitment type the intermediary can ensure this will not occur with the appropriate choice of disclosure policy.

Implication (iii) is that when ε is small, the disclosure policy gives very informative signals in the following sense. When conditioning his choice on the signal, a buyer makes the appropriate choice against the firm with probability that is close to one. The appropriate choice here is trust against a firm if and only if the firm chooses cooperation. Implication (iii) is related to (i), in that a highly informative signal is required in order for the buyer to be induced to pay $\tau^* - \eta$ for information.

Item (ii) gives an upper bound on the probability of cooperation in the equilibrium. It is important to mention that this particular feature of the equilibrium results in economic inefficiency. In particular, disclosure policies exist that can improve on the ex-ante expected surplus generated by the equilibrium of Theorem 3.1.[16] In fact, the following claim is proved in Section 3.5.

CLAIM 3.1: *There exists a disclosure policy, (k, S, σ), such that $G(k, S, \sigma)$ has a PT-PBE in which the buyer chooses trust in every period, and in every period this is met with cooperation by the firm.*

The policy from Claim 3.1 eliminates all uncertainty and achieves the highest possible surplus. (Compare this to (ii) from Theorem 3.1).

Recall here that each disclosure policy, (k, S, σ), initiates a subgame, $G(k, S, \sigma)$, between the firm, the buyers, and the intermediary. Behavior in each of these must be consistent with some *PT-PBE* strategies within the subgame. The overall game is initiated by the intermediary's choice of disclosure policy. The point made by Claim 3.1 is that there is some $G(k, S, \sigma)$ in which the *PT-PBE* of the subgame delivers the first-best outcome. The problem, however, is that this efficient policy will never be adopted by a self-interested intermediary.

Although the setup here imposes some highly specific assumptions, the inefficiency of the adopted disclosure policy will hold more generally, as long as the intermediary has a monopoly on the sale of information. Notice in particular, that the buyer would never pay a positive price for information in the efficient situation described in Claim 3.1.

Perturbations of the model might result in noisier characterizations than the one given in Theorem 3.1. However, the basic problem of the intermediary is the

[16]The ex-ante expected surplus is the expected social surplus given the strategies, and the prior probability distribution over the types of firms.

same in any case. That is, in order to extract a high price for information the intermediary has available two basic channels. He can provide accurate signals (Theorem 3.1, item (iii)), and he can introduce excess uncertainty at the ex-ante stage (Theorem 3.1, item (ii)). The degree to which he can do one or the other will be constrained by the particulars of the environment.

3.3.1 The Intermediary Optimal Disclosure Policy

Let us consider more closely the disclosure policy delivering the results enumerated in Theorem 3.1. A good place to begin is with the willingness of a buyer to pay for information.

Notice first that the ex-ante expected value of information to the buyer (equation (3.1)) is determined by his ex-ante belief that he will avoid a mistake by observing the signal. Here, a “mistake” refers to one of two possibilities: 1) the buyer chooses to trust against a firm that defects, or 2) the buyer withholds trust against a firm that cooperates.

Then, observe that for a given ex-ante expected value of information, the buyer's willingness to pay for the signal is greater the lower is his ex-ante expected payoff from choosing trust (equation (3.3)). The reason for this is that paying for the signal must be at least as good as simply making the uninformed choice to trust.

Willingness to pay, however, is greatest when the buyer's ex-ante expected payoff from trusting is exactly zero—lowering the ex-ante payoff past this point depresses the value of information to the buyer. (See the discussion before and after equation (3.4).)

The optimal disclosure policy from the perspective of the intermediary is hence one that accomplishes two ends. It minimizes the probability of mistakes, thus maximizing the value of information to the buyer. At the same time it makes the buyer ex-ante indifferent between trusting and not trusting. In order to attain this indifference condition, the intermediary suppresses the ex-ante likelihood that the firm will choose cooperation. The end result is inefficiency.

What type of disclosure policies are adopted in the equilibrium by the intermediary, hence delivering the outcomes described in the main theorem? Consideration of a simple policy with two signals will suffice, a *good rating* and a *bad rating*. [17]

[17] There is certainly more than one disclosure policy that gives the intermediary the highest payoff.

Let us consider the limiting case in which the firm is a strategic type with probability one ($\varepsilon = 0$).

Given some fixed k, say that the firm has a *good history* if along the k-length partial history there are no instances of trust met with defection, and at least one instance of trust met with cooperation. The firm has a *bad history* otherwise. [18] Now consider a disclosure policy with the following features, given a partial history of length k.

I. *A firm with a bad history is given a bad rating with probability one.*

That is, the policy punishes defection by the firm against a trusting buyer, but punishes also firms that fail to realize a surplus with buyers.

Suppose, for the sake of the argument, that buyers choose trust if and only if the firm has a good rating. The set of k-length histories with no instances of $Coop/Trust$. is then an absorbing set, in the sense that if the firm reaches a history in the set he stays forever within it, receiving a persistent bad rating.

Notice, moreover, that upon defecting on a trusting buyer, the firm gets shut out of the market permanently. Hence, in the steady state only a firm that has never cheated a buyer will obtain a good rating. The stationarity assumption implies that such firms, moreover, meet trust with cooperation at all histories occurring along the path in the steady state. Hence, in the steady state after observing the good rating the buyer knows the firm will make the cooperative choice.

Recall that the rate of discount δ is greater than θ_2, the highest possible cost of effort to the firm (Assumption 3.1). Now consider the following.

If the policy assigns a good rating to a firm that has a good history with probability one, then every type of firm will cooperate in every period (this follows from Assumption 3.1). The value of the signal to buyers would in this case be zero. In order to extract a positive price from the buyer the intermediary will thus introduce some noise in the ratings. This is the second feature of the intermediary optimal policy.

II. *Suppose the firm has a good history, with an instance of $Coop./Trust$ having occurred less than k periods in the past. Then, for some $\rho \in (0, 1)$ the firm gets a*

[18] This is analogous to situation in the credit market in which borrowers with no credit, as well as those with bad credit, are considered to have unfavorable credit histories.

good rating with probability $1 - \rho$*, and a bad rating with probability* ρ*.*

Since a firm with a good history can get a bad rating, a well-behaved firm might inadvertently lapse into a persistent bad rating regime. This would destroy the value of the signal to buyers since in the stationary limit every type of firm would have eventually reached this absorbing state. In order to prevent this, the policy has the following attribute.

III. *A firm with a good history, with the last instance of* $Coop./Trust$ *having occurred* k *periods ago is given a good rating with probability one.*

Now consider the following. Given that buyers choose trust if and only if the firm has a good rating, the choice of k, and ρ (in item II above) determines a threshold type of firm satisfying the following. For every firm with cost of effort less than the threshold, it is a best response to cooperate when a good rating is drawn. For every remaining type of firm it is a best response to defect in every period. In the steady state the probability that the buyer trusts a defecting firm will thus be zero.

Notice that there is no benefit to the firm from choosing cooperation when given a bad rating, even if it has a good history. Therefore, by conditioning on the signal, the probability that a buyer chooses not-trust when the firm in fact cooperates is also zero. This aspect of the equilibrium relies crucially on the assumption that the firm can perfectly condition its choice on the signal. If the firm's observation of the signal were noisy, for example, such a high degree of coordination could not be possible.

Given that the signal reduces the probability of mistakes to zero, the proposed strategy of the buyer, in the manner it conditions on the signal, is therefore a best response to the given firm strategies. In order to obtain the maximal payoff for himself, the intermediary simply chooses ρ in item II so that the resulting threshold type of firm implies an ex-ante probability of cooperation equal to $1/(1+u)$. In the *PT-PBE* the buyer pays $u/(1+u)$ for the signal in every period with probability one.

Notice that in the intermediary optimal policy the buyer and firm never miscoordinate. In particular, buyers never suffer a loss from trusting a defecting firm. Nevertheless the outcome is inefficient. In the equilibrium there is excess defection by the long run agent, brought about by giving bad ratings to good firms, and thus a failure to realize some surplus.

The proof of the main theorem uses a disclosure policy with the above three features. It must be slightly modified, however, in order to accommodate the possibility that the firm is the commitment type. Moreover, it will have to be proved that when ε is positive the payoff to the intermediary is uniquely determined .

3.4 Conclusion and Extensions

This paper studies the third party provision of information in a dynamic reputation model. Information is provided to consumers by a profit maximizing intermediary with monopoly access to information about a long run firm. The choice of information disclosure policy by the intermediary determines the incentives of the long run firm. Two conclusions are drawn.

First, it is shown that the intermediary can alleviate the inefficiencies arising from the asymmetric information and moral hazard in the environment. There are limits, however, on what can be achieved. In particular, the disclosure policy chosen by the intermediary introduces welfare diminishing noise, resulting in bad signals drawn with positive probability for well-behaved firms. Noise is used by the intermediary in order to bolster the value to consumers of the signal by generating ex-ante uncertainty.

The paper studies the steady state of a model in which the short run agents believe they arrive to the market in the long run steady state. Focusing on the steady state greatly simplifies the analysis and allows sharp results to be obtained. The basic flavor of the insights should carry over to richer environments, however, provided the intermediary has monopoly power, and commits to a disclosure rule. The specifics of the welfare reducing noise might vary with perturbations of the model. For example, in some other environment it might be optimal for the intermediary to have bad firms draw occasional good signals.

We assume in our model that the long run firm can condition its choice on the signal. This allows the intermediary to coordinate the decisions of the firm and consumers. Relaxing this assumption would lead to noisier results but there is no reason to believe it would change them drastically. The assumption seems reasonable, in any case. In the personal credit market, for example, a borrower can observe his credit rating before applying for a loan.

The existence of commitment types delivers the uniqueness of equilibria in the model. Without a commitment type, for every disclosure policy, there is an equilibrium of the game in which the value of information to the buyer is zero. In such an equilibrium a strategic firm never cooperates, the offer to observe the signal is rejected by the buyer, and the buyer takes the non-trusting action in every period.

The price taking assumption seems natural in the current environment. Every buyer's ex-ante value for the signal is the same. If buyers are unwilling to accept from the intermediary a price offer below this value, the offer must convey information about the firm, hence altering the buyer's assessment of the signal's value. However, the intermediary has no incentive to alter the buyer's ex-ante belief with a deviating offer. All that matters to the intermediary is how much the buyer is willing to pay for information.

The model can be extended in several interesting directions.

In the optimal disclosure rule here, once the firm cheats a buyer it draws persistently bad signals, and hence a zero continuation payoff. Some interesting issues arise when the firm has a positively valued outside option that can be exercised by shutting down its current enterprise. The intermediary relies on the firm's continued operation in order to generate his own profits. If the firm has an outside option then the intermediary might prefer to forgive occasional lapses in good behavior. Clearly, this will lower the value of the signal. However, an additional complication arises. A Bayesian buyer can make inferences about the tenure of the firm conditioning on the knowledge that the firm is operating. This additional information can further depress the value of the intermediary's information. In particular, the buyer might reason that the firm is a low cost type, since it has stayed in business at least up until the current period.

Another interesting extension would be to permit firm types to change exogenously. Borrower characteristics change over time, for example. An individual might frequently miss credit payments as a poor college student. Once the individual is gainfully employed his default risk diminishes, all things equal. This is perhaps why credit agencies give defaulting borrowers a fresh start after 7 years. In this extension, the interplay between the persistence of the firm's type and the potential value of information to the buyer also merits attention.

Another direction inviting further study involves perturbations to the ex-ante information of buyers. One possibility is to allow buyers to have private information about their own payoffs. The problem of the intermediary is then one of mechanism design—to provide the appropriate menu of disclosure policies, and prices, in order to screen the buyers of information. The issue is complicated, of course, by the required consideration of the effect policies will have on the reputational incentives of the long run agent. Along this vein is also the case in which the buyer observes a private signal about the firm.

3.5 Proofs of the Main Results

3.5.1 Proof of Claim 3.1

The proof is by construction. First consider the appropriate disclosure policy. Fix any $k > 1$, and then set $S = \{s^*, s_*\}$, for some arbitrary signals, s^*, and s_*. Define $\sigma : R_k \longrightarrow S$ in the following way. $\sigma(r) = ((s^*, s_*)\,,\,(1, 0))$ if r contains only instances of $Coop/Trust$. $\sigma(r) = ((s^*, s_*)\,,\,(0, 1))$ otherwise.

Next consider the required equilibrium strategies for $G(k, S, \sigma)$. The intermediary charges zero for information in every period. The buyer observes the signal with probability one if the price offer is zero, and rejects the offer if it is positive. He chooses trust after observing the signal if and only if he observes s^*. In the event he rejects the offer for the signal, he chooses trust outright. Now for the firm strategies. Let r^* denote the k-length history with all instances of $Coop/Trust$. For each $\theta \in [\theta_1, \theta_2]$, let $\alpha_\theta(r, s) = Coop.$ if and only if $s = s^*$, and r is r^*.

The strategy of each strategic type of firm is a best response to the strategy of the buyers. Consider, for example, the strategic firm with cost of effort θ. The payoff to this firm from choosing to cooperate always at (r^*, s^*) is $(1-\theta)/(1-\delta)$. If the firm defects instead it will be shut out permanently from the market, obtaining the bad signal, s_* for the remainder of the game. The deviation payoff to the firm from choosing defection at (r^*, s^*) is hence 1. Assumption 3.1 then ensures that the choice to cooperate is optimal at (r^*, s^*) for the firm, given the remaining players' strategies. Clearly, it is a best response for the firm to defect when the signal is s_*, since whether or not he defects, the buyer chooses $Not\text{-}trust$, and thus the signal in the subsequent period is the bad one, and the next buyer will again choose $Not\text{-}trust$.

Next, define steady state beliefs, μ. Let $\mu(r^*|s^*) = 1$, and $\mu(r_*|s_*) = 1$, where r_* consists of all *Not-trust/Defect* entries. Let also $\mu(r^*, s^*|t) = 1$ for all $t \in [\theta_1, \theta_2] \cup \{B\}$. The above strategy of the buyers is a best response to the firm strategy, given these beliefs. Notice also that rejection of a positive offer from the intermediary is a best response for the buyer since information has to value in the equilibrium. Given appropriately defined equilibrium behavior in the early periods, and the above defined strategies, the beliefs given will coincide with the implied steady state distribution over states. The strategy, and belief profile is clearly an equilibrium of the subgame $G(k, S, \sigma)$. It delivers the required conditions stated in Claim 3.1. The proof of Claim 3.1 is complete.

3.5.2 Part I of Proof of Theorem 3.1

The first step of the proof of Theorem 3.1 is to establish the following result.

LEMMA 3.1: *Claim (i) of Theorem 3.1 implies the claims (ii) through (iv) of the theorem.*

Proof: Assume claim (i) of the Theorem 3.1 holds. Given $\eta > 0$, let $\tilde{\eta}$ be such that

$$\eta = \max\left\{\tilde{\eta}\,,\ \frac{\tilde{\eta}}{u}\,,\ \tilde{\eta}\cdot(1+u)\,,\ \tilde{\eta}\cdot\frac{(2+u)\cdot(1+u)}{u}\right\}.$$

Then, choose $\varepsilon > 0$ such that in the equilibrium of the game the intermediary is guaranteed to receive $\tau \geq \tau^* - \tilde{\eta}$ in every period (notice that this implies $\tau \geq \tau^* - \eta$. Next, let $P\{\cdot\}$ be the distribution corresponding to the buyers' equilibrium beliefs about signals and firm choices. Recall that $P\{a\}$ is used throughout to denote the probability the firm will take action $a \in \{(C)oop., (D)efect\}$. Conditional probabilities derived from $P\{\cdot\}$ are described analogously.

The definition of τ^* (equation (3.5)), and the discussion after it, imply together that

$$\frac{u}{1+u} - \tilde{\eta} \ \leq\ \min\{u\cdot P\{C\}\,, P\{D\}\}\,.$$

Therefore,

$$P\{C\} \leq \frac{1}{1+u} + \tilde{\eta}, \quad \text{and} \quad P\{D\} \geq \frac{u}{1+u} - \tilde{\eta}, \tag{3.6}$$

and, moreover,

$$P\{C\} \geq \frac{1}{1+u} - \frac{\tilde{\eta}}{u}, \quad \text{and} \quad P\{D\} \leq \frac{u}{1+u} + \frac{\tilde{\eta}}{u}. \tag{3.7}$$

The first inequalities in each of (3.6) and (3.7), with the definition of η, give

$$P\{C\} \in \left[\frac{1}{1+u} - \eta \ , \ \frac{1}{1+u} + \eta\right],$$

which is claim (ii) from the main theorem.

Next, recall, the value of information to the buyer (expressed first in (3.1), with S^* defined there too): $(u \cdot P\{C|S^*\} - P\{D|S^*\}) \cdot P\{S^*\}$. Since the buyer is rational, it follows that

$$\frac{u}{1+u} - \tilde{\eta} \leq (u \cdot P\{C|S^*\} - P\{D|S^*\}) \cdot P\{S^*\}.$$

An application of Bayes' theorem then gives

$$\frac{u}{1+u} - \tilde{\eta} \leq u \cdot P\{S^*|C\} \cdot P\{C\} - P\{S^*|D\} \cdot P\{D\}.$$

Using here the bounds expressed in (3.6) yields, after some algebra,

$$\frac{u}{1+u} - \tilde{\eta} \leq \frac{u}{1+u}\left(P\{S^*|C\} - P\{S^*|D\}\right) + \tilde{\eta} \cdot (1+u).$$

Hence,

$$P\{S^*|C\} - P\{S^*|D\} \geq 1 - \tilde{\eta} \cdot \left[\frac{(2+u)\cdot(1+u)}{u}\right].$$

The left hand side is the probability the buyer makes the correct choice when conditioning his choice on the signal. The definition of η gives that this quantity is bounded below by $1 - \eta$. This is implication (iii) of the theorem.

Finally, the bounds in (3.6) imply that

$$u \cdot P\{C\} - P\{D\} \leq \tilde{\eta} \cdot (1+u).$$

The definition of η gives $u \cdot P\{C\} - P\{D\} \leq \eta$, which is claim (iv) from the theorem. ■

3.5.3 Auxiliary Results for Part II of Proof of Theorem 3.1

We proceed next to prove some auxiliary lemmas concerning *PT-PBE* (defined in Section 3.2.2). In the light of Lemma 3.1 the goal ultimately is to prove claim (i) of the main theorem.

The first result here is an immediate consequence of the price taking built into the definition of a *PT-PBE*. It claims that in any price taking equilibrium, if the intermediary makes a positive price offer, then the buyer observes the signal with probability that is either one or zero. The second claim, which is rather obvious, but nonetheless useful, simply states that the buyer pays a positive price for a signal only if it is informative.

LEMMA 3.2: *For all (k, S, σ), in any PT-PBE of $G(k, S, \sigma)$, if the equilibrium offer by the intermediary is positive, then the buyer observes the signal with probability that is either one, or zero.*

Proof: Consider a subgame $G(k, S, \sigma)$ with a *PT-PBE* in which $P2$ accepts the positive equilibrium offer, τ, with probability $\beta \in (0, 1)$. The expected period payoff to the intermediary is then $\tau \cdot \beta$. By the definition of a *PT-PBE*, interim and posterior beliefs do not vary with deviation offers by $P3$. Since the buyer is just indifferent between accepting and rejecting τ, it follows that he will accept with probability one any deviating offer $\tau' < \tau$. Hence, the intermediary can improve his period payoff by making an offer $\tau' < \tau$ such that $\tau' > \tau \cdot \beta$. ■

LEMMA 3.3: *Suppose ξ is a PT-PBE of $G(k, S, \sigma)$. Let $U(s)$ denote the equilibrium expected payoff to the buyer from choosing trust, conditional on having observed $s \in S$. Let $P\{\cdot\}$ denote the joint distribution over signals and firm choices corresponding to the buyer's equilibrium beliefs about signals and firm choices. If the equilibrium offer by the intermediary is positive and accepted with positive probability by the buyers, then there are signals $s, s' \in S$ occurring along the equilibrium path of ξ such that $U(s) > 0$, and $U(s') < 0$.*

Proof: Fix a $G(k, S, \sigma)$, and consider a *PT-PBE* of $G(k, S, \sigma)$, ξ. Let $U(s)$ be as described in the statement of the lemma given the equilibrium. Suppose the price offer by the intermediary in ξ is positive and accepted with positive probability by the buyer.

Suppose first, by way of contradiction, that $U(s) \geq 0$ for all s occurring in the steady state along the path of ξ. Then, an improving deviation by the buyer is to reject the positive offer by the intermediary, and to simply choose the trusting action. Hence ξ cannot be an equilibrium. On the other hand, if $U(s) \leq 0$ for every s on the path of ξ, it is then an improving deviation for the buyer to reject the equilibrium price offer, and to simply choose the non-trusting action, and hence ξ cannot be an equilibrium. ■

The converse of the above result also holds when S contains two signals.

LEMMA 3.4: *Suppose ξ is a PT-PBE of $G(k, S, \sigma)$, where S contains two signals. Let $U(s)$ denote the equilibrium expected payoff to the buyer from choosing trust, conditional on having observed $s \in S$. Let $P\{\cdot\}$ denote the joint distribution over signals and firm choices corresponding to the buyer's equilibrium beliefs about signals and firm choices. Suppose for one of the signals s, that $U(s) \cdot P\{s\} > 0$, while for the other signal, $U(s') \cdot P\{s'\}$. Then the buyer pays a positive price for the signal in the equilibrium.*

Proof: Let $U(\cdot)$ and $P\{\cdot\}$ be as described in the statement of the lemma. In the *PT-PBE* of $G(k, S, \sigma)$, the expected payoff to the buyer from choosing trust, conditional on s, is then

$$U(s) \equiv u \cdot P\{C|s\} - P\{D|s\}. \tag{3.8}$$

(Recall that $P\{a\}$ is the probability of choice $a \in \{Coop., Defect\}$.)

The price taking assumption concerning ξ implies the following. If for some $s \in S$, and strictly positive $\hat{\tau}$,

$$\begin{aligned} U(s) \cdot P\{s\} - \hat{\tau} &> 0, \text{ and} \\ U(s) \cdot P\{s\} - \hat{\tau} &> U(s) \cdot P\{s\} + U(s') \cdot P\{s'\}, \end{aligned} \tag{3.9}$$

then $P2$ would accept the offer of $\hat{\tau}$ from the intermediary with probability one. This completes the proof. ■

The next result is that for each subgame, $G(k, S, \sigma)$, where S contains two signals, there is essentially at most two *PT-PBE* in which the buyer pays a positive amount for the signal.

LEMMA 3.5: *Suppose* $S = \{s^*, s_*\}$. *For all* (k, S, σ) *the subgame,* $G(k, S, \sigma)$, *has (essentially) at most two PT-PBE in which the intermediary makes a positive price offer that is accepted with positive probability by the buyers.*

Proof: First some notation: Fix a subgame, $G(k, S, \sigma)$, and consider a price taking equilibrium, ξ. Let $s(\xi) \subseteq S \cup \{\emptyset\}$ be such that, in the equilibrium ξ, the buyer chooses trust with probability one after observing a signal in $s(\xi)$. Notice next that, for each $(r, s) \in R_k \times S$, if any strategic firm type $\theta > \theta_1$ chooses to cooperate at (r, s), then there is a cutoff type, $\theta(r, s)(\xi) \in (0, \theta_2)$, such that along the path of ξ every strategic type $\theta < \theta(r, s)(\xi)$ makes the cooperative choice with probability one at (r, s), and any $\theta > \theta(r, s)(\xi)$ chooses to defect with probability one.

Now consider a pair of *PT-PBE* of $G(k, S, \sigma)$, $\xi = (\{\alpha_\theta\}, \beta, \tau, \mu)$, and $\xi' = (\{\alpha'_\theta\}, \beta', \tau', \mu')$, such that the price offers τ, and τ' are positive and accepted with positive probability by the $P2$'s. Observe that Lemma 3.2 implies the short run agent observes the signal with probability one in both ξ, and ξ'. Notice also that Lemma 3.3 ensures $s(\xi)$, and $s(\xi')$, are singleton and non-empty. Hence, the result is proved by showing that if $s(\xi) = s(\xi')$, then ξ and ξ' are essentially the same. With this in mind, in the remainder, suppose $s(\xi) = s(\xi')$.

For each strategic type θ, at every $(r, s) \in R_k \times S$, the continuation payoff to θ under ξ and ξ' must be the same. To see this, suppose instead that the payoff to θ at (r, s) in the equilibrium ξ exceeds his payoff from ξ'. Since $s(\xi) = s(\xi')$, and since the buyers observe the signal with probability one in both equilibria, it follows that along the path of ξ' the θ type can induce the same joint distribution over signal and histories as in ξ by deviating to his strategy in ξ. Clearly continuation payoffs under ξ and ξ' must be equal for every cost type, at every (r, s). This implies $\theta(r, s)(\xi) = \theta(r, s)(\xi')$ for all $(r, s) \in R_k \times S$, and thus firm behavior in ξ and ξ' is the same except perhaps for the set of strategic firm types $\{\theta(r, s)(\xi),\ (r, s) \in R_k \times S\}$, which has measure zero. The price taking assumption then gives that $\tau = \tau'$. Lemmas 3.2 and 3.3 give that buyer behavior is the same in both equilibria, and thus ξ, and ξ' are essentially the same. ■

3.5.4 Intermediary Optimal Disclosure Policies

In view of Lemma 3.1, in order to prove Theorem 3.1, it suffices to show the following. For every η, there is a sufficiently small ε such that if the firm is a commitment type with probability ε, then the intermediary can guarantee himself a period payoff of $\tau^* - \eta$. Toward this end, we next introduce a simple class of disclosure policies that guarantee, in equilibrium, the achievement of the payoff bound. The set of signals in these policies consists of a "good" signal, and a "bad" signal. The good signal is drawn only if the firm has consistently engaged in good behavior—that is, if the firm has chosen to reciprocate, with cooperation, every instance of trust by the consumer. The bad signal is drawn whenever the partial history reveals the firm has been opportunistic, i.e., defected when the buyer trusted. However, there is noise. With some probability, the bad signal is drawn for a firm with a clean record. A formal definition will first require some notation.

Let R_k^* denote the k-length histories along which: 1) there are no instances of $Trust/Defect$, and 2) there is at least one instance of $Trust/Coop$. Denote by R_k^1 the set of k-length histories in R_k^* along which there is exactly one occurrence of $Trust/Coop.$ and it occurs in the last position of the history (i.e., k periods in the past). Let r_k^{nc} denote the k-length history consisting of all $Not\text{-}trust/Coop.$ entries. Consider the following.

Definition 3.2: *The disclosure policy $(k, S, \sigma) \in \Sigma$ is a noisy bad rating policy (NBR policy) if S consists of two signals, say s^*, and s_*, and σ is such that, for some $\rho \in (0,1]$, and $\psi \in [0,1]$,*

$$\begin{aligned} \sigma(r) &= ((s^*,\, s_*)\,,\, (\rho,\, 1-\rho))\,,\ \textit{for each}\ \ r \in R_k^* \setminus R_k^1, \\ \sigma(r) &= ((s^*,\, s_*)\,, (\,1,\, 0))\,,\ \textit{for each}\ r \in R_k^1, \\ \sigma(r_k^{nc}) &= ((s^*,\, s_*)\,, (\,\psi,\, 1-\psi))\,,\ \textit{and} \\ \sigma(r) &= ((s^*,\, s_*)\,, (\,0,\, 1))\,,\ \textit{for each}\ \ r \in R_k \setminus \{R_k^* \cup \{r_k^{nc}\}\}\,. \end{aligned} \tag{3.10}$$

In the remainder the focus is on these *NBR* policies. Notice that each of these can be identified completely with a triple (k, ρ, ψ). We therefore sometimes use the notation $[k, \rho, \psi]$ to denote the *NBR* on R_k with ρ and ψ assuming the role given in Definition 3.2. The notation is extended to the subgame induced by $[k, \rho, \psi]$, denoted now $G[k, \rho, \psi]$. In the interest of conciseness, the letters T, and N will be used to denote the choices to trust, and not trust, respectively, by the buyer.

Similarly, C, and D will represent the choices of the firm to cooperate, and defect, respectively.

The first result here is that for sufficiently small ε there is at most one *PT-PBE* in which the buyers pay a positive amount for information when the subgame has a *NBR* disclosure rule.

LEMMA 3.6: *Consider an NBR disclosure policy,* (k, S, σ)*. If*

$$u \cdot \varepsilon - (1 - \varepsilon) < 0, \tag{3.11}$$

then there is essentially at most one PT-PBE of $G(k, S, \sigma)$ *in which the intermediary proposes a positive price that is accepted by the buyer. Moreover, in any PT-PBE of* $G(k, S, \sigma)$ *where the buyer pays a positive amount for the signal, the buyer chooses* T *with probability one after observing* s^**, and* N *with probability one after observing* s_**.*

Proof: Consider ξ, a *PT-PBE* of $G(k, S, \sigma)$, in which the offer by $P3$ is positive and accepted by the buyer with positive probability. Again, let $s(\xi)$ denote the set of signals such that along ξ the buyer chooses trust with probability one after observing any $s \in s(\xi)$. In the light of Lemma 3.5 and the argument in its proof, the claim can be established by showing that $s(\xi) = s^*$ for any *PT-PBE* of $G(k, S, \sigma)$ in which the buyer pays a positive amount for the signal. With that in mind suppose instead that $s(\xi) \neq s^*$. Lemma 3.3 gives that $s(\xi) = s_*$. In the remainder it is shown this results in a contradiction.

If $s(\xi) = s_*$ it is optimal for a strategic firm to choose defection whenever the signal is s_*. This is because a *NBR* rule maps any history with an instance of $Trust/Defect$ to the signal s_*, and hence the strategic firm can ensure that the buyer chooses trust again in the next period by choosing defection currently. This implies cooperation occurs at s_* only if the firm is the behavioral type, which has prior probability ε. It follows that the expected value to the buyer from choosing to trust when observing s_* is at most $u \cdot \varepsilon - (1 - \varepsilon)$, which is negative (assumed in equation (3.11)), contradicting that ξ is an equilibrium. ■

DEFINITION 3.3: *For each profile of buyer and intermediary stationary strategies* $(\beta = (\beta_1, \beta_2), \tau)$*, define*

$$\begin{aligned}\phi^*(\beta,\tau) =&\beta_2(s^*)\cdot\beta_1(\tau)+\beta_2(\emptyset)\cdot(1-\beta_1(\tau))\\ \phi_*(\beta,\tau) =&\beta_2(s_*)\cdot\beta_1(\tau)+\beta_2(\emptyset)\cdot(1-\beta_1(\tau)).\end{aligned} \tag{3.12}$$

For example, $\phi^*(\beta,\tau)$ is the probability the buyer chooses T when the signal is s^*, and the buyer and intermediary adopt the strategies, β, and τ, respectively.

LEMMA 3.7: *Consider a NBR disclosure policy, (k, S, σ). Suppose*

$$u\cdot\varepsilon-(1-\varepsilon)<0,$$

and that $(\{\alpha_\theta\},\beta,\tau,\mu)$ is a PT-PBE of $G(k,S,\sigma)$. Then,

$$\frac{\phi^*(\beta,\tau)-\phi_*(\beta,\tau)}{1-\delta}>\theta_1. \tag{3.13}$$

Proof: Let $U(s)$ denote the equilibrium expected payoff to the buyer from choosing trust, conditional on having observed $s\in S$. Let $P\{\cdot\}$ denote the joint distribution over signals and firm choices correponding to the buyer's equilibrium beliefs about signals and firm choices.

If τ is positive in the equilibrium and accepted by the buyer with positive probability, then the result is immediate in view of Lemma 3.3. Therefore, suppose, by way of contradiction that

$$\frac{\phi^*(\beta,\tau)-\phi_*(\beta,\tau)}{1-\delta}\le\theta_1, \tag{3.14}$$

and that the buyer rejects any positive offer from $P3$ in the equilibrium.

Notice now that

$$\phi^*(\beta,\tau)-\theta+\delta\cdot\frac{\phi^*(\beta,\tau)}{1-\delta}\le\phi_*(\beta,\tau)+\delta\cdot\frac{\phi_*(\beta,\tau)}{1-\delta}\iff$$

$$\theta\ge\frac{\phi^*(\beta,\tau)-\phi_*(\beta,\tau)}{1-\delta}.$$

The first line implies that for the strategic firm with cost of effort θ, choosing C at s^* is always strictly worse than choosing D. Equation (3.14) then implies that every strategic firm chooses D always at s^*. If every strategic firm chooses D at s^*, then in the steady state s^* is realized with positive probability only if the firm is the commitment type. Therefore, $U(s^*)=u$. Observe next that, if every strategic firm defects at s^*, it must also defect always at s_*. Therefore, where $\{t = B\}$ denotes the event "the firm is the commitment type",

$$\begin{aligned} P\{C|s_*\} &= P\{t = B|s_*\} \\ &= P\{s_*|t = B\} \cdot \frac{\varepsilon}{P\{s_*\}} \\ &\leq \varepsilon. \end{aligned}$$

To see why the last line follows notice that $P\{s_*|t = B\} \leq P\{s_*\}$ because in the equilibrium $P\{s_*|t \neq B\} = 1$ (since every strategic firm defects always). In any case, $U(s_*) \leq u \cdot \varepsilon - (1-\varepsilon)$, which is negative by the hypotheses of the lemma. But $U(s^*) = u$, and $U(s_*) < 0$, which contradicts that the buyer rejects every positive offer (since $P\{s^*\} > 0$ and $P\{s_*\} > 0$ (Lemma 3.4)). ■

Now consider the following definition—

Definition 3.4: *For each disclosure policy $d = (k, S, \sigma)$, and each stationary profile of buyer and intermediary strategies, $(\beta = (\beta_1, \beta_2), \tau)$ let $V(\theta, d, (\beta, \tau))(r)$ denote θ's expected continuation payoff at $r \in R_k$ (before the signal is drawn) in the subgame, $G(k, S, \sigma)$, given that it adopts a stationary best response to (β, τ) in $G(k, S, \sigma)$.*

Given k, enumerate the positions in the k-length histories $j = 1, \ldots, k$, where $j = 1$, for example, corresponds to the most recent event occurring along the partial history.

Lemma 3.8: *Suppose $d = (k, S, \sigma)$ is a NBR disclosure policy. Then, the following holds for every strategic firm type, $\theta \in [\theta_1, \theta_2]$, and every profile of stationary buyer and intermediary strategies, (β, τ).*

- *V1. $V(\theta, d, (\beta, \tau))(r) = V(\theta, d, (\beta, \tau))(r')$ for every $r, r' \in R_k$ with (D, T) in position $j = 1$.*
- *V2. Recall that R_k^* are the k-length histories with at least one occurrence of (C, T), and no occurrences of (D, T). $V(\theta, d, (\beta, \tau))(r) = V(\theta, d, (\beta, \tau))(r')$ for each $r, r' \in R_k^*$ with (C, T) in position $j = 1$.*

Proof: Refer to the definition of an *NBR* policy (Definition 3.2). For V1 notice that in a *NBR* policy when (D, T) occurs, the signal s_* persists at least until the occurrence of the event (D,T) is cleared from the k-length history. V1 then follows from the stationarity assumption on the strategies. The proof of V2 follows similar lines. Every history in R_k^* with (C,T) in position $j=1$ induces the same distribution over signals s_*, and s^*. The stationarity assumption on strategies implies moreover that after every history in R_k^* with (C,T) in position $j=1$, the strategies will induce the same distribution over sequences of signals to follow. ■

LEMMA 3.9: *Consider a fixed k. For every $\eta > 0$ there is a $\hat{\psi}$ (depending on k) such that for each NBR disclosure policy, $d = [k, \rho, \psi]$, with $\psi \in (0, \hat{\psi})$ the following holds, for all stationary buyer and intermediary strategy profiles, (β, τ).*

V3. For all $r, r' \in R_k$ with no instances of (C,T),

$$|V(\theta, d, (\beta, \tau))(r) - V(\theta, d, (\beta, \tau))(r')| < \eta.$$

V4. Recall the definition of $\phi_*(\beta, \tau)$ given in Definition 3.3. For all r with (D,T) in position $j = 1$,

$$\left| V(\theta, d, (\beta, \tau))(r) - \frac{\phi_*(\beta, \tau)}{1-\delta} \right| < \eta.$$

Proof: Notice that all the histories from V4 and V5 in the claim lie outside of R_k^*. Then, note that $[k, \rho, \psi]$ maps every history $r \in R_k \setminus R_k^*$ to s_* with probability one except for the history that has (C,N) in every position. The history with (C,N) in every position is mapped to s_* with probability $1-\psi$. Clearly, the result holds for sufficiently small ψ. ■

3.5.5 Part II of the Proof of Theorem 3.1

In the light of Lemma 3.1, in order to prove Theorem 3.1 it suffices to establish item (i) there. The proof of the following result accomplishes this. Lemmas 3.6-3.9 above are used in the proof.

LEMMA 3.10: *There exists a NBR disclosure policy $[k^*, \rho^*, \psi^*]$ for which the following holds. Suppose the prior probability that the firm is the commitment type* *satisfies*

$$u \cdot \varepsilon - (1-\varepsilon) < 0.$$

Then $G[k^*, \rho^*, \psi^*]$ has a unique PT-PBE. Moreover, in the PT-PBE the buyer pays the intermediary at least $\tau^* - \varepsilon$ in every period for information (τ^* defined in (3.5)).

Proof: The result is established in several parts. First, the desired disclosure policy is constructed. Then its claimed properties are proved.

Part A: The k^* from $[k^*, \rho^*, \psi^*]$ in the lemma.
Let k^* be the smallest integer $k \geq 2$ such that

$$\frac{1}{k} \; < \; \frac{1}{1+u}. \tag{3.15}$$

Part B: The ρ^* from $[k^*, \rho^*, \psi^*]$ in the lemma.
For each strategic firm θ let $V^*(\theta, \rho)(r)$ be the expected payoff to θ at $r \in R^*_{k^*}$ when each of the following hold—

I. The firm chooses C at (r', s^*), and D at (r', s_*) for every $r' \in R^*_{k^*}$.

II. The buyer observes the signal in every period with probability one and chooses C if and only if the signal observed is s^*.

III. The signal is s^* with probability one at histories in $R^*_{k^*}$ that have exactly one instance of (C, T), where this occurs, moreover, at position $j = k$. The signal is s^* with probability ρ, and s_* with probability $1-\rho$ at every other history in $R^*_{k^*}$.

Notice that I-III result in the firm staying forever in the set $R^*_{k^*}$ once the set is reached, and thus the payoff function $V^*(\cdot,\cdot)(\cdot)$ is well defined.

Now, let $M(\rho)$ denote the transition matrix over the states $R^*_{k^*} \times S$ when the firm is type θ, given that I-III hold for some given ρ. $M(\rho)$ gives the transition probabilities of a positive recurrent Markov chain with states $R^*_{k^*} \times S$. There is hence a unique invariant distribution $\pi(\rho)$ on the states, i.e., such that $M(\rho) \cdot \pi(\rho) = \pi(\rho)$.

Write

$$\pi^*(\rho) = \sum_{r \in R^*_{k^*}} \pi(\rho)(r, s^*). \tag{3.16}$$

This is the probability of s^* implied by the invariant distribution $\pi(\rho)$. Notice that $\pi^*(\rho)$ is continuous and strictly increasing in ρ, and moreover that $\pi^*(0) = 1/k^*$, and $\pi^*(1) = 1$.

Next define

$$\theta^*(\rho) = \left\{ \max_{\theta}[\theta_1, \theta_2] : V^*(\theta, \rho)(r) \geq 1 \text{ for all } r \in R^*_{k^*} \right\}. \tag{3.17}$$

If no θ satisfies the required condition of (3.17), set $\theta^*(\rho) = \theta_1$. Notice that Assumption 3.1 implies $\theta^*(1) = \theta_2$.

Let $\rho^* \in (0, 1)$ be such that

$$\left(\frac{\theta^*(\rho^*) - \theta_1}{\theta_2 - \theta_1} + \varepsilon \right) \cdot \pi^*(\rho^*) \ = \ \frac{1}{1+u}. \tag{3.18}$$

The definition of k^* ((3.15)) ensures such a ρ^* exists since: 1) $\pi^*(0) = 1/k$, and 2) $\pi^*(1) = 1$, and $\theta^*(1) = \theta_2$.

Part C: The ψ^* from $[k^*, \rho^*, \psi^*]$ in the lemma.

Fix $\psi^* > 0$ so that for the *NBR* disclosure policy, $[k^*, \rho^*, \psi^*]$, the conclusions of Lemma 3.9 (V3 and V4) hold with η there equal to $\rho^* \cdot (\theta_1 \cdot (1-\delta))^2/2$. That is, with $d = [k^*, \rho^*, \psi^*]$, for every strategic firm θ, and every profile of stationary buyer and intermediary strategies (β, τ),

$$\begin{aligned} &\left|V(\theta, d, (\beta, \tau))(r) - V(\theta, d, (\beta, \tau))(r')\right| < \frac{\rho^* \cdot (\theta_1 \cdot (1-\delta))^2}{2}, \\ &\qquad \text{for all } r, r' \in R_k \text{ with no instances of } (C, T), \text{ and} \\ &\left|V(\theta, d, (\beta, \tau))(r) - \frac{\phi_*(\beta, \tau)}{1-\delta}\right| < \frac{\rho^* \cdot (\theta_1 \cdot (1-\delta))^2}{2}, \\ &\qquad \text{for all } r \in R_k \text{ with an instance of } (D, T). \end{aligned} \tag{3.19}$$

Part D: If $G[k^*, \rho^*, \psi^*]$ has a PT-PBE, then it is unique.

Suppose $\xi = (\{\alpha_\theta\}, \beta, \tau, \mu)$ is a *PT-PBE* of $G[k^*, \rho^*, \psi^*]$. In view of Lemma 3.6 it suffices to show that the buyer pays a positive price for information in ξ.

Let $U(s)$ denote the equilibrium expected payoff to the buyer from choosing trust, conditional on having observed $s \in S$. Let $P\{\cdot\}$ denote the joint distribution over signals and firm choices corresponding to the buyer's equilibrium beliefs about signals and firm choices. With Lemma 3.4 on hand, in order to show that the buyer pays a positive price in ξ, it suffices to show that

$$U(s^*) = u, \quad \text{and} \quad U(s_*) < 0. \tag{3.20}$$

We proceed then to establish (3.20).

In the remainder, for each $r \in R_k$, and each $(a, a') \in \{T, N\} \times \{C, D\}$, we use the notation $[r, a, a']$ to denote the k-length history obtained when the firm, and buyer choose a, and a', respectively, at r.

Recall the definition of $\phi^*(\beta, \tau)$, and $\phi_*(\beta, \tau)$ given in Definition 3.3. We drop the explicit reference to the argument (β, τ) since we fix ξ now, the putative equilibrium of $G[k^*, \rho^*, \psi^*]$. Similarly, we drop the argument $(d, (\beta, \tau))$ in $V(\theta, d, (\beta, \tau))(\cdot)$, and simply write instead $V(\theta)(\cdot)$.

Consider a history $r \in R^*_{k^*}$. Suppose that in the equilibrium ξ of $G[k^*, \rho^*, \psi^*]$, the strategic firm θ chooses D with positive probability at (r, s^*), then (recall here Definition 3.4 of $V(\theta)(\cdot)$)

$$\begin{aligned} \phi^* - \theta + &\delta \cdot \phi^* \cdot V(\theta)([r, C, T]) + \delta \cdot (1 - \phi^*) \cdot V(\theta)([r, C, N]) \\ &\leq \phi^* + \delta \cdot \phi^* \cdot V(\theta)([r, D, T]) + \delta \cdot (1 - \phi^*) \cdot V(\theta)([r, D, N]), \text{ and thus,} \\ &\qquad \theta \;\geq\; \delta \cdot \phi^* \cdot [V(\theta)([r, C, T]) - V(\theta)([r, D, T])] . \end{aligned} \tag{3.21}$$

The last line follows from a rearrangement of the first inequality, and then noting that $V(\theta)([r, C, N]) - V(\theta)([r, D, N]) \geq 0$, since $\psi^* > 0$ by assumption (refer here to the definition of a *NBR* disclosure policy).

Now consider a history $r' \in R_k$ with no instances of (C, T). In an *NBR* disclosure policy such a history is mapped with probability one to s_*. Suppose now that the firm θ choose C at (r', s_*) with positive probability in the equilibrium ξ. Proceeding as in (3.21) gives

$$\begin{aligned} \theta \quad &\leq \quad \delta \cdot \phi_* \cdot [V(\theta)([r', C, T]) - V(\theta)([r', D, T])] \\ + \quad &\delta \cdot (1 - \phi_*) \cdot [V(\theta)([r', C, N]) - V(\theta)([r', D, N])] . \end{aligned} \tag{3.22}$$

Now Lemma 3.8 (V1) implies $V(\theta)([r,D,T]) = V(\theta)([r',D,T])$. V2 from the same result implies $V(\theta)([r,C,T]) \geq V(\theta)([r',C,T])$ (notice that this holds with equality if r' has no instances of (D,T)). Using these facts and combining the last line of (3.21) with (3.22) gives, after some algebra

$$\begin{aligned}
&\phi^* \cdot [V(\theta)([r,C,T]) - V(\theta)([r,D,T])] \leq \\
&\phi_* \cdot [V(\theta)([r,C,T]) - V(\theta)([r,D,T])] + \\
&(1-\phi_*) \cdot \left[V(\theta)([r',C,N]) - V(\theta)([r',D,N])\right], \quad \text{and thus} \\
&(\phi^* - \phi_*) \cdot [V(\theta)([r,C,T]) - V(\theta)([r,D,T])] \leq \\
&(1-\phi_*) \cdot [V(\theta)([r,C,T]) - V(\theta)([r,D,T])] \leq \\
&\qquad V(\theta)([r,C,T]) - V(\theta)([r,D,T]).
\end{aligned} \tag{3.23}$$

Observe next that

$$\begin{aligned}
V(\theta)([r,C,T]) &\geq \rho^* \cdot \phi^* + (1-\rho^*) \cdot \phi_* + \delta \cdot \frac{\phi_*}{1-\delta} \\
&= \rho^* \cdot (\phi^* - \phi_*) + \frac{\phi_*}{1-\delta}.
\end{aligned} \tag{3.24}$$

This is because, by assumption, r is in $R^*_{k^*}$, and therefore $[r,C,T]$ is also in $R^*_{k^*}$.

Combining the bottom half of (3.23) with (3.24) yields

$$\begin{aligned}
&(\phi^* - \phi_*) \cdot \left[\rho^* \cdot (\phi^* - \phi_*) + \frac{\phi_*}{1-\delta} - V(\theta)([r,D,T])\right] \leq \\
&\qquad V(\theta)([r,C,T]) - V(\theta)([r,D,T]), \text{ and thus} \\
&(\phi^* - \phi_*) \cdot [\rho^* \cdot (\phi^* - \phi_*)] \leq \\
&\qquad V(\theta)([r,C,T]) - V(\theta)([r,D,T]) + \left(V(\theta)([r,D,T]) - \frac{\phi_*}{1-\delta}\right).
\end{aligned} \tag{3.25}$$

Using the result of Lemma 3.7 (equation (3.13)) gives

$$\begin{aligned}
&\rho^* \cdot (\theta_1 \cdot (1-\delta))^2 \leq \\
&V(\theta)([r,C,T]) - V(\theta)([r,D,T]) + \left(V(\theta)([r,D,T]) - \frac{\phi_*}{1-\delta}\right).
\end{aligned} \tag{3.26}$$

Notice now that (3.26) contradicts (3.19) given the choice of ψ^*. Hence, what has been shown thus far (beginning around equation (3.21)) is that, in any *PT-PBE* of $G[k^*, \rho^*, \psi^*]$, if that strategic firm chooses defection at (r, s^*) for any $r \in R^*_{k^*}$, then the firm never chooses cooperation at (r', s_*) for any $r' \notin R^*_{k^*}$. This implies that any firm choosing D at any (r, s^*), with $r \in R^*_{k^*}$, will never again draw the signal s^*. It thus follows that $U(s^*) = u$, since in the invariant distribution only firms that choose C with probability one on $R^*_{k^*} \times \{s^*\}$ will ever draw the signal s^*.

Next observe that since $U(s^*)$, ϕ^* must equal one in the equilibrium. This means that no strategic firm will choose C at any (r, s_*), where $r \in R^*_{k^*}$. The reason is that there is nothing to be gained form cooperation at such a history. The firm is guaranteed to stay within $R^*_{k^*}$, when $\phi^* = 1$ and it cooperates on $R^*_{k^*} \times \{s^*\}$, even when it defects at $R^*_{k^*} \times \{s_*\}$. (Recall that in a *NBR* any $r \in R^*_{k^*}$ with exactly one (C, T) that occurs in position $j = k^*$ gets mapped to s^* with probability one, but that all the other histories in $R^*_{k^*}$ induce the same distribution over the signals.) It thus follows that $U(s_*) = u \cdot \varepsilon - (1 - \varepsilon)$ because only the commitment type firm ever chooses C at s_*, and thus $U(s_*) < 0$ by the hypotheses of the lemma. This completes the proof that $G[k^*, \rho^*, \psi^*]$ has at most one equilibrium.

Part E: $G[k^*, \rho^*, \psi^*]$ ***has a PT-PBE.***

Recall I-III from Part B of this proof, and the definitions of $\pi^*(\rho)$, and $\theta^*(\rho)$ given there. Assume II for the moment:

II. The buyer observes the signal with probability one and then chooses T if and only if he observes s^*.

Now consider the following stationary strategy α^*_θ for the firm of type $\theta \leq \theta^*(\rho^*)$ in $G[k^*, \rho^*, \psi^*]$.

$$\begin{aligned} &\alpha^*_\theta(r, s^*) = ((C, D); (0, 1)), \text{ for all } r \in R^*_{k^*}, \\ &\alpha^*_\theta(r, s_*) = ((C, D); (1, 0)), \text{ for all } r \in R^*_{k^*}, \text{ and} \\ &\alpha^*_\theta(r, s) \text{ is a best response to II on } \{R_k \setminus R^*_{k^*}\} \times S. \end{aligned} \tag{3.27}$$

For each $\theta > \theta^*(\rho^*)$, let α^*_θ be a best response for θ to II in $G[k^*, \rho^*, \psi^*]$. Notice that such best responses in stationary strategies are sure to exist against II.

As discussed in Part B of this proof, the strategy α^*_θ for $\theta \leq \theta^*(\rho^*)$, together with II, imply a unique invariant distribution $\mu^*(r, s|\theta)$ on $R_k \times S$. The same is true for each $\theta > \theta^*(\rho^*)$. These imply a unique joint distribution $P\{\cdot\}$ over signals and histories.

As argued above, if it is a best response in $G[k^*, \rho^*, \psi^*]$ for a strategic type of firm to choose defection against II at (r, s^*), where $r \in R^*_{k^*}$, then it is a best response to choose D always at s_*. Against II this means that after defecting at at (r, s^*), where $r \in R^*_{k^*}$, a firm draws the signal s_* for the remainder of the game, and thus gets zero in the continuation game after the defection. The period payoff to a firm from choosing D at such a (r, s^*) is 1. It follows then, that the α^*_θ specified for $\theta \leq \theta^*(\rho^*)$ is a best response in $G[k^*, \rho^*, \psi^*]$ to II (see the definition of $\theta^*(\rho)$, Definition 3.18).

Notice now that each $\theta > \theta^*(\rho^*)$ chooses defection at (r, s^*) for at least one $r \in R^*_{k^*}$. It thus follows that $\mu(r, s^* | \theta) = 0$ for every $\theta > \theta^*(\rho^*)$ (see the discussion in Part D). We thus have

$$\begin{aligned} P\{C | s^*\} &= 1, \quad \text{and} \\ P\{C | s_*\} &= \varepsilon. \end{aligned} \tag{3.28}$$

Equation (3.28) implies

$$\begin{aligned} U(s^*) &= u, \quad \text{and} \\ U(s_*) &= u \cdot \varepsilon - (1 - \varepsilon), \end{aligned} \tag{3.29}$$

where $U(s)$ is the buyer's expected payoff from T conditional on having observed $s \in S$, when the strategic firms adopt the strategies $\{\alpha^*_\theta\}$. Hence the part of II asserting that the buyer chooses T if and only if he observes s^* is in fact a best response to $\{\alpha^*_\theta\}$.

What remains now is to determine the equilibrium price offer by the intermediary. With that in mind (and in view of (3.29), and (3.9) from Lemma 3.4), write

$$\hat{\tau} = \min \{u \cdot P\{s^*\} \;,\; (1 - \varepsilon - u \cdot \varepsilon) \cdot P\{s_*\}\}. \tag{3.30}$$

Let the intermediary offer $\hat{\tau}$ in every period. Let the buyer accept with probability one any offer $\tau \leq \hat{\tau}$, and reject with probability one very offer $\tau > \hat{\tau}$.

Part F: $\hat{\tau}$ in the PT-PBE constructed in Part E is equal to $\tau^* - \varepsilon$.

Refer to equation (3.30) where $\hat{\tau}$ is defined. Notice that in the equilibrium constructed in Part E

$$P\{s^*\} = \left(\frac{\theta^*(\rho^*) - \theta_1}{\theta_2 - \theta_1} \right) \cdot \pi^*(\rho^*).$$

(See equation (3.18), and the definition of $\theta^*(\rho^*)$ in equation (3.17).) ρ^* was chosen so that the right hand side of this expression is equal to $1/(1+u)$. Plugging $P\{s^*\} = 1/(1+u)$ into (3.30) gives the desired result. ■

Chapter 4

Efficiency and Complexity in Decentralized Markets

4.1 Introduction

Consider a dynamic market in which agents are thrown together in every period in order to haggle over the terms of trade. Buyers are matched to sellers in a decentralized fashion. Traders might seek each other out, for instance, or they might just bump into each other at random. Whatever process brings the agents together, the end result is exchange that is separated along temporal and spatial dimensions.

There are several important markets that operate in this manner. Some examples closely fitting this description are the markets for housing, financial derivatives, and labor. An important question then is: Under what conditions will the competitive price and allocation materialize in such a decentralized market?

In an interesting contribution, Gale and Sabourian (2005) (henceforth GS) introduce a refinement that selects competitive equilibria in a finite-agent matching and bargaining model of decentralized trade. [1]

In particular, their argument goes, if agents are averse to complexity, other things equal, then in equilibrium each player will choose the least complex strategy

[1] A competitive equilibrium here refers to an equilibrium of the matching and bargaining game in which all exchange occurs at market clearing prices, and in which the resulting allocation maximizes the gains from trade. This will also be referred to as a *Walrasian equilibrium*.

that earns the equilibrium payoff; as a result, simple behavior will be adopted, which, in turn, ensures Walrasian exchange.

This argument provides a rationale for selecting competitive equilibria within an important framework that describes exchange as resulting from strategic bargaining between pairwise matches. [2] This aspect of their work merits emphasis as it has been observed, since Rubinstein and Wolinsky (1990), that strategic behavior in such settings can result in non-competitive outcomes—even in the absence of trading frictions such as search costs and asymmetric information.

Of central importance to the above argument is the result that simple behavior gives rise to competitive exchange. The appeal of the GS refinement then relies on the extent to which this result holds more generally in frictionless markets. Thus, it is worthwhile to ask if simple behavior will deliver Walrasian outcomes in environments that are close to the one examined by GS.

With that in mind, the current paper studies an extension of their model in which sellers might enter the market with any number of units, rather than just one. This extension is worth considering since there is no *a priori* reason for restricting agents to buying or selling one unit only. [3]

Moreover, by examining the more general case, this paper delivers insights that are drastically different from those obtained by GS. In particular, it is argued here that simple behavior—resulting from complexity aversion, or for any other reason—does not suffice to establish competitive outcomes. In fact, it is shown that when behavior is consistent with a perfect equilibrium simple behavior can rule out efficiency altogether (see Claim 4.2 below).

Allowing an agent to trade several times introduces a strategic aspect that otherwise would not arise. In particular, a several-unit seller must consider, not only the price at which he currently trades, but also the effect of his exchange on future market conditions. Consider a market populated by a finite set of traders with distinct valuations, and in which exchange occurs sequentially over time.

Suppose each buyer leaves the market permanently after making a purchase, and that each seller exits upon having exhausted his supply. In such an environment,

[2] These models are intended to give an explicit account of the inner workings of a decentralized market. See, for example, Rubinstein and Wolinsky (1985), Gale (1986a, 1986b), Rubinstein and Wolinsky (1990), and the recent contribution, Lauermann (2012).

[3] Here the focus is on multi-unit sellers and buyers that demand one unit. The goal is not analyze the most general case but to call attention to interesting strategic aspects that arise when an agent can trade several times in a market.

the execution of certain trades will have an effect on prevailing market clearing prices—those derived from extant supply and demand schedules.

Suppose then, that buyers adopt simple price-taking behavior, whereby each buyer with a sufficiently high valuation attempts to trade at some market clearing price determined by the valuations of the currently active players. Then, a seller with sufficient supply might initially make inefficient exchanges to manipulate supply and demand schedules in order to sell his remaining units at more favorable competitive prices. [4]

Simple behavior here assumes a paradoxical role. If it promotes a tendency toward competitive behavior among buyers, then it paves the way for inefficient price manipulation by a seller with sufficient supply.

The above argument implicitly makes the assumption that buyers do not switch sides in order to resell a purchased unit of the good. Admitting resale introduces the following complication. A seller that sells off a unit in order to manipulate demand, will potentially generate a new competitor, thus nullifying his desired effect on the price.

Whether or not the possibility of resale plays a decisive role depends on the particular market. There are several markets in which resale is simply not feasible. Markets for service goods are perhaps the most extreme example of this. In other cases, licensing restrictions can preclude reselling, or at least make it less profitable (e.g., the market for computer software).

Even when resale is technically and legally feasible, however, in some cases the discrepancy between the prices of new goods and used goods is significant and the secondary market is effectively a market for a different good. A stark example of this is found in the market for autos. A brand new car will depreciate a significant amount as soon as it is driven off the lot. In any case, firms can continue to manipulate demand even when there is a secondary market by simply destroying goods, or otherwise restricting supply.

[4] In a finite agent market, an exchange between a seller and buyer whose valuations are both less than all currently competitive prices will shift the supply schedule leftward but will not affect the demand curve in the relevant region.

4.2 The Environment

The model here extends the one in GS by allowing sellers to trade more than one unit. Relative to the setup there, very little is needed in the way of new objects or notation. The required solution concepts also carry over without modification. Nevertheless, in order that the current paper be self-contained, the model is laid out here in full detail.

Consider a market in which an indivisible good is exchanged for money. At the outset there are finite sets of buyers and sellers, denoted B and S, respectively. Buyer $i \in B$ wishes to consume at most one unit of the good and is willing to pay at most $v_i \geq 0$ to do so. Seller $k \in S$ initially has y_k units available for exchange. His valuation for each of these is a constant, $w_k \geq 0$.[5] Assume, without loss of generality, that the number of buyers is equal to the total initial supply of the good.

Exchange takes place in discrete time as agents bargain over the terms of trade. Buyers leave the market after a purchase, and a seller exits upon exhausting his supply. Hence, the prevailing market data—the active buyers and sellers, and the supply of the good available for exchange—will change over time. Let ξ^* represent the market data initially. Write $\mathscr{E}$ for the set of all submarkets of ξ^*, i.e., data obtained as players trade and exit.[6] Let $\mathscr{E}$ include also the empty market $\emptyset$. Generic elements of $\mathscr{E}$ will be denoted by the letter ξ with subscripts and other embellishments added when needed.

Exchange is governed by a set of bargaining rules and a sequence of deterministic matching functions, $\Pi = \{\pi_t\}_{t=1}^{\infty}$. At date t the function π_t matches one buyer to one seller, from those remaining in the market, and assigns the roles of *proposer* and *responder*. Throughout, matches will be written $\langle m, n \rangle$, where the first coordinate is a proposer. Thus, for each $\xi \in \mathscr{E}$, $\pi_t(\xi) = \langle m, n \rangle$ implies that m and n are active players in ξ—on opposing sides of the market—and additionally that m is the t-th proposer, while n is the t-th responder. At time $t + 1/3$ the proposer in the t-th match names a price for the exchange of one unit. At $t + 2/3$ his partner either *accepts* (A) or *rejects* (R) the offer. When an offer is accepted, trade occurs at

[5] A constant valuation is assumed for ease of exposition. The results here do not require this restriction.

[6] One way to describe a submarket is $\xi = (\{v_i\}_{i \in B'}, \{w_k, y'_k\}_{k \in S'})$, where $B' \subseteq B$ and $S' \subseteq S$, each y'_k satisfies $1 \leq y'_k \leq y_k$, and also $\sum_{k \in S'} y'_k = |B'|$. Market data will be identified with the market itself. For instance, ξ^* will be referred to as the initial market, and data that arises as agents trade and exit will be referred to as a submarket of ξ^*.

the agreed to price; if an offer is rejected, no exchange takes place. At the end of negotiations (exchange or rejection) matches dissolve and a new pair of players is chosen from the remaining ones.

The exchange protocol determines a dynamic matching and bargaining game $g = (\xi^*, \Pi)$ with histories of the form

$$[(\pi_1, p_1, r_1), \ldots, (\pi_{t-1}, p_{t-1}, r_{t-1}), e_t],$$

where, for $\tau = 1, \ldots, t-1$, π_τ is the τ-th match, $p_\tau \in \mathbb{R}_+$ is the τ-th offer, $r_\tau \in \{A, R\}$ is the response to p_τ, and $e_t \in \{\pi_t, (\pi_t, p_t), (\pi_t, p_t, r_t)\}$. It will be useful to identify histories according to their underlying markets. Given a market ξ, say that history h is in ξ if the sequence of events occurring along h lead to the market data ξ. Then, for each $\xi \in \mathscr{E}$, and each π matching players that are active in ξ, let $H(\xi; \pi)$ denote the histories in ξ that end in π.

There is no uncertainty, and furthermore all players observe the entire history describing previous events. A player m strategy is then a function f_m such that $f_m(h) \in \mathbb{R}_+$ if the history h ends with m matched as a proposer, and $f_m(h) \in \{A, R\}$ when m is a responder at h, which ends in a price offer. As in GS there is no discounting. When i is a buyer his utility from trading at p, currently or in the future, is simply $v_i - p$. Similarly, when k is a seller that trades s units at prices, $p_1, \ldots, p_s$, his payoff is $\sum_{r=1}^{s}(p_r - w_k)$. A player's continuation payoff is zero if he does not trade in a particular subgame. Use $V_m(f)$ to denote m's payoff under the profile of strategies f, and let $V_m(f)(h)$ be m's continuation payoff after h according to f.

The analysis to follow will characterize the trading outcomes (the allocation and the exchange prices) induced by a certain class of strategy profiles. The focus will be on the standard criteria of efficiency. An outcome is *efficient* in ξ if it results in the optimal gains from trade possible in ξ. Additionally, it is *Walrasian* (or competitive) in ξ if it is efficient in ξ and furthermore all exchange occurs at market clearing prices given the supply and demand schedules derived from ξ.

The above matching and bargaining game can have perfect equilibria that support non-Walrasian outcomes. More troubling, however, when valuations are heterogeneous, inefficient outcomes can also be supported.[7] GS address this matter

[7] This is true for markets in which sellers have one, or many units, and under various matching protocols (for deterministic matching this is shown by GS, Proposition 13; for random matching it is shown by Gale and Sabourian (2006)).

with a complexity aversion refinement that selects Walrasian equilibria when each agent leaves the market after a single exchange. The rest of this section is devoted to describing this refinement. First, a definition of simple behavior and an ordering of strategies by complexity is needed.

DEFINITION 4.1: *Consider a market* $\xi \in \mathscr{E}$. *Suppose players* m *and* n *are active in* ξ *and on opposing sides of the market. The strategy* f_m *is simple proposing to* n *in* ξ *if it always prescribes the same offer when* m *is a proposer to* n *in* ξ, *that is,* $f_m(h) = f_m(h')$, *for all* $h, h' \in H(\xi; \langle m, n \rangle)$. *Similarly,* f_m *is simple responding to* p *from* n *in* ξ *if it always gives the same response to* n*'s offer of* p *in* ξ, *that is,* $f_m(h) = f_m(h')$, *for all* $h, h' \in H(\xi; \langle n, m \rangle) \times \{p\}$.

Now, one strategy is less complex than the other if they prescribe the same behavior except in some instances where the first is simple and the latter is not.

DEFINITION 4.2: *Consider a player* m. *Say that* f'_m *is less complex than* f_m *if, for some market* ξ *(in which* m *is active) and player* n *(also an active player in* ξ*), either of the following is true: (1)* f'_m *is simple proposing to* n *in* ξ *while* f_m *is not, but otherwise the strategies are identical, or (2) for some* $p \in \mathbb{R}_+$, f'_m *is simple responding to* p *from* n *in* ξ, *while* f_m *is not, but otherwise the strategies are identical.*

A complexity averse equilibrium is then one in which each player chooses the least complex strategy that gives him the equilibrium payoff.

DEFINITION 4.3: *A Nash equilibrium with complexity costs (NEC) is a Nash equilibrium* $f = \{f_m\}_{m \in B \cup S}$ *where, for each player* m, *if* f'_m *is less complex than* f_m, *then* $V_m(f) > V_m(f'_m, f_{-m})$. *A perfect equilibrium with complexity costs (PEC) is a NEC that is also a perfect equilibrium.*

4.3 Complexity Averse Equilibria

GS give a complete characterization of PEC under the following richness condition on matching.

R1: *Consider* $\xi \in \mathscr{E}$ *and any two matches,* π, π', *pairing players in* ξ. *Then, for each* T *there exists a* $t \geq T$ *such that* $\pi_t(\xi) = \pi$ *and* $\pi_{t+1}(\xi) = \pi'$.

Given R1, and when each seller has one unit, every PEC induces a Walrasian outcome in every subgame of the above described matching and bargaining game (this is the main result of GS). In the remainder, say that the strategy f_m is *simple* if it is minimally complex—i.e., it is everywhere simple proposing and responding, for every p and every player on the other side of the market that m might be paired with. A simple strategy is one that conditions only on the underlying market, the current trading partner, and (when responding) the current price offer.

THEOREM 4.1: *Suppose each seller in ξ^* has one unit available for exchange. Let g be the matching and bargaining game with initial data ξ^* and a matching technology that satisfies R1. Then, in any PEC of g, each player's behavior is simple. Furthermore, in every PEC, for every $\xi \in \mathscr{E}$ and after any history in ξ the continuation outcome is Walrasian in ξ with all trade occurring at a uniform price.*

In contrast to Theorem 4.1, when sellers may be endowed with any number of units, simple behavior (and hence a PEC) no longer guarantees Walrasian outcomes or even efficiency. Before giving an example, a convenient notation for market data is established. In the remainder markets will be represented by supply and demand schedules in the following manner:

$$\xi = \begin{pmatrix} v_1 & \dots & v_{s_1} & v_{s_1+1} & \dots & v_{s_2} & \dots \begin{bmatrix} v_{i_m} \\ w_m \end{bmatrix} \dots & v_{s_{\kappa-1}+1} & \dots & v_{s_\kappa} \\ w_1 & & w_1 & w_2 & & w_2 & & w_\kappa & & w_\kappa \end{pmatrix},$$

where buyer valuations have been enumerated so that $v_1 \geq v_2 \geq \cdots \geq v_{s_\kappa}$, seller valuations enumerated so that $w_1 \leq w_2 \leq \cdots \leq w_\kappa$, and $s_r = \sum_{k=1}^{r} y'_k$, for each $s = 1, \ldots, \kappa$, where y'_k is seller k's supply in ξ. The brackets designate the marginal players in ξ, i.e., $v_{i_m} \geq w_m$ but $v_{i_m+1} < w_m$ or $v_{i_m+1} < w_{m+1}$, depending on whether w_m or w_{m+1} is under v_{i_m+1} in the above listing. Observe that for an outcome to be efficient all agents listed to the left of the marginal players and the marginal players themselves—but only these—engage in trade. Furthermore, observe that p is a market clearing price in ξ if and only if $\max\{v_{i_m+1}, w_m\} \leq p \leq \min\{\tilde{w}, v_{i_m}\}$, where $\tilde{w} = w_m$ if $s_m > i_m$, and $\tilde{w} = w_{m+1}$ if $s_m = i_m$.

It is straightforward to construct a market with PEC yielding non-efficient outcomes. The following example is the simplest one that exposes the main idea. It does not require R1, and thus the following weaker condition (implied by R1) will be used.

R2: *For each $\xi \in \mathscr{E}$, and each π matching players in ξ, there is a t for which $\pi_t(\xi) = \pi$.*

CLAIM 4.1: *Consider*

$$\xi^* = \begin{pmatrix} v_1 & \begin{bmatrix} v_2 \\ w_1 \end{bmatrix} & v_3 \\ w_1 & & w_2 \end{pmatrix},$$

where $v_2 > w_2$ and $v_3 \geq w_1$. Let g be the matching and bargaining game with initial data ξ^*, and a matching sequence that satisfies R2. Then g has many PEC supporting inefficient outcomes. Specifically, for each $p \in [w_2, v_2]$ there is a PEC in which seller 2 trades with one buyer $i \in \{1, 2\}$ at p, and seller 1 trades with the remaining buyer in $\{1, 2\}$ at p, and with buyer 3 at a price of v_3.

The proof is relegated to the Section 4.6. The claim is established there by constructing equilibria in simple strategies that support the desired outcomes. These suffice to prove the claim, since simple strategies are minimally complex.

The intuition behind the example is straightforward. First, in markets where seller 1 has one remaining unit, a PEC must deliver a Walrasian outcome (an implication of Theorem 4.1). The market clearing prices prevailing after seller 1 trades initially with buyer 1 or 2 are those in $[v_3, w_2]$. Then, if the continuation price in these markets is low enough, seller 1 would rather trade off a unit to buyer 3 than trade with buyer 1 or 2 initially. The reason is that by doing so he would make seller 2 marginal, permitting the sale of his last unit for a more favorable market clearing price in $[w_2, v_2]$.

There are simple equilibria (and thus PEC) that support Walrasian outcomes in the game of Claim 4.1. When an individual seller is permitted to have more than two units, on the other hand, PEC can preclude efficiency altogether. This is because the payoffs obtained through manipulation—even after trading away units at low prices—might be strictly better than selling at *any* market clearing price as given. There are two separate effects to be considered as an individual seller's supply increases. First, a higher number of units allows him to marginalize more and more sellers with increasingly higher valuations and thus to have a greater influence on market clearing prices. Second, with more supply there is more to be gained from any given price increase. In sum, the motives for price manipulation become more pronounced as a seller is endowed with additional units. The following claim formalizes the intuition.

Claim 4.2: *Consider*

$$\xi^* = \begin{pmatrix} v_1 & v_2 & \begin{bmatrix} v_3 \\ w_1 \end{bmatrix} & v_4 & v_5 \\ w_1 & w_1 & & w_2 & w_3 \end{pmatrix},$$

where $v_3 > w_3$ *and* $v_4 + v_5 + w_3 > 3 \cdot w_2$*. Let* g *be the matching and bargaining game with initial data* ξ^**, and a matching sequence that satisfies R1. Suppose* f *is a PEC of* g*. Then the outcome induced by* f *is not efficient.*

An implication of this example is that sometimes non-simple behavior is needed to deliver efficiency.[8] A sketch of the proof is provided here and formal arguments deferred to Section 4.6. First, in order to exclude seller 2 from trade in a PEC, exchange cannot occur at a price greater than w_2 (established in the formal proof). Additionally, if buyers 4 and 5 are excluded from trade in a PEC, then the equilibrium is such that they will always agree to trade at any price less than their valuations (also established in the formal proof). Suppose then a PEC yields efficient exchange. The condition $v_4 + v_5 + w_3 > 3 \cdot w_2$ ensures that seller 1 will wish to deviate by trading first with buyers 4 and 5—at prices sufficiently close to their respective valuations—in order to marginalize seller 3. This is because by doing so he can sell his last unit for at least w_3 (by Theorem 4.1).

4.4 Walrasian Equilibria and Perfection

In view of Claims 4.1 and 4.2, it is natural to ask if the result of Theorem 4.1 can be restored by considering, perhaps, an alternative formulation of complexity aversion, or of simple behavior. In this connection, the current section argues that perfection may have to be sacrificed in order to guarantee competitive behavior both on and off the equilibrium path. In particular, this is shown for the game from Claim 4.2. There, in order to obtain a Walrasian outcome, agents must first behave in a manner that is consistent with a competitive equilibrium—given the initial market—but must then abandon the competitive equilibrium in subgames resulting after some non-competitive deviations. Here, as in the previous examples, the tension between equilibrium and competition is derived from a seller's capacity for price manipulation.

[8] Lemma 4.4 (in Section 4.6) proves—for the multi-unit seller model considered here—that PEC are simple when R1 holds.

Claim 4.3: *Let g be the game from Claim 4.2. Consider a strategy profile f. Suppose for each $\xi \in \mathscr{E}$, and after each h in ξ, f induces a Walrasian outcome in ξ. Then f is not a perfect equilibrium.*

Proof: By way of contradiction, suppose f is a perfect equilibrium that induces a Walrasian outcome in every subgame of g. In view of the hypotheses of Claim 4.2, choose $\epsilon > 0$ so that $v_4 + v_5 + w_3 - \epsilon > 3 \cdot w_2$. Consider the following deviation by seller 1. Recall that R2 is assumed here (as it is in Claim 4.2).

In ξ^* seller 1 refuses to trade until meeting buyer 4 as a proposer and then proposes $p_1 = v_4 - \epsilon/2$ to buyer 4. Let ξ be the market obtained if seller 1 trades with buyer 4 in ξ^*. After any history in ξ, if buyer 5 has not traded, seller 1 refuses all exchanges until matched as a proposer with buyer 5 and then proposes $p_2 = v_5 - \epsilon/2$ to buyer 5. Otherwise seller 1 behaves as prescribed by his original strategy in profile f.

Now, observe that the set of market clearing prices in ξ^* is $[v_4, w_2]$ so that, by assumption, seller 1's payoff according to f is at most $3 \cdot (w_2 - w_1)$. Since $p_1 + p_2 + w_3 > 3 \cdot w_2$ (by choice of ϵ), to contradict that f is perfect it suffices to show that seller 1's payoff under the deviation is at least $p_1 + p_2 + w_3 - 3 \cdot w_1$. To see this, assume in the remainder that seller 1 behaves according to this deviation, and then note the following.

First, seller 1 eventually trades with buyer 4 at p_1 in ξ^*. This is because f yielding efficiency in every subgame implies: (1) no exchanges occur while seller 1 waits to propose to buyer 4, and (2) by assumption, f is such that buyer 4's payoff is zero after every history in ξ^*, and therefore he must accept any price less than v_4 in ξ^* (by perfection).

Then, after seller 1 and buyer 4 trade in ξ^*, seller 2 becomes the marginal seller and buyer 3 becomes the marginal buyer. Hence buyer 5 does not trade in any subsequently efficient outcome. Recall now that, by assumption, f yields competitive exchange in every subgame. Thus, when seller 1 trades with buyer 4 in ξ^* and then waits to propose to buyer 5, he eventually does sell to buyer 5 at p_2—for the same reasons he could initially trade with buyer 4 at p_4.

Now, any exchange occurring while seller 1 waits to meet buyer 5 is efficient with respect to the market reached when seller 1 and buyer 4 trade in ξ^* (because f induces competitive exchange in all subgames). Thus, after seller 1 and buyer 5 agree to an exchange, seller 3 becomes marginal in the resulting market, and hence market clearing prices are subsequently bounded below by w_3. It follows that seller 1 sells his last unit for at least w_3, i.e., after he trades with buyer 5 his behavior is as under his original strategy, which, given the other players' strategies, induces a Walrasian outcome in the continuation game (by assumption). Clearly, seller 1 is then better off by deviating, and thus f cannot be an equilibrium. ■

4.5 Conclusion

In contrast to GS, the current paper argues that simple behavior—resulting from complexity aversion, or for any reason—does not always ensure efficiency. In particular, the examples here show that non-simple behavior is needed in order to discourage a form of price manipulation whereby a seller alters future market clearing prices by initially engaging in non-efficient exchanges. The arguments in this paper highlight an important tension between equilibrium and Walrasian outcomes in dynamic models of exchange. Although the environment considered here is quite special, this tension will be present in many plausible models of sequential trade among heterogeneous agents. It is worth mentioning, in this regard, that several interesting papers have advanced the notion that simple behavior delivers competitive outcomes in matching and bargaining games. A popular assumption made in this literature is that each agent wishes to trade at most once. The examples considered in this paper show that such an assumption may not be innocuous.[9]

Several aspects of the above arguments merit additional comments. First, consider the robustness of the results here to varying assumptions about matching. There is no reason to believe Claim 4.3 should fail under reasonable alternative matching specifications. For instance, an analogous result will hold when bargaining pairs are chosen randomly in each period, the result being that no equilibrium yields Walrasian exchange in every subgame, with probability 1. In any case, R2 (assumed for Claims 4.1 and 4.3) appears to be a minimal restriction on the set of deterministic matching sequences. Claim 4.2 does require rich and deterministic

[9]For example, in a model where each buyer demands one unit and each seller supplies one unit, Rubinstein and Wolinsky (1990) show that when agents condition their acts only on the set of remaining agents and on time, then a Walrasian equilibrium emerges uniquely. Their result is robust to the perturbation considered in this paper, but only because the agents on each side of the market have homogeneous valuations.

matching.[10] However, similar assumptions will be needed in order to model frictionless trade. In this connection, it is worth noting that the absence of frictions makes price manipulation more, not less, plausible, and thus, in this sense, the results here are robust. As a case in point, consider a setting where sellers are free to solicit bargaining partners, rather than having to wait to be matched with a particular buyer. In such an event it would be easier to carry out price manipulations since the seller would not have to wait for the trades that alter supply and demand schedules in a favorable way, but could solicit them directly.

A second aspect concerns the question of how these results should be interpreted as markets become large—in particular, for a sequence of markets whose supply and demand schedules approximate continuous functions. When supply and demand curves are close to continuous, and each seller's endowment is small relative to the size of the market, the efficiency loss resulting from an individual seller's price manipulation must necessarily be small. Hence, it may be possible that there are simple equilibria giving approximately maximal gains from trade in large enough markets. It is not clear, however, that all simple equilibria can deliver such approximate efficiency. Additionally, even approximately continuous markets will contain many submarkets similar to the ones considered here, and in these simple behavior can obviate efficiency.

A final aspect concerns how the price manipulation here compares to the behavior of a classical monopolist. Recall that a classical monopolist restricts supply in order to increase the price, and then sells to those buyers with the highest valuations. Manipulation here involves trading away units to low valuation buyers in order to "marginalize" higher valuation sellers, and thus put upward pressure on market clearing prices.

4.6 Proofs

Claims 4.1 and 4.2 are proved in this section. The former is established by construction; the latter relies on Lemma 4.1, which will be stated and proved below.

[10] The main result of GS itself requires similar restrictions on matching, that is, without them it is possible to obtain inefficient simple equilibria even when every seller initially has one unit to trade. For the case of sequential deterministic matching see GS, Proposition 13. For the case in which trade is sequential but matching is random see Gale and Sabourian (2006).

4.6.1 Proof of Claim 4.1

A minimally complex equilibrium that yields the desired result will be constructed. To that end, fix $p \in [w_2, v_2]$. With an eye toward conciseness, the phrase "player m tries to trade at s with n" will be used here as shorthand for—when m is a seller, for instance—"m offers s to n, accepts any $q \geq s$ from n, and rejects any $q < s$ from n". Consider the following strategy profile.

1. At all histories in ξ^*: Seller 1 tries to trade at v_3 with buyer 3. He always offers some $q > p$ to buyers 1 and 2. He rejects $p' \leq p$, and accepts any $p' > p$ when buyer 1 or buyer 2 makes the offer. Seller 2 tries to trade at p with every buyer. Buyers 1 and 2 try to trade at p with every seller. Buyer 3 tries to trade at v_3 with every seller.

2. Suppose ξ is any market that could potentially result after seller 1 trades with buyer 3 in ξ^*, or after seller 2 trades with any buyer in ξ^*. At all histories in ξ: Seller 1 tries to trade at p with any remaining buyer in $\{1, 2\}$. If buyer 3 is active in ξ, then seller 1 tries to trade at v_3 with him. If seller 2 is active—i.e., seller 1 traded with buyer 3 at ξ^*—he tries to trade at p with buyers 1 and 2. If buyer $i \in \{1, 2\}$ is active, he tries to trade, with any seller, at p. Buyer 3 tries to trade at v_3 with any seller.

3. If ξ is any market that could result after seller 1 trades with buyer in 1 or with buyer 2 in ξ^*, then at all histories in ξ: Seller 1 tries to trade at v_3, with any remaining buyer. The remaining buyer in $\{1, 2\}$ tries to trade at v_3. Seller 2 tries to trade at w_2. Buyer 3 always offers some $q < v_3$, accepts any $p' < v_3$, and rejects any $p' \geq v_3$.

Clearly this profile supports the inefficient outcome described in Claim 4.1. The only trades that could occur along the equilibrium path in ξ^* are between seller 1 and buyer 3, or between seller 2 and either buyer 1 or buyer 2. All these trades lead to histories from item 2, where the subsequent exchanges complete the desired outcome.

Since each player's above described strategy is minimally complex, in order to establish that the profile is a PEC, it suffices to prove that it is a perfect equilibrium. With that in mind, consider first the histories from items 2 and 3. It is easy to see that no player will wish to deviate from the strategies described for them there, e.g., in 3 the buyers' trading price, v_3, is less than seller 2's reservation price and buyer 3 can not make seller 1 an improving offer that also improves his own payoff.

Clearly, seller 2, and buyers 1,2, and 3 have no incentive to deviate from the strategies described for them at the histories from item 1. At histories from item 1 any payoff altering deviation by seller 1 involves him selling to buyer 1, or to buyer 2. In order to do so he must make the trade at some $q \leq p$. Subsequently the market would transition to one of those from item 3, where he would sell his last unit at v_3. The deviation is then no better than the original strategy under which seller 1 sells one unit for v_3 and another for p. This proves that f is a PEC and thus completes the proof of Claim 4.1.

4.6.2 Proof of Claim 4.2

The following result (Lemma 4.1) is used in verifying the claim. It states that if matching satisfies R1, then payoffs in a PEC have a certain *Markov* property. The result follows immediately from Lemma 4.2 and Lemma 4.4, which are stated and proved in the last parts of this section.

LEMMA 4.1: *Suppose f is a PEC of a matching and bargaining game with a matching sequence that satisfies the richness condition R1. Then, for each $\xi \in \mathscr{E}$, and each player m, $V_m(f)(h) = V_m(f)(h')$, for each h, h' in ξ.*

Let g be as described in the statement of Claim 4.2, and suppose, in contradiction to what is stated there, that f is a PEC yielding an efficient outcome in g.

As a first step, it will be proved that trading prices along the equilibrium path of f are bounded above by w_2. To see this, suppose instead buyer i pays $p > w_2$ along the equilibrium path. Then, consider a history $h \in H(\xi^*; \pi)$, where π matches seller 2 as a proposer to i (such a history assuredly exists by R1). If seller 2 were to propose $p' \in (w_2, p)$ at h, then perfection implies buyer i would accept the offer, i.e., Lemma 4.1, with the assumption that i trades at p along the equilibrium path, implies that i's payoff is $v_i - p$ everywhere on ξ^*. But perfection then implies that seller 2's continuation payoff at h is greater than zero, which contradicts Lemma 4.1 because (by assumption) seller 2 does not trade along the path of f, and hence Lemma 4.1 implies his payoff is identically zero on ξ^*. It has thus been proved that all exchange prices occurring along the equilibrium path are bounded above by w_2, in any PEC delivering efficiency in g.

In order to verify that no PEC yields efficiency in g it will now be shown that seller 1's payoff according to f must be greater than $3 \cdot (w_2 - w_1)$. With that in mind, consider the market

$$\xi = \begin{pmatrix} v_1 & v_2 & \begin{bmatrix} v_3 \\ w_2 \end{bmatrix} & v_5 \\ w_1 & w_1 & & w_3 \end{pmatrix},$$

which is obtained by seller 1 trading with buyer 4 in ξ^*. In view of Lemma 4.1, let V denote seller 1's payoff in ξ according to f.

It will next be established that $V \geq v_5 + w_3 - 2 \cdot w_1$. The first step in this direction is to show that buyer 5's f induced payoff in ξ is equal to zero. Toward that end, suppose, by way of contradiction, that buyer 5 obtains a positive payoff in ξ. Then, since $v_5 < w_2$ and $v_5 < w_3$, perfection implies that buyer 5 trades with seller 1. Let i be the other buyer that seller 1 trades with, and suppose they trade at p. By Lemma 4.1, i's payoff is identically $v_i - p$ on ξ. Perfection then gives that he accepts any price less than p whenever it is offered in ξ. Observe now that seller 2 is marginal in the market reached by seller 1 trading with i in ξ, and moreover that Theorem 4.1 applies in this market since each seller would then have one remaining unit. Thus, seller 1's continuation payoff is bounded below by $p + w_2 - 2 \cdot w_1$ after any history in ξ where he is matched as a proposer with i. Lemma 4.1 then yields $V \geq p + w_2 - 2 \cdot w_1$. But then, seller 1 could not have traded with buyer 5 and with buyer i at p, because $v_5 < w_2$ (by the hypotheses of the claim). This contradicts what was assumed initially, and thus proves that buyer 5's payoff is zero in ξ.

Now, in order to prove that $V \geq v_5 + w_3 - 2 \cdot w_1$ consider the history $h \in H(\xi; \pi)$, where π matches seller 1 as a proposer to buyer 5. Since buyer 5 gets zero in ξ, if seller 1 were to propose any $p < v_5$ at h, then buyer 5 would accept the offer. After such an exchange, seller 3 would become marginal in the resulting market. Having sold to buyer 5, seller 1 could then sell his last unit for at least w_3 (by Theorem 4.1). Clearly then, seller 1's continuation payoff at h is bounded below by $v_5 + w_3 - 2 \cdot w_1$, and hence Lemma 4.1 gives $V \geq v_5 + w_3 - 2 \cdot w_1$.

Now, let V^* denote seller 1's payoff on ξ^* according to f. Since buyer 4's payoff is zero everywhere on ξ^* (i.e., by assumption he does not trade along the path of f), he accepts any $p < v_4$ whenever it is offered in ξ^*. By considering a history in ξ^* at which seller 1 is matched to propose to buyer 4, arguing as above gives $V^* \geq v_4 - w_1 + V$. Along with the previous arguments, this implies $V^* \geq v_4 + v_5 + w_3 - 3 \cdot w_1$.

Recall now the hypotheses of the claim, in particular, $v_4 + v_5 + w_3 > 3 \cdot w_2$. Clearly now, $V^* > 3 \cdot (w_2 - w_1)$. This contradicts that f delivers efficiency (according to the first part of this proof), and thus completes the proof of Claim 4.2.

4.6.3 Proof of Lemma 4.1

Lemma 4.1 follows immediately from Lemmas 4.2 and 4.4, which are stated and proved in the remainder of this section (an intermediate result, Lemma 4.3, is used in the proof of Lemma 4.4).

Recall now that a strategy is simple if is everywhere simple proposing and responding (see Definition 4.1), for every price and every player on the other side of the market an agent might be paired with—a strategy profile is called simple if each of its component strategies is simple. Lemma 4.2 states that when the matching sequence satisfies R1 payoffs in a simple equilibrium have a Markov property; Lemma 4.4 states that when matching satisfies R1 all PEC are simple. In the remainder some additional notation will be required. Recall that $H(\xi;\pi)$ denotes the histories in ξ that end in match π. Then, let $H(\xi;\pi,\pi')$ be histories in $H(\xi;\pi)$ where the match π' follows π when the latter ends in rejection.

Lemma 4.2: *Suppose f is a simple perfect equilibrium in a matching and bargaining game with a matching sequence that satisfies R1. Then, for each $\xi \in \mathscr{E}$, and for each player m, $V_m(f)(h) = V_m(f)(h')$ for all h, h' in ξ.*

Fix an ξ and let f be a subgame perfect simple profile. Let $Q \geq 1$ denote the amount of the tradable good available in ξ. The proof is established by induction on the submarkets of ξ ordered by the number of remaining units of the good. With that in mind, let $\mathscr{E}_q \subseteq \mathscr{E}$ be the set of submarkets of ξ at which extant supply is $q \leq Q$. The following induction claim will now be proved:

For $Q > 1$ suppose that $V_m(f)(\cdot)$ is constant in ξ' for each $\xi' \in \mathscr{E}_{Q-1}$, and each player m, that is, for every agent m, $V_m(f)(h) = V_m(f)(h')$, for all h and h' in ξ'. Then $V_m(f)(\cdot)$ is constant in ξ.

Assume in the following that m is active at ξ (if not the result is trivial) and that the induction hypotheses is true for $Q - 1$. Consider histories, h and h' in ξ, where

$\xi \in \mathscr{E}_Q$. To prove the induction claim it will be shown that $V_k(f)(h) = V_k(f)(h')$ for a seller k—the desired result for buyers goes through similarly. Consider two cases separately.

Case I: No trade in the subgame after h or after h'. If no trade occurs after h, then (since f is simple) no trade occurs after h'. Clearly then, in this case $V_k(f)(h) = V_k(f)(h') = 0$.

Case II: Trade after h and after h'. Now, let $\langle m, n\rangle$ be the first match that trades after h, and let $\langle m', n'\rangle$ denote the first match that trades after h'. Suppose $\langle m, n\rangle$ and $\langle m', n'\rangle$ agree to trade at p and p', respectively. In the remainder, for any buyer-seller pair active in ξ, write ξ^{jl} for the market obtained when j and l agree to trade in ξ. By the induction hypothesis, $V_k(f)(\cdot)$ is constant on ξ^{mn} and on $\xi^{m'n'}$. With that in mind, let V denote k's continuation payoff in ξ^{mn} and let V' be his payoff in $\xi^{m'n'}$, so that,

$$\begin{aligned} V_k(f)(h) &= [p\ - w_k] \cdot I\ + V, \quad \text{and} \\ V_k(f)(h') &= [p' - w_k] \cdot I' + V', \end{aligned} \tag{4.1}$$

where $I = 1 (I' = 1)$ if k is one of the players that trade at $h (h')$, and zero otherwise.

Now, consider, for some buyer i in ξ, the following histories:

$$h^1 \in H(\xi;\ \langle i, k\rangle, \langle m, n\rangle) \ \text{ and } \ h^2 \in H(\xi;\ \langle i, k\rangle, \langle m', n'\rangle),$$

which are sure to exist by R1. Since f is simple and such that m and n trade at p when they are matched at h, it follows that $\langle m, n\rangle$ always trades at p in ξ; similarly $\langle m', n'\rangle$ always trades at p' in ξ. Thus, for all $\tilde{p}$,

$$\begin{aligned} V_k(f)(h^1, \tilde{p}, R) &= [p\ - w_k] \cdot I\ + V, \\ V_k(f)(h^2, \tilde{p}, R) &= [p' - w_k] \cdot I' + V', \text{ and} \\ V_k(f)(h^1, \tilde{p}, A) &= V_k(f)(h^2, \tilde{p}, A) = \tilde{p} - w_k + V^{ik}, \end{aligned} \tag{4.2}$$

where V^{ik} is k's continuation payoff at histories in ξ^{ik}. Now, since f is subgame perfect, k must accept any offer $\tilde{p} > \theta$ at h^1, where

$$\theta - w_k + V^{ik} = [p - w_k] \cdot I + V.$$

Conversely, he will reject any $\tilde{p} < \theta$ there. Similarly, when matched at h^2, he will accept any $\tilde{p} > \theta'$—and reject any $\tilde{p} < \theta'$—where

$$\theta' - w_k + V^{ik} = [p' - w_k] \cdot I' + V'.$$

Observe now that f_k being simple—since k is matched in the same way at h^1 and h^2—implies $\theta = \theta'$. To see that this implies $V_k(f)(h) = V_k(f)(h')$ observe that by definition

$$\theta - \theta' = ([p - w_k] \cdot I + V) - ([p' - w_k] \cdot I' + V'),$$

which (in view of (4.1)) gives $\theta - \theta' = V_k(f)(h) - V_k(f)(h')$. This establishes the induction claim for Case II.

All that remains is to show that the claim is true for ξ when $Q = 1$. In this case there is exactly one buyer and one seller active in ξ (recall that total initial supply is equal to the number of buyers initially). The same arguments used in Case I apply if the remaining buyer and seller do not trade. In the event they do trade, f being simple implies they always trade at the same price in ξ—i.e., set $V = V' = V^{ik} = 0$ and $I = I = 1$ in equations (4.1) and (4.2) and use the arguments from Case II.

In light of Lemma 4.2, in order to complete the proof of Lemma 4.1 what is required now is to show that PEC are simple. The definition of a NEC yields the result immediately for markets that do not occur along the equilibrium path—there is no return to complexity at such markets and thus NEC must be simple there. Verifying the claim for markets that do occur along the path uses an induction argument, which is the content of Lemma 4.3. GS establish the result for the case in which each $y_k = 1$. Here, the proof modifies their techniques to allow for instances in which sellers might trade several units. In the following, say that a profile f is simple at ξ if each component strategy f_m is simple proposing and responding in ξ, for every price and against every player in ξ on the opposite side of the market. Also, for each $\xi \in \mathscr{E}$, say that ξ' is a strict submarket of ξ if ξ' is any market reached after the execution of at least one trade in ξ.

LEMMA 4.3: *Suppose f is a PEC of a matching and bargaining game with a matching sequence that satisfies R1. Let ξ be a market occurring along the path of f. If f is simple in all strict submarkets of ξ, then it is simple in ξ.*

Proof: In the following, let f and ξ be as described in the statement of the lemma. In view of the statement of the lemma, assume in all the following that f is simple in

ξ', for each strict submarket ξ' of ξ. Importantly, Lemma 4.2 then implies that for each such ξ', $V_m(f)(h) = V_m(f)(h')$, for every m active in ξ', and for all h and h' in ξ' (this fact will be used several times in the this proof). The lemma is established in several parts.

PART 1. *Suppose m and n are players active in ξ, on opposing sides of the market. Then, f_n is simple responding to p from m in ξ, for all $p \in \mathbb{R}_+$.*

Proof of Part 1. Suppose not. Then, there exists a price p, and histories $h, h' \in H(\xi; \langle m, n\rangle)$ such that $f_n(h, p) = A$ and $f_n(h', p) = R$. To see that f cannot be a PEC consider two cases separately.

First, suppose the match $\langle m, n\rangle$ does not result in trade at p in ξ along the equilibrium path of f. Consider a deviation f'_n identical to f_n except that f'_n always rejects p when m offers it in ξ. Clearly (f'_n, f_{-n}) and f induce the same equilibrium path so their payoffs are identical and hence f is not a NEC, because f'_n is less complex than f_n.

Next, suppose that $\langle m, n\rangle$ does trade at p in ξ along the equilibrium path of f. Consider f'_n identical to f_n except that f'_n always accepts p from m in ξ. Now, let h^* denote the first f equilibrium history in ξ, and write $f' = (f'_n, f_{-n})$. Since (by assumption) m and n trade at p along the equilibrium path of f, Lemma 4.2 implies $V_n(f')(h, p, A) = V_n(f)(h^*)$ for all $h \in H(\xi; \langle m, n\rangle)$ (i.e., f and f' are identical everywhere except possibly when m proposes p to n in ξ). Then, since h^* occurs along the path of both f' and f, these profiles must yield the same payoffs, and therefore f cannot be a NEC, because by construction f'_n is less complex than f_n.

PART 2. *Suppose the match $\langle m, n\rangle$ trades at p in ξ along the equilibrium path of f. Then, m and n trade at p whenever $\langle m, n\rangle$ occurs in ξ.*

Proof of Part 2. Suppose not, that is, $\langle m, n\rangle$ trades at p in ξ along the equilibrium path but does not always trade at p in ξ. Player n is a simple responder in ξ (shown above), and hence $f_m(h) \neq p$ for some $h \in H(\xi; \langle m, n\rangle)$. Consider a deviation f'_m identical to f_m except that it always offers p to n in ξ. Clearly, f'_m is less complex than f_m. Then, write $f' = (f'_m, f_{-m})$, and let h^* denote the first f equilibrium history in ξ. Lemma 4.2 implies $V_m(f')(h, p) = V_m(f)(h^*)$ for each $h \in H(\xi; \langle m, n\rangle)$—i.e., n always accepts the offer of p and the exchange then leads to the submarket of ξ that occurs along the path of f. Since f and f' both induce the history h^*, it follows they yield the same payoffs, which contradicts that f is a NEC, because f'_m is less complex than f_m.

PART 3. *Suppose the match π occurs in ξ along the path of f, but does not end in trade. Then, this match never results in trade at any history obtained by taking an equilibrium history in $H(\xi;\pi)$ and varying the offers made in ξ but not the responses.*

Proof of Part 3. First, a formal statement of the claim of Part 3 will first be given. To that end, enumerate the matches occurring in ξ along the path of f in order of appearance—e.g., $\pi_1 = \langle m_1, n_1\rangle$ is the first match, $\pi_2 = \langle m_2, n_2\rangle$ is the second, ..., and $\pi_S = \langle m_S, n_S\rangle$ is the last (the one that ends in trade). Let h^* be the first equilibrium history in ξ. Define sets of histories, $H_1, \dots, H_{S-1}$, where

$$\begin{aligned} &H_1 = [h^*, \pi_1], \quad \text{and for each } s = 2, \dots, S, \\ &H_s = \bigcup_{(p_1,\dots,p_{s-1})\in\mathbb{R}_+^{s-1}} [h^*; (\pi_1, p_1, R), \dots, (\pi_{s-1}, p_{s-1}, R),\ \pi_s]. \end{aligned} \tag{4.3}$$

Here, for instance, $[h^*; (\pi_1, p_1, R), \dots, (\pi_{s-1}, p_{s-1}, R),\ \pi_s]$ is the history obtained when, at h^*, p_1 is offered and rejected, then p_2 is offered and rejected, and so on, until finally π_s is matched. The claim made in Part 3 is that, for each $s = 1, \dots, S-1$,

$$f_{m_s}(h) = p \implies f_{n_s}(h, p) = R, \text{ for all } h \in H_s. \tag{4.4}$$

In order to prove (4.4) proceed by induction. With that in mind, fix σ such that $1 \le \sigma \le S-1$. The induction claim is the following. If (4.4) holds for $s = \sigma+1, \dots, S-1$, then it holds for σ. Assume the induction hypothesis and suppose, by way of contradiction, there is some $h \in H_\sigma$, and some p, such that $f_{m_\sigma}(h) = p$ and $f_{n_\sigma}(h, p) = A$.

First, observe that perfection and Lemma 4.2 together imply that m_σ's payoff at h must be at least $V_{m_\sigma}(f)(h^*)$. This is because if m_σ were to elicit rejection at h, the matches $\pi_{\sigma+1}, \dots, \pi_{S-1}$ would subsequently end in rejection (by the induction hypothesis, since $h \in H_\sigma$), and then π_S would trade into the submarket that occurs after ξ along the equilibrium path of f (by Part 2).

Then, observe that $V_{m_\sigma}(f)(h, p) \le V_{m_\sigma}(f)(h^*)$—otherwise, m_σ would initiate the trade with n_σ in ξ along the equilibrium path by offering p (Part 1 implies n_σ would accept the offer).

Clearly now, $V_{m_\sigma}(f)(h, p) = V_{m_\sigma}(f)(h^*)$. Additionally, Lemma 4.2 and Part 1 give $V_{m_\sigma}(f)(h, p) = V_{m_\sigma}(f)(h', p)$, for all $h' \in H(\xi; \pi_\sigma)$, i.e., since n_σ is a simple responder (Part 1) the offer of p from m_σ is always accepted by n_σ in ξ, and then a strict submarket of ξ is reached. It then follows that

$$V_{m_\sigma}(f)(h^*) = V_{m_\sigma}(f)(h', p), \text{ for all } h' \in H(\xi; \pi_\sigma). \tag{4.5}$$

Next, note that m_σ cannot be a simple proposer in ξ. To see this, recall first that $f_{m_\sigma}(h) = p$ and $f_{n_\sigma}(h, p) = A$ by assumption. Then, note that n_σ is a simple responder (Part 1). Thus, if m_σ were a simple proposer, the match π_σ would always result in trade at p in ξ, but, by assumption, it does not result in trade along the equilibrium path.

Finally, consider a deviation f'_{m_σ} identical to f_{m_σ} except that it always offers p to n_σ in ξ. Let $f' = (f'_{m_\sigma}, f_{-m_\sigma})$. Then, f' and f induce the same path to H_σ, and thus equation (4.5) implies that $V_{m_\sigma}(f) = V_{m_\sigma}(f')$, which contradicts that f is a NEC, because the deviation is less complex than the original strategy. This proves the induction claim.

To complete proof of Part 3 it now suffices to establish (4.4) for $s = S-1$. This is done by setting σ in the above arguments to $S-1$.

Part 4. *For each pair of players m and n in ξ, on opposing sides of the market, f_m is simple proposing to n in ξ.*

Proof of Part 4. By way of contradiction, suppose there is a match $\langle m, n\rangle$ and histories $h, h' \in H(\xi; \langle m, n\rangle)$ for which $f_m(h) \neq f_m(h')$. Consider two cases separately.

Assume first that $\langle m, n\rangle$ does not occur in ξ along the equilibrium path. Clearly f_m must be simple proposing to n in ξ. Otherwise m could choose a less complex strategy that gives the same payoff.

Now suppose $\langle m, n\rangle$ does occur in ξ along the equilibrium path. Part 2 then implies this match does not trade in ξ along the equilibrium path (i.e., if $\langle m, n\rangle$ traded at $\tilde{p}$ along the equilibrium path, then it would always trade at $\tilde{p}$ in ξ, which requires m to always propose $\tilde{p}$ to n in ξ).

Next, choose any p offered by m to n in ξ along the equilibrium path (recall that n will always reject this offer). Consider f'_m that always offers p to n in ξ but is otherwise identical to f_m. Clearly, f'_m is less complex than f_m.

Let $H = \bigcup_{s=1}^{S} H_s$ (the H_s's are defined in the formal statement of Part 3), and write $f' = (f'_m, f_{-m})$. Now, observe the following. First, Parts 2 and 3, together with the induction hypothesis and Lemma 4.1, imply that $V_m(f')(h') = V_m(f)(h')$

for any $h' \in H$. Second, f and f' induce the same path to H. Clearly then, $V_m(f') = V_m(f)$, and therefore f cannot be a NEC, because f'_m is less complex than f_m. ■

LEMMA 4.4: *Suppose f is a PEC of a matching and bargaining game with a matching sequence that satisfies R1. Then f is simple in ξ for every $\xi \in \mathscr{E}$.*

Proof: Let $\xi_1, \ldots, \xi_R$ be the markets occurring along the path of f, enumerated in order of appearance. It is immediate that each player's strategy is simple in any $\xi \notin \{\xi_1, \ldots, \xi_R\}$ (otherwise, f could not be a NEC since there is no return to non-simple behavior off the equilibrium path). Then, in view of Lemma 4.3 it suffices to show that f is simple in ξ_R (or ξ_{R-1} in the event ξ_R is empty). To see this is the case proceed as follows.

By assumption there are no trades in ξ_R along the equilibrium. Therefore, if m is active in ξ_R, then $V_m(f)(h) = 0$ for each equilibrium history h in ξ_R, and there is no benefit to m from non-simple behavior in ξ_R. If ξ_R is empty, then the one seller in ξ_{R-1} has one unit to exchange, and hence the subgames in ξ_{R-1} satisfy the hypotheses of Theorem 4.1. Theorem 4.1 then implies f is simple in ξ_{R-1}.■

Bibliography

[1] Simon Baron-Cohen, Alan M. Leslie, and Uta Frith. Does the autistic child have a “theory of mind”? *Cognition*, 21(1):37–46, 1985.

[2] Simon Baron-Cohen, Sally Wheelwright, Richard Skinner, Joanne Martin, and Emma Clubley. The autism-spectrum quotient (aq): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, 31(1):5–17, 2001.

[3] Meghana Bhatt and Colin F. Camerer. Self-referential thinking and equilibrium as states of mind in games: fmri evidence. *Games and Economic Behavior*, 52(2):424–459, 2005.

[4] Patrick Billingsley. *Probability and Measure*. John Wiley and Sons, 2 edition, 1986.

[5] Colin F Camerer, Teck-Hua Ho, and Juin-Kuan Chong. Sophisticated experience-weighted attraction learning and strategic teaching in repeated games. *Journal of Economic Theory*, 104(1):137–198, 2002.

[6] Gary B Charness and Matthew Rabin. Understanding social preferences with simple tests. *Quarterly Journal of Economics*, 117(3):817–869, 2002.

[7] Dorothy L. Cheney and Robert M. Seyfarth. *How Monkeys See the World: Inside the Mind of Another Species*. University of Chicago Press, 1990.

[8] Vincent P. Crawford and Nagore Iriberri. Level-k auctions: Can a nonequilibrium model of strategic thinking explain the winner’s curse and overbidding in private-value auctions? *Econometrica*, 75(6):1721–1770, 2007.

[9] Vincent P. Crawford and Joel Sobel. Strategic information transmission. *Econometrica*, 50(6):1431–1451, 1982.

[10] Johann Elker, David Pollard, and Winfried Stute. Glivenko-cantelli theorems for classes of convex sets. *Advances in Applied Probability*, 11(4):820–833, 1979.

[11] Lawrence E. Fouraker and Sidney Siegel. *Bargaining Behavior*. MacGraw Hill, 1963.

[12] Douglas Gale. Bargaining and competition part I: Characterization. *Econometrica*, 54(4):785–806, 1986a.

[13] Douglas Gale. Bargaining and competition part II: Existence. *Econometrica*, 54(4):807–818, 1986b.

[14] Douglas Gale and Hamid Sabourian. Complexity and competition. *Econometrica*, 73(3):739–769, 2005.

[15] Douglas Gale and Hamid Sabourian. Markov equilibria in dynamic matching and bargaining games. *Games and Economic Behavior*, 54(2):336–352, 2006.

[16] Jacob K. Goeree, Thomas R. Palfrey, Brian W. Rogers, and Richard D. McKelvey. Self-correcting information cascades. *The Review of Economic Studies*, 74(3):733–762, 2007.

[17] John C. Harsanyi. Games with incomplete information played by 'Bayesian' players, I-III. *Management Science*, 14(3,5,7):159–182, 320–334, 486–502, 1967-68.

[18] Sergiu Hart. Evolutionary dynamics and backward induction. *Games and Economic Behavior*, 41(2):227–264, 2002.

[19] Robert S.E. Hurley, Molly Losh, Morgan Parlier, Stephen J. Reznick, and Joseph Piven. The broad autism phenotype questionnaire. *Journal of Autism and Developmental Disorders*, 37(9):1679–1690, 2007.

[20] Gil Kalai. Learnability and rationality of choice. *Journal of Economic Theory*, 113(1):104–107, 2003.

[21] Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *Journal of Political Economy*, 101(6):2590–2615, 2011.

[22] Daniel T. Knoepfle, Colin F. Camerer, and Joseph Tao Yi Wang. Studying learning in games using eye-tracking. *Journal of the European Economic Association*, 7(2/3):388–398, 2009.

[23] Stephan Lauermann. Dynamic matching and bargaining games: A general approach. *American Economic Review*, 103(2):663–689, 2013.

[24] Qingmin Liu. Information acquisition and reputation dynamics. *Review of Economic Studies*, 78:1400–1425, 2001.

[25] Alessandro Lizzeri. Information revelation and certification intermediaries. *RAND Journal of Economics*, 30(2):214–231, 1999.

[26] Kevin A. McCabe, Stephen J. Rassenti, and Vernon L. Smith. Reciprocity, trust, and payoff privacy in extensive form bargaining. *Games and Economic Behavior*, 24(1):10–24, 1998.

[27] Paul R. Milgrom and Robert J. Weber. A theory of auctions and competitive bidding. *Econometrica*, 50(5):1089–1122, 1982.

[28] Erik Mohlin. Evolution of theories of mind. *Games and Economic Behavior*, 75(1):299–318, 2012.

[29] Kristine H. Onishi and Renee Baillargeon. Do 15-month-old infants understand false beliefs? *Science*, 308:255–258, 2005.

[30] Marco Ottaviani and Andrea Prat. The value of public information in monopoly. *Econometrica*, 69(6):1673Ű1683, 2001.

[31] Arthur J. Robson. Why would Nature give individuals utility functions? *Journal of Political Economy*, 109(4):900–914, 2001.

[32] Arthur J. Robson and Hillard Kaplan. The evolution of human longevity and intelligence in hunter-gatherer economies. *American Economic Review*, 93(1):150–169, 2003.

[33] Robert W. Rosenthal. Sequences of games with varying opponents. *Econometrica*, 47(6):1353–1356, 1979.

[34] Ariel Rubinstein and Asher Wolinsky. Equilibrium in a market with sequential bargaining. *Econometrica*, 53(5):1133–1150, 1985.

[35] Ariel Rubinstein and Asher Wolinsky. Decentralized trading, strategic behavior and the walrasian outcome. *Review of Economic Studies*, 57(1):63–78, 1990.

[36] Jórg Oechssler & Burkhard Schipper. Can you guess the game you are playing? *Games and Economic Behavior*, 43(1):137–152, 2003.

[37] Vernon L. Smith. Microeconomic systems as an experimental science. *American Economic Review*, 72(5):923–955, 1982.

[38] Michael Spence. Job market signaling. *Quarterly Journal of Economics*, 87:355–374, 1973.

[39] Dale O. Stahl. Evolution of smart$_n$ players. *Games and Economic Behavior*, 5(4):604–617, 1993.

Appendix A

for “The Evolution of Theory of Preferences”

A.1 Experiment Instructions

The following instructions were given, in written form, to all subjects participating in our experiments.

Page 1

In this experiment you will participate in a series of two person decision problems. The experiment will last for a number of rounds. Each round you will be randomly paired with another individual. The joint decisions made by you and the other person will determine how much money you will earn in that round.

Your earnings will be paid to you in cash at the end of the experiment. We will not tell anyone else your earnings. We ask that you do not discuss your earnings with anyone else.

If you have a question at any time, please raise your hand.

Page 2

You will see a diagram similar to one on your screen at the beginning of the experiment. You and another person will participate in a decision problem shown in the diagram.

One of you will be Person 1 (orange). The other person will be Person 2 (blue). In the upper left corner, you will see whether you are Person 1 or Person 2.

You will be either a Person 1 or a Person 2 for the entire experiment.

Page 3

Notice the four pairs of squares with numbers in them; each pair consists of two earnings boxes. The earnings boxes show the different earnings you and the other person will make, denoted in Experimental Dollars. There are two numbers, Person 1 will earn what is in the orange box, and Person 2 will earn what is in the blue box if that decision is reached.

In this experiment, you can only see the earnings in your own box. That is, if you are Person 1 you will only see the earnings in the orange boxes, and if you are Person 2 you will only see the earnings in the blue boxes. Both boxes will be visible, but the number in the other person's box will be replaced with a "?".

However, for each amount that you earn, the amount the other person earns is fixed. In other words, for each amount that Person 1 sees, there is a corresponding, unique amount that will always be shown to Person 2.

For example, suppose Person 1 sees an earnings box containing "12" in round 1. In the same pair, suppose Person 2 sees "7". Then, at any later round, anytime Person 1 sees "12", Person 2 will see "7".

Together, you and the other person will choose a path through the diagram to an earnings box. We will describe how you make choices next.

Page 4

A node, displayed as a circle and identified by a letter, is a point at which a person makes a decision. Notice that the nodes are color coded to indicate whether Person 1 or Person 2 will be making that decision. You will always have two options.

If you are you Person 1 you will always choose either “Right" or “Down", which will select a node at which Person 2 will make a decision.

If you are Person 2 you will also choose either “Right" or “Down" which will select a pair of earnings boxes for you and Person 1.

Once a pair of earnings boxes is chosen, the round ends, and each of you will be able to review the decisions made in that round.

Page 5

In each round all pairs will choose a path through the same set of nodes and earnings boxes. This is important because at the end of each round, in addition to your own outcome, you will be able to see how many pairs ended up at each other possible outcome.

While you review your own results from a round, a miniature figure showing all possible paths through nodes and to earnings boxes will be displayed on the right hand side of the screen.

The figure will show how many pairs chose a path to each set of earnings boxes.

The Payoff History table will update to display your payoff from the current period.

Page 6

We have provided you with a pencil and a piece of paper on which you may write down any information you deem relevant for your decisions. At the end of the experiment, please return the paper and pencil to the experimenter.

At the end of the experiment, we will randomly choose 2 rounds for payment, and your earnings from those rounds will be summed and converted to $CAD at a rate of 1 Experimental Dollar = $2.

Important points:

- You will be either a Person 1 or a Person 2 for the entire experiment.
- Each round you will be randomly paired with another person for that round.
- Person 1 always makes the first decision in a round.
- Person 1's payoff is in the orange earnings box and Person 2's in the blue earnings box.
- Each person will only be able to see the numbers in their own earnings box.
- Earnings always come in unique pairs so that for each amount observed by Person 1, the number observed by Person 2 will be fixed.
- In a given round, all pairs will choose a path through the same set of nodes and earnings boxes.
- After each round you will be able to see how many pairs ended up at each outcome.
- We will choose 2 randomly selected periods for payment at the end of the experiment.

Any questions?

A.2 Screenshots

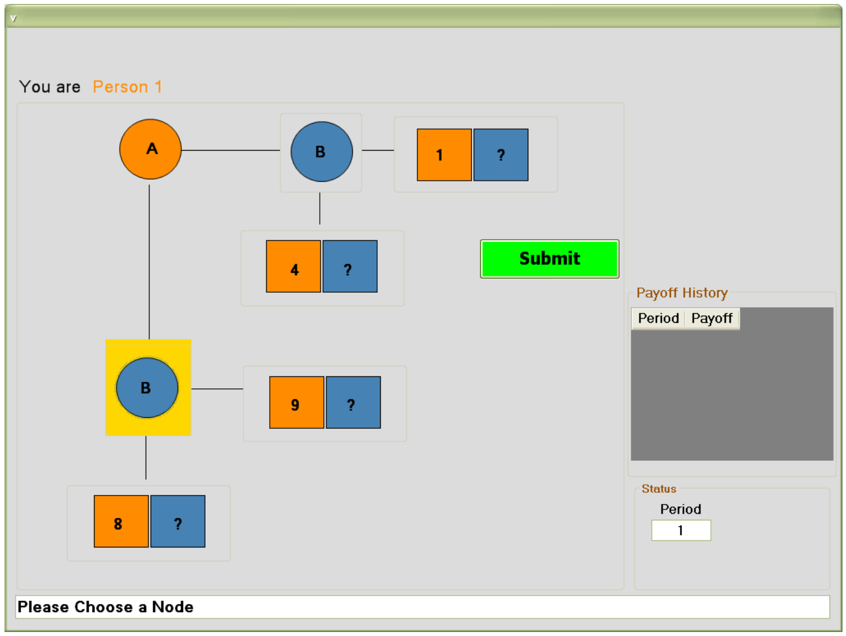

Figure A.21: Screenshot for Player 1.

This figure shows the screen as player 1 sees it prior to submitting his choice of action. The yellow highlighted node indicates that player 1 has provisionally chosen the corresponding action, but the decision is not final until the submit button is clicked. While waiting for player 1 to choose, player 2 sees the same screen except that she is unable to make a decision, provisional choices by player 1 are not observable, and the “Submit" button is invisible.

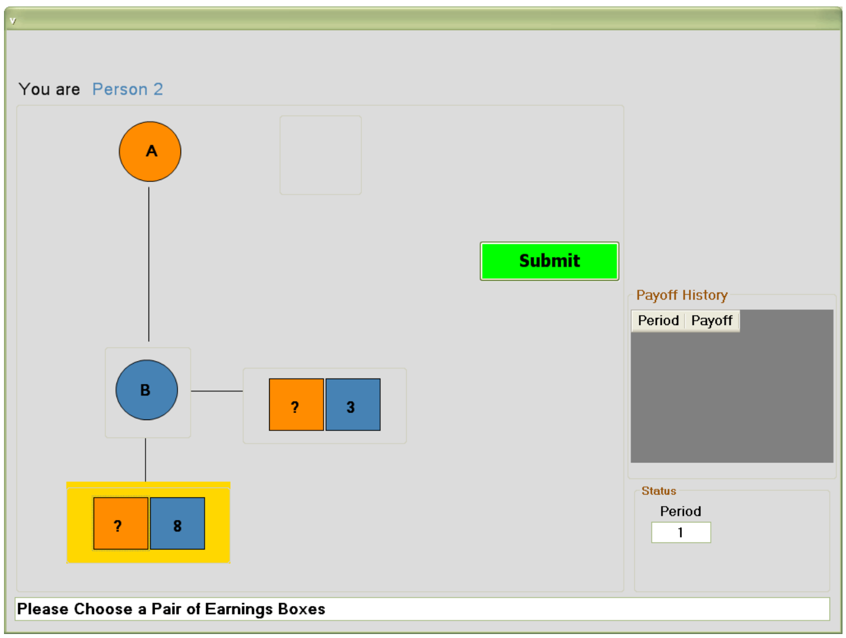

Figure A.22: Screenshot for Player 2.
This figure shows the screen as player 2 sees it after player 1 has chosen an action. Here, player 1 chose to move down, so the upper right portion of the game tree is no longer visible. While player 2 is making a decision, player 1 sees an identical screen except that he is unable to make a decision and the “Submit" button is invisible.

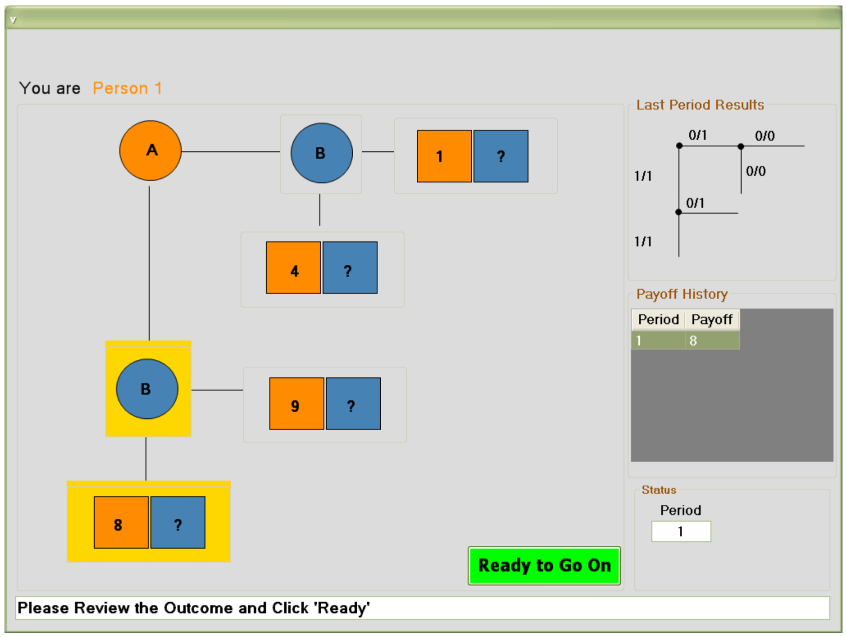

Figure A.23: Screenshot of Post-Decision Review.
This figure shows the final screen subjects see in each period after both player 1 and player 2 have made their decisions. The smaller game tree in the upper right portion of the figure displays information about how many pairs ended up at each outcome. For the purposes of the screenshot, the software was run with only one pair, but in a typical experiment, subjects learned about the decisions of 4 pairs (3 other than their own).

A.3 Autism-Spectrum Quotient Questionnaire

This test is reproduced as per the terms and conditions of the Autism Research Center. It can be found at http://www.autismresearchcentre.com/arc_tests.

1.	I prefer to do things with others rather than on my own.	[1]	[2]	[3]	[4]
2.	I prefer to do things the same way over and over again.	[1]	[2]	[3]	[4]
3.	If I try to imagine something, I find it very easy to create a picture in my mind.	[1]	[2]	[3]	[4]
4.	I frequently get so absorbed in one thing that I lose sight of other things.	[1]	[2]	[3]	[4]
5.	I often notice small sounds when others do not.	[1]	[2]	[3]	[4]
6.	I usually notice car number plates of similar strings of information.	[1]	[2]	[3]	[4]
7.	Other people frequently tell me that what I've said is impolite, even though I think it is polite.	[1]	[2]	[3]	[4]
8.	When I'm reading a story, I can easily imagine what the characters might look like.	[1]	[2]	[3]	[4]
9.	I am fascinated by dates.	[1]	[2]	[3]	[4]
10.	In a social group, I can easily keep track of several different people's conversations.	[1]	[2]	[3]	[4]
11.	I find social situations easy.	[1]	[2]	[3]	[4]
12.	I tend to notice details that others do not.	[1]	[2]	[3]	[4]
13.	I would rather go to a library than a party.	[1]	[2]	[3]	[4]
14.	I find making up stories easy.	[1]	[2]	[3]	[4]
15.	I find myself drawn more strongly to people than to things.	[1]	[2]	[3]	[4]
16.	I tend to have very strong interests, which I get upset about if I can't pursue.	[1]	[2]	[3]	[4]
17.	I enjoy social chit-chat.	[1]	[2]	[3]	[4]
18.	When I talk, it isn't always easy for others to get a word in edgeways.	[1]	[2]	[3]	[4]
19.	I am fascinated by numbers.	[1]	[2]	[3]	[4]
20.	When I'm reading a story I find it difficult to work out the characters' intentions.	[1]	[2]	[3]	[4]
21.	I don't particularly enjoy reading fiction.	[1]	[2]	[3]	[4]
22.	I find it hard to make new friends.	[1]	[2]	[3]	[4]
23.	I notice patterns in things all the time.	[1]	[2]	[3]	[4]
24.	I would rather go to the theatre than a museum.	[1]	[2]	[3]	[4]
25.	It does not upset me if my daily routine is disturbed.	[1]	[2]	[3]	[4]
26.	I frequently find that I don't know how to keep a conversation going.	[1]	[2]	[3]	[4]
27.	I find it easy to "read between the lines" when someone is talking to me.	[1]	[2]	[3]	[4]
28.	I usually concentrate more on the whole picture, rather than the small details.	[1]	[2]	[3]	[4]
29.	I am not very good at remembering phone numbers.	[1]	[2]	[3]	[4]
30.	I don't usually notice small changes in a situation, or a person's appearance.	[1]	[2]	[3]	[4]
31.	I know how to tell if someone listening to me is getting bored.	[1]	[2]	[3]	[4]
32.	I find it easy to do more than one thing at once.	[1]	[2]	[3]	[4]
33.	When I talk on the phone, I'm not sure when it's my turn to speak.	[1]	[2]	[3]	[4]
34.	I enjoy doing things spontaneously.	[1]	[2]	[3]	[4]
35.	I am often the last to understand the point of a joke.	[1]	[2]	[3]	[4]
36.	I find it easy to work out what someone else is thinking or feeling just by looking at their face.	[1]	[2]	[3]	[4]
37.	If there is an interruption, I can switch back to what I was doing very quickly.	[1]	[2]	[3]	[4]
38.	I am good at social chit-chat.	[1]	[2]	[3]	[4]
39.	People often tell me that I keep going on and on about the same thing.	[1]	[2]	[3]	[4]
40.	When I was young, I used to enjoy playing games involving pretending with other children.	[1]	[2]	[3]	[4]
41.	I like to collect information about categories of things (e.g. types of car, types of bird, types of train, types of plant, etc.	[1]	[2]	[3]	[4]
42.	I find it difficult to imagine what it would be like to be someone else.	[1]	[2]	[3]	[4]
43.	I like to plan any activities I participate in carefully.	[1]	[2]	[3]	[4]
44.	I enjoy social occasions.	[1]	[2]	[3]	[4]
45.	I find it difficult to work out people's intentions.	[1]	[2]	[3]	[4]
46.	New situations make me anxious.	[1]	[2]	[3]	[4]
47.	I enjoy meeting new people.	[1]	[2]	[3]	[4]
48.	I am a good diplomat.	[1]	[2]	[3]	[4]
49.	I am not very good at remembering people's date of birth.	[1]	[2]	[3]	[4]
50.	I find it very easy to play games with children that involve pretending.	[1]	[2]	[3]	[4]

1 = definitely agree, 2 = slightly agree, 3 = slightly disagree, 4 = definitely disagree

A.4 Additional Tables and Figures

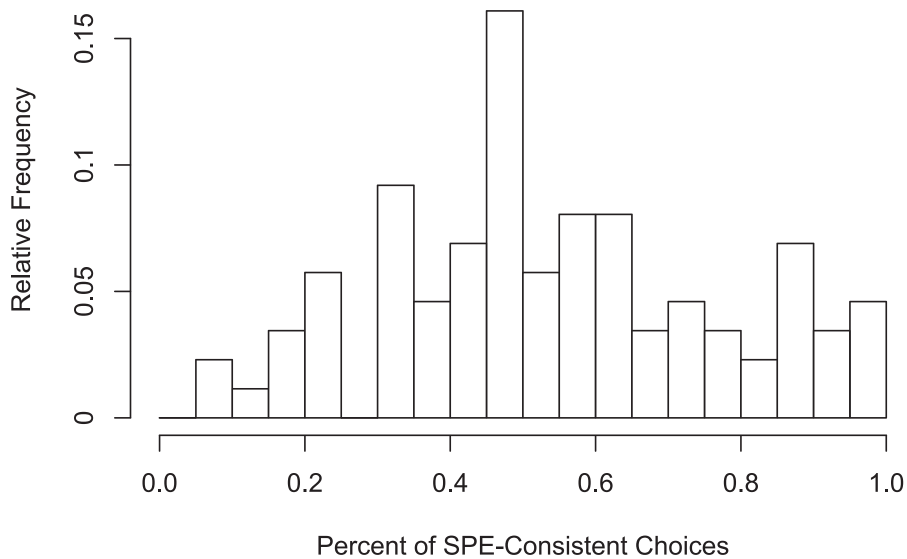

Figure A.41: Histogram of the Individual Rates of SPE-consistent Choices. The figure excludes all periods in which the player had a dominant strategy and in which choice under the rule of thumb corresponded to the SPE.

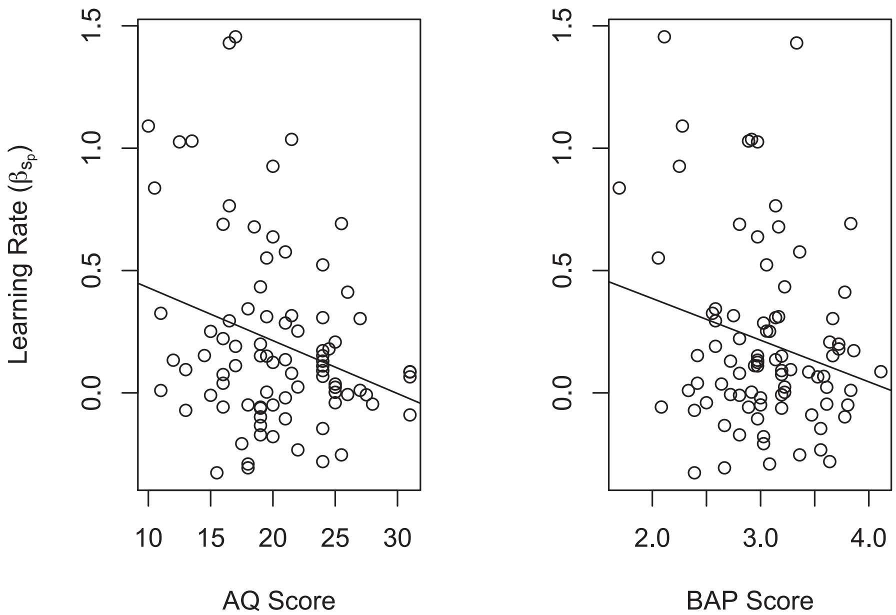

Figure A.42: Scatterplots Comparing Learning Rates to Theory of Mind Measures from Psychology.
The solid lines plot OLS fits of the data.

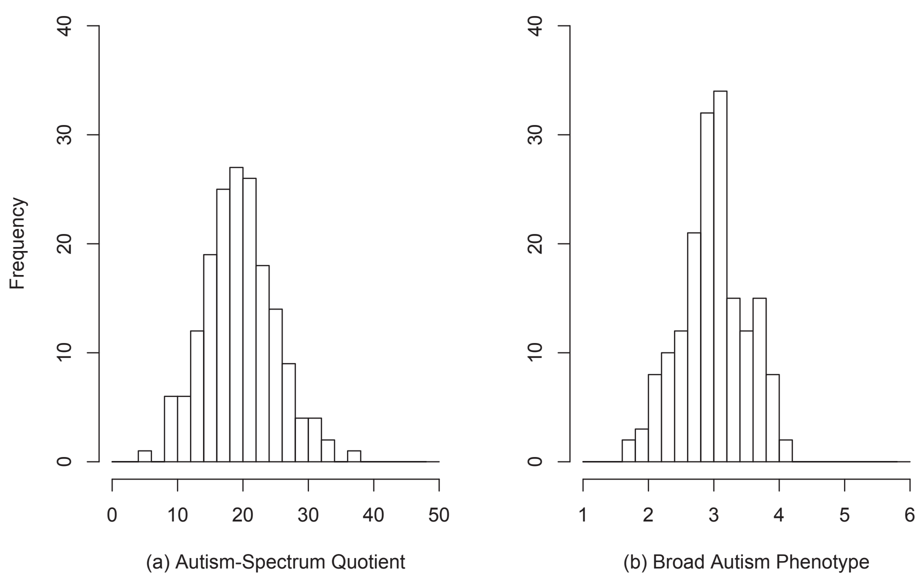

Figure A.43: Histograms of AQ and BAP Scores.
Each panel includes the entire range of feasible scores.

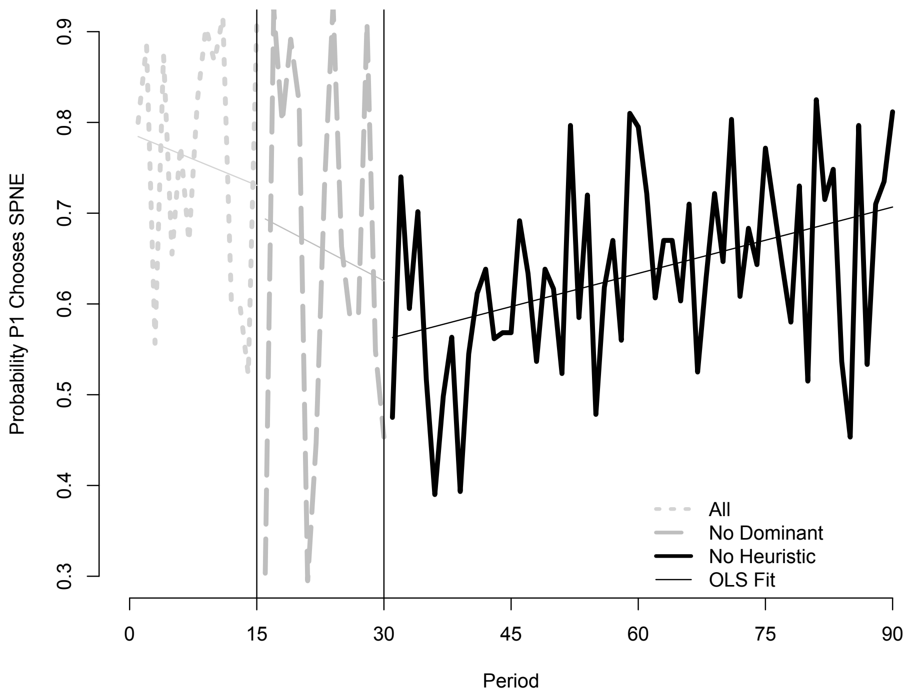

Figure A.42: Aggregate Learning Data.

means	8-1	8-2	10-1	10-2	8-3	6-1	8-4	10-3	8-5*	10-3*	10-4	10-5	8-6	10-6	8-7	8-8	8-9	8-10	10-7	8-11
Pairs SPE	0.68	0.54	0.62	0.61	0.63	0.44	0.51	0.46	0.68	0.50	0.32	0.66	0.53	0.38	0.54	0.52	0.57	0.64	0.25	0.71
P1 SPE	0.72	0.59	0.64	0.71	0.67	0.54	0.61	0.54	0.70	0.59	0.50	0.72	0.64	0.50	0.67	0.61	0.63	0.70	0.36	0.72
P2 SPE | P1 SPE	0.94	0.91	0.98	0.87	0.94	0.83	0.84	0.86	0.96	0.85	0.63	0.92	0.83	0.77	0.80	0.85	0.91	0.92	0.71	0.98
P1 Heur	0.36	0.40	0.42	0.38	0.42	0.45	0.34	0.55	0.43	0.43	0.55	0.38	0.46	0.47	0.44	0.52	0.49	0.45	0.59	0.34
P1 SPE | heur ≠ SPE	0.64	0.52	0.55	0.58	0.55	0.36	0.49	0.37	0.64	0.45	0.25	0.64	0.44	0.31	0.46	0.43	0.51	0.58	0.21	0.68
P1 SPE | heur = SPE	0.81	0.60	0.86	0.74	0.88	0.71	0.58	0.75	0.78	0.73	0.57	0.70	0.73	0.59	0.75	0.77	0.78	0.82	0.52	0.88
P1 SPE | No Heur	0.71	0.69	0.65	0.72	0.70	0.49	0.56	0.57	0.74	0.66	0.63	0.79	0.60	0.51	0.73	0.78	0.73	0.78	0.52	0.81
P2 Dom	0.94	0.90	0.98	0.86	0.94	0.84	0.85	0.82	0.97	0.87	0.63	0.93	0.86	0.78	0.80	0.83	0.91	0.93	0.66	0.98

Table A.41: Observed Probability of Outcomes by Session.

Each entry is a probability that we observed a particular outcome in a particular session. Pairs SPE refers to the probability that a pair ended at the *SPE*. P1 SPE is the probability that player 1's choice was consistent with the *SPE*. P1 Heur is the probability that Player 1 chose in a manner consistent with the "highest mean" rule of thumb (heuristic). P1 SPE | heur ≠ SPE is the probability that player one followed the *SPE* when it did not correspond to the "highest mean" rule of thumb. P1 SPE | heur = SPE is the probability that player one followed the *SPE* when it *did* correspond to the "highest mean" rule of thumb. P1 SPE | No Heur is the probability that player 1 followed the *SPE* when the rule of thumb was inapplicable (i.e. equal means). P2 Dom is the probability that player 2 chose the dominant strategy. P2 SPE | P1 SPE is the conditional probability of player 2 choosing the dominant strategy given that player 1 followed the *SPE*. Sessions are labeled in the format # of Subjects – Session ID so that 10-2 corresponds to the 2nd session with 10 subjects. * indicates sessions in which the payoff set was {1,2,3,4,8,9,10}, rather than {1,2,3,4,5,6,7}.