SOPHISTICATED PLAY BY IDIOSYNCRATIC AGENTS

DAVID P MYATT AND CHRIS WALLACE

January 1999

Address for Correspondence:

DAVID P MYATT, NUFFIELD COLLEGE, OXFORD OX1 1NF, UK

ABSTRACT. The payoffs of a symmetric 2×2 coordination game are perturbed by agent-specific heterogeneity. Individuals observe a (possibly sampled) history of play, which forms the initial hypothesis for an opponent's behaviour. *Seeding beliefs* in this manner, they iteratively reason toward a Bayesian Nash equilibrium. Realised actions augment history and *context* evolves. Both *risk-dominance* and *generalised risk-dominance* determine equilibrium selection as heterogeneity vanishes. When sampling is sufficiently incomplete, the risk-dominant equilibrium is played irrespective of the history observed.

1

1. Introduction

Say first, of God above or man below,

What can we reason but from what we know?

Alexander Pope: *An Essay on Man*

Many interesting games possess multiple Nash equilibria. Absent a suitable refinement concept, many theorists encourage the examination of the *context* in which a game is played.[1] In particular, the environment may provide a guideline for players. If this guideline specifies a particular Nash equilibrium, then an agent can do no better than to play their part in it. In a coordination game, for example, a convention to play a particular equilibrium is self-enforcing. Such advice is unsatisfactory, however, as it does not explain how such an environment might arise.

Players may look to history to inform their decisions and this can, therefore, provide a suitable context. Of particular relevance are the actions taken previously by others. Observation of these actions allow an agent to infer the typical play of a game. An individual will then select a strategy contingent on this observation. The realised action becomes part of history and history thus evolves.

Jointly modelling the interdependence of history and action choice allows selection between equilibria. To some extent, this is the approach of the adaptive learning literature epitomised by the work of Kandori, Mailath and Rob (1993).[2] They specify a dynamic in which entrants respond to an existing population's strategy frequency. Such a dynamic is path-dependent as the initial configuration of history determines long-run behaviour. In response the authors achieve ergodicity via the introduction of mutations — agents make mistakes when choosing their strategies. The invariant distribution of the resulting process is examined as mutations vanish and an equilibrium is selected irrespective of initial conditions.

---

[1] See for example Binmore (1994) or Schelling (1960).
[2] Henceforth KMR (1993).

The adaptive learning approach is subject to a number of critiques. Players form beliefs naïvely and also fail to choose optimal actions given those beliefs. The latter objection is highlighted by Myatt and Wallace (1997)[3] in which agents *differ* rather than *err*. Ergodicity of the adaptive learning process requires occasional contrarian behaviour, where entrants choose against an established convention. A more convincing explanation for such behaviour is *idiosyncrasy* on the part of individuals, and this is the approach employed here. The focus in the present paper will be on the former objection. If a player repeatedly interacts with randomly selected members of the population, then adopting a best response to the incumbent strategy frequency is not unreasonable. In contrast, when an agent expects a single play against an opponent who simultaneously selects a strategy, this is suspect. Such a framework is envisioned by Young (1993). His model specifies pairs of entrants who, following a (sampled) observation of history, independently choose strategies and play a single game. In this situation, an intelligent agent would reason more carefully.

This paper recognises and responds to this critique. Observed frequencies no longer provide a theoretical opponent. Rather, context *seeds the beliefs* of a player. An agent initially conjectures that their opponent will act optimally given their observation yielding a new hypothesis for their opponent's play. Bearing the agent's reaction to this in mind, their opponent can calculate a new best response. Anticipating such a thought experiment, the agent constructs a best response to *this* strategy. Thus, entrants engage in an iterative best response process whose limit crucially depends upon the starting point. This starting point is provided by history. Such intelligent reflection upon ones environment and the reasoning procedure of others is *sophisticated play.*

Full rationality (in the sense commonly used by economists) might seem a more appropriate response to the naïveté inherent in the adaptive learning literature. It is not. Observed strategy frequencies would be ignored by "rational" players as they do not have a direct effect on payoffs. Agents would proceed directly to an equilibrium. Again, the

---

[3] Henceforth MW (1997).

modeller is left without a mechanism for selecting between equilibria and players have no way to coordinate. The suggestion here is that they *can* use observations of past play to coordinate on an equilibrium via the use of some reasonable thought experiment. This is motivated by a desire to model the way in which individuals might think strategically. Unlike full rationality, sophisticated play is a behavioural postulate — a common feature of both adaptive and evolutionary research. Such research might question the computational ability required of the agents in this model. Although players reason indefinitely, the intention here is to capture a situation in which agents think *to some extent* about their opponent's behaviour. In fact, many of the conclusions are largely unaffected by allowing less sophistication on the part of the players.[4]

When sampling is complete, agents observe identical strategy frequencies and this is common knowledge. The iterative reasoning process converges to a (history dependent) Bayesian Nash equilibrium of the underlying stage game. As strategies in such Bayesian equilibria are trigger rules, either action may be realised. Allowing the individual-specific heterogeneity to vanish, these equilibria correspond to the pure Nash equilibria of the unperturbed game. With vanishing noise, only one of the actions is played almost all of the time. Thus an equilibrium is selected. The selection criterion depends crucially upon the *depth* and *width* of the *basins of attraction*. If an equilibrium is both *risk-dominant* (Harsanyi and Selten 1988) and *generalised risk-dominant* (MW 1997) then it is selected.

When sampling is incomplete, selection results may radically differ. Agents no longer observe common histories, and hence they must carefully consider not only their opponent's action but also what their opponent might have seen. Such deliberation, however, must also pay heed to their opponent's beliefs about what *they* might have seen. Much of the epistemological literature focuses on this issue, see for example Morris, Rob and Shin (1995). An infection argument applies. An agent initially conjectures a best response to the observed strategy frequency. Upon further reflection they recognise that, with high

---

[4]Agents are allowed to reason indefinitely purely for mathematical convenience.

probability, an opponent will observe a different frequency. Such observations may generate a contrary best response. Mindful of this, the initial player changes strategy. With sufficiently many iterations of reasoning, a single strategy may be adopted by all entrants irrespective of their observations. In fact, when sampling is sufficiently incomplete all players adopt the risk-dominant equilibrium *regardless* of history. Moreover, this result holds without resorting to perturbation of the non-ergodic Markov process.

The argument proceeds as follows. Section 2 outlines the model. The trembled stage game is introduced (2.1) followed by the dynamic (2.2). In Section 2.3 the iterative reasoning process employed by sophisticated agents is both described and justified. These components are illustrated by a motivating example of PC adoption in Section 2.4. The analysis takes place in Section 3. Results are first obtained for the reasoning process with full sampling (3.2) and its associated dynamic (3.3). Secondly, Sections 3.4 and 3.5 present the results for incomplete sampling. Returning to the example in Section 4, the results are discussed and illustrated. These elements are drawn together along with some concluding remarks in Section 5. For convenience, omitted results are collected in Appendix A.

2. The Model

In this section the model is described. First, the trembled stage game of MW (1997) is reviewed. Concepts of risk-dominance and generalised risk-dominance are re-introduced. Turning to the updating procedure, Section 2.2 constructs the specific dynamic driving the population's evolution. The method by which individuals make decisions is left open — this is dealt with in Section 2.3. Here, the reasoning process which connects decision to context is both introduced and defended. Finally, a motivating example is presented in Section 2.4.

2.1. **The Trembled Stage Game.** The key elements of MW (1997) are retained here. In particular, the analysis centres on the familiar symmetric 2×2 strategic form game

with generic payoffs:

	1	2
1	a $\quad$ a	c $\quad$ b
2	b $\quad$ c	d $\quad$ d

Notice that this game can be represented by a 2×2 matrix.

Definition 1. *The* mean payoff matrix *is defined as:*

$$\Lambda = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

A player equipped with payoffs Λ*, entering a population is a* mean payoff entrant*.*

Coordination games are of particular interest, and will be the focus of subsequent analysis. This is when $a > c$ and $d > b$. Such a game has two pure Nash equilibria, $(1, 1)$ and $(2, 2)$ with associated security payoffs b and c. The symmetric mixed equilibrium entails mixing probabilities of $[x^*, 1 - x^*]$ where:

$$x^* = \frac{(d - b)}{(d - b) + (a - c)}$$

Note that equilibrium $(1, 1)$ risk-dominates $(2, 2)$ if and only if $a - c > d - b$, corresponding exactly to $x^* < \frac{1}{2}$. Unless otherwise stated, and without loss of generality, assume $x^* < \frac{1}{2}$ throughout.

The payoffs Λ may be viewed as the mean payoffs for any entering player. Any particular agent has heterogeneous preferences which are generated by the addition of payoff trembles. Each payoff is subject to an independent Gaussian disturbance. The variances of these disturbances may be strategy profile specific, with a common scaling factor which is allowed to vanish for limiting results. Clearly these disturbances have a fully-parametric form. This, however, is a natural representation of differing payoffs across

players. In particular, one might view differences over a particular payoff to be the resulting sum of many individual idiosyncratic factors, yielding the normal distribution as a natural specification. Furthermore, this formulation allows clear closed-form results to be obtained. Bergin and Lipman (1996) show that full generality of trembles leads to inconclusive results. Hence the parametric approach taken here is justified.

Definition 2. *Define the* payoff heterogeneity matrix Ψ *as:*

$$\Psi = \begin{bmatrix} \sigma_a & \sigma_b \\ \sigma_c & \sigma_d \end{bmatrix}$$

An entrant has trembled payoff matrix $\widetilde{\Lambda}$ *where:*

$$\widetilde{\Lambda} = \begin{bmatrix} \tilde{a} & \tilde{b} \\ \tilde{c} & \tilde{d} \end{bmatrix} = \begin{bmatrix} a + \sigma\varepsilon_a & b + \sigma\varepsilon_b \\ c + \sigma\varepsilon_c & d + \sigma\varepsilon_d \end{bmatrix}$$

where $\varepsilon_i \sim N\left(0, \sigma_i^2\right)$, *with* $E\left[\varepsilon_i\varepsilon_j\right] = 0$ *for* $i \neq j$ *and* σ *is a common scaling factor.*

Note that the payoff heterogeneity matrix Ψ determines the *relative* size of the payoff trembles. The overall size is determined by the scaling factor σ. For later convenience, the following definitions are introduced.

Definition 3. *The* payoff balance *of* Λ *and the* tremble balance *of* Ψ *are respectively:*

$$\lambda = 1 - x^* = \frac{(a - c)}{(a - c) + (d - b)} \quad and \quad \psi = \frac{(\sigma_a^2 + \sigma_c^2)}{(\sigma_a^2 + \sigma_c^2) + (\sigma_b^2 + \sigma_d^2)}$$

A game has balanced *trembles if* $\psi = \frac{1}{2}$. *Otherwise it has* unbalanced *trembles.*

Definition 4. *Strategy* 1 generalised-risk-dominates *strategy* 2 *whenever:*

$$\frac{a - c}{\sqrt{\sigma_a^2 + \sigma_c^2}} > \frac{d - b}{\sqrt{\sigma_b^2 + \sigma_d^2}}$$

Using the notions of balance, this is equivalent to $\lambda/(1 - \lambda) > \sqrt{\psi/(1 - \psi)}$. Assembling these components yields the trembled stage game:

Definition 5. *Define the* trembled stage game $\mathcal{G}$ *as the triple:*

$$\mathcal{G} = \langle \Lambda, \Psi, \sigma \rangle$$

Note that $\mathcal{G}$ is a Bayesian game of incomplete information.

2.2. **The Evolution of Context.** Take a finite population of n players. At the beginning of a period each player simultaneously updates their strategy. Alternatively, this can be interpreted as a new group of players replacing last period's entire population. Each new entrant (or player updating) observes the strategy frequency of the population in a sample of size $s \leq n$ taken from last period's play. The agent considers playing a game with a prospective opponent who also observes a (possibly different) sample of size s. In particular, the iterative procedure outlined in Section 2.3 below takes place in the agent's mind, and a strategy is selected. This agent, along with their new strategy, then becomes part of the updated population, and the whole process repeats itself. Denote the number of individuals playing strategy 1 as z, a member of the finite state space $Z = \{0, ..., n\}$. Computation of the probability that the agent chooses either strategy leads to the construction of a Markov chain on Z.

In the dynamic described above, all players simultaneously update. This coincides with the approach of KMR (1993). If agents were to update alone and enter a population that was essentially static in strategy frequency, it might be more reasonable for them to simply play a best response to the present state (as in KMR (1993) and Young (1993)). This is the methodology of MW (1997) with players updating sequentially. When there is simultaneous updating, however, it is more reasonable to assume that players would think very carefully about their planned action. In particular they would be concerned that a prospective opponent might also base their decision upon observations of history. To some extent, Young (1993) adopts such an idea. Taking this a step further, a *sophisticated* player might act in the way described in Section 2.3.

The dynamic presented here makes more sense in a sophisticated world, enables a transparent analysis and helps to focus on the crucial components of the argument.

2.3. **The Reasoning Process.** Agents who play adaptively (MW 1997, KMR 1993, Young 1993 *inter alia.*) reason very simply indeed. On observation of incumbent strategy frequencies, individuals adopt a simple best response. This naïve behaviour might be justified in a scenario in which an entrant faces a static population, perhaps meeting a randomly selected opponent for a single interaction. Alternatively, this approach is not unreasonable if an entrant repeatedly plays within a slowly evolving population. Notice that these conditions are not met by many of the models mentioned earlier. In particular, KMR (1993) envision a dynamic in which the entire population revises simultaneously. An intelligent player might realise this and adjust their behaviour accordingly — they may anticipate the updated actions of their opponents. In addition, the players of Young (1993) see only a *sample* of history to which they play a best response. More sophistication on their part would lead them to a consideration of the (possibly different) observations of an opposing agent.

These are precisely the issues to be addressed here. To some extent, agents recognise the interactive nature of their decisions. They begin (as in previous work) by hypothesising a best reply to the strategy frequency of their sample. The agent conjectures a similar response from their opponent. Having calculated this, the agent can formulate a best response to *this* postulated behaviour. An analogous thought experiment is anticipated for their opponent. Bearing in mind the possibly different sample of a prospective opponent an agent will iteratively reason until convergence. More formally, agents play a static Bayesian game where their private information is both the realised payoffs and the observed sample. A myopic strategy profile initialises an iterative best response process which leads to a Bayesian Nash equilibrium. This equilibrium generates the actual behaviour of the agent.

This is a reasonable approach. By their very nature, evolutionary game theory and the adaptive learning literature make behavioural assumptions. The sophisticated reasoning process suggested here is an appropriate way to capture the method by which individuals

might think about each others' actions. For mathematical convenience reasoning continues indefinitely. In fact, the results of a process in which agents only iterate finitely coincide with those presented here. This is an established characteristic of human behaviour. Psychologists have shown that people are capable of interactive reasoning to a great extent; indeed, this is one of distinguishing features of the human mind. Studies such as Kinderman, Dunbar and Bentall (1998) have shown that 4 levels of reasoning are not uncommon among adults.[5]

The sophisticated agent also uses observations of their environment to initialise such an iterative reasoning process. This is payoff irrelevant information and hence there is no specific reason for a purely rational player to take heed. Nevertheless, this is a reasonable behavioural hypothesis. Despite the payoff irrelevance of historical precedent, individuals appear to attach great significance to it. Schelling (1960, Chapter 3) agrees: "Precedent seems to exercise an influence that greatly exceeds its logical importance or legal force." Focal points indicate a possible way to play and history provides the focal point. The focus of history is sharp as it describes actual previous play rather than merely being an arbitrary coordination device. The sophisticated agent goes further; instead of naïvely following in their predecessors' footsteps they are guided toward a new mode of play.

Given the high level of sophistication the agents in the model display, a natural suggestion might be the imposition of full rationality. This would not be constructive. The adaptive learning literature is in part motivated by the problem of equilibrium selection. This is precisely because the "rational" approach has been unable to provide a convincing answer. Such an imposition thus results in a return to the predictionless vacuum that this literature attempts to fill. As argued above, evolutionary game theory by its very nature makes *behavioural* assumptions. Few social scientists seriously believe that rationality (in the narrow sense often used by economists) provides a reasonable behavioural hypothesis. People do not behave "rationally"; at best it is an approximation. An alternative defence

---

[5]Certain professions are conducive to an even greater degree of recursive empathy. Novelists, for example, may be able to achieve five or more levels of interactive reasoning as they place themselves in the minds of their characters. For further details see the survey of Dunbar (1996) and the references therein.

is that agents act "as if" they are rational. By playing their part in a Nash equilibrium individuals are choosing strategies that mimic those of a rational counterpart.[6] This may justify equilibrium play; it cannot select between equilibria. Precisely because the focus here is on such selection, the remit of the "as if" approach is exceeded. Hence the only rôle for rationality is one of approximation and here the approximation is explicit. Any thought experiment that an agent might reasonably engage in will take note of their environment. In the tradition of English empiricism no inference can be made without reference to experience. To quote Hume (1764):

> In vain, therefore, should we pretend to determine any single event, or infer
>
> any cause or effect, without the assistance of observation and experience.

2.4. **An Illustrative Example.** MW (1997) outlined a simple example to illustrate the main results. This example is again employed here. In spirit, it follows a leading example of Kandori and Rob (1995). Consider a population of n academics in a research institution. All members use personal computers (PCs) to conduct their work, and may adopt either the IBM or Apple Macintosh (Mac) standards. Institution members interact during the course of their work, and receive payoffs according to their PC and that adopted by their colleagues. The strategic form game for mean payoff agents is:

	IBM	Mac
IBM	5 / 5	2 / 4
Mac	4 / 2	6 / 6

The payoffs are chosen as a stylised representation of the following criteria: Players benefit from compatibility; given compatibility, Mac adoption results in higher productivity than IBM adoption; the loss from incompatibility is less severe for IBM users than Mac users due to wider outside support for the IBM standard. This game has two pure strategy

---

[6] Aumann and Brandenburger (1995) provide a detailed account of the conditions necessary for rational players to adopt Nash equilibrium strategies.

Nash equilibria corresponding to the two standards. Although Mac is payoff dominant, IBM is risk-dominant ($x^* = \frac{2}{5} < \frac{1}{2}$). The game is thus a Rousseau (1755) stag-hunt.

Using the formal notation of Section 2.1, this becomes the trembled stage game $\mathcal{G}_{PC} = \langle \Lambda_{PC}, \Psi_{PC}, \sigma \rangle$. Of particular interest are the cases $\Psi_{PC} \in \{\Psi_B, \Psi_U\}$ where:

$$\Psi_B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \text{ and } \Psi_U = \begin{bmatrix} 0 & \frac{3}{4} \\ 2 & 0 \end{bmatrix}$$

corresponding to balanced and unbalanced respectively. The second configuration is specified to reflect the increased risk of being a lone Mac user.

Each period a randomly selected member leaves the institution and a replacement chooses their computer system. As an alternative interpretation, researchers periodically replace their PCs. At replacement time, however, it is assumed that the requirements of the researcher are different from those at the time of the original purchase, and hence payoffs are a fresh draw from the payoff distribution.

3. ANALYSIS

The analysis begins by examining the response of an idiosyncratic agent to an observed strategy frequency.

3.1. **Entrant Response and Bayesian Nash Equilibria.** When selecting a strategy a player is concerned with the strategy choice of an opponent. A sufficient statistic for a player's beliefs about an opponent's behaviour is x, the probability with which strategy 1 is played. The expected payoffs from strategies 1 and 2 when an opponent plays 1 with probability x are respectively:

$$y_1 = x\tilde{a} + (1-x)\tilde{b} = xa + (1-x)b + x\sigma\varepsilon_a + (1-x)\sigma\varepsilon_b$$

$$y_2 = x\tilde{c} + (1-x)\tilde{d} = xc + (1-x)d + x\sigma\varepsilon_c + (1-x)\sigma\varepsilon_d$$

so that:

$$y_1 - y_2 = x(a - c) - (1 - x)(d - b) + x\sigma(\varepsilon_a - \varepsilon_c) + (1 - x)\sigma(\varepsilon_b - \varepsilon_d) \qquad (1)$$

If a player believes an opponent will play strategy 1 with probability x then it is optimal to reply with the same whenever $y_1 > y_2$, or equivalently $y_1 - y_2 > 0$ — a natural trigger rule. Rearranging Equation (1) this occurs whenever:

$$x(\varepsilon_c - \varepsilon_a) + (1 - x)(\varepsilon_d - \varepsilon_b) < \frac{x(a - c) - (1 - x)(d - b)}{\sigma} \qquad (2)$$

The left hand side of Equation (2) is normally distributed with zero mean and variance $x^2 \left(\sigma_a^2 + \sigma_c^2\right) + (1 - x)^2 \left(\sigma_b^2 + \sigma_d^2\right)$. The best response to a conjecture of x will therefore be strategy 1 with probability:

$$\Pr\left[1|x\right] = \Phi\left(\frac{x(a - c) - (1 - x)(d - b)}{\sigma\sqrt{x^2\left(\sigma_a^2 + \sigma_c^2\right) + (1 - x)^2\left(\sigma_b^2 + \sigma_d^2\right)}}\right) \qquad (3)$$

where Φ represents the standard Gaussian distribution. The following is useful later on and can also be found in MW (1997).

Definition 6. *Define the* basin depth *as $\kappa(x)^2$ where $\kappa(x)$ satisfies:*

$$\kappa(x) = \frac{x(a - c) - (1 - x)(d - b)}{\sqrt{x^2\left(\sigma_a^2 + \sigma_c^2\right) + (1 - x)^2\left(\sigma_b^2 + \sigma_d^2\right)}} \qquad (4)$$

Using this notation, the above can be summarised in a convenient Lemma:

Lemma 1. *If an agent plays strategy 1 with probability x, then with probability $\rho(x, \sigma) = \Phi\left(\kappa(x)/\sigma\right)$ the optimal response will be strategy 1.*

Ignoring for the moment the observations of agents, consider the game in isolation. The static game $\mathcal{G}$ is one of incomplete information with uncertainty over the payoffs. In such a game, the Bayesian Nash equilibria correspond to fixed points of the mapping $x \rightarrow \rho(x, \sigma)$. Since these equilibria play an important rôle in the following analysis a short examination of their properties is necessary. In particular, interest lies in the case where the Harsanyian perturbations are small — $\sigma \rightarrow 0$. Lemma 2 investigates:

Lemma 2. *For σ sufficiently small there are three Bayesian Nash equilibria of $\mathcal{G}$. These correspond to fixed points of $x \mapsto \rho(x, \sigma)$ local to 0, x^* and 1. They converge to these points as $\sigma \to 0$.*

Proof. Fixed points of $\rho(x, \sigma)$ correspond to roots of:

$$f(x) = \Phi\left(\frac{\kappa(x)}{\sigma}\right) - x$$

Notice that $f'(x) = \phi(\kappa(x)/\sigma)\kappa'(x)/\sigma - 1$. When $\sigma \to 0$, $f(x) \to 1 - x$ if $x > x^*$, and $f(x) \to -x$ if $x < x^*$, so there cannot be a fixed point unless it is local to $\{0, x^*, 1\}$. Consider the interval $0 \leq x \leq \varepsilon$. For sufficiently small σ, $f(x)$ is decreasing in this interval. Moreover, $f(0) \geq 0$ and $f(\varepsilon) < 0$. Therefore there is exactly one root in this interval. It is immediate that the fixed point converges to 0 in the limit. A similar argument applies to $1 - \varepsilon \leq x \leq 1$.

Now consider $x^* - \varepsilon \leq x \leq x^* + \varepsilon$. Then $f(x^* - \varepsilon) < 0$ and $f(x^* + \varepsilon) > 0$. Again there is at least one root in this interval. $\Phi(\kappa(x)/\sigma)$ is strictly increasing. A fixed point of $\Phi(\kappa(x)/\sigma)$ corresponds to a fixed point of its inverse. Local to x^* the derivative of the inverse is less than one. This locality expands as σ gets small. Within this region there can be only one fixed point of the inverse and hence in this interval the root of $f(x)$ is unique. Convergence in the limit for $\sigma \to 0$ is again immediate. $\square$

This lemma is illustrated in Figure 1. As $\sigma \to 0$, $\Phi(\kappa(x)/\sigma)$ crosses the 45 degree line at three points local to 0, x^* and 1.

3.2. **The Reasoning Process with Full Sampling.** Lemma 2 establishes the existence and number of Bayesian Nash equilibria in $\mathcal{G}$. To select between them, an iterative best response procedure is constructed with the *sophisticated* nature of the players in mind — see Section 2.3. Note there is still no private information other than payoffs at this stage since agents obtain the same (complete) sample of the population. Hence a symmetric strategy profile can be summarised by the probability x that an individual plays strategy 1. Suppose an opponent initially is conjectured to play strategy 1
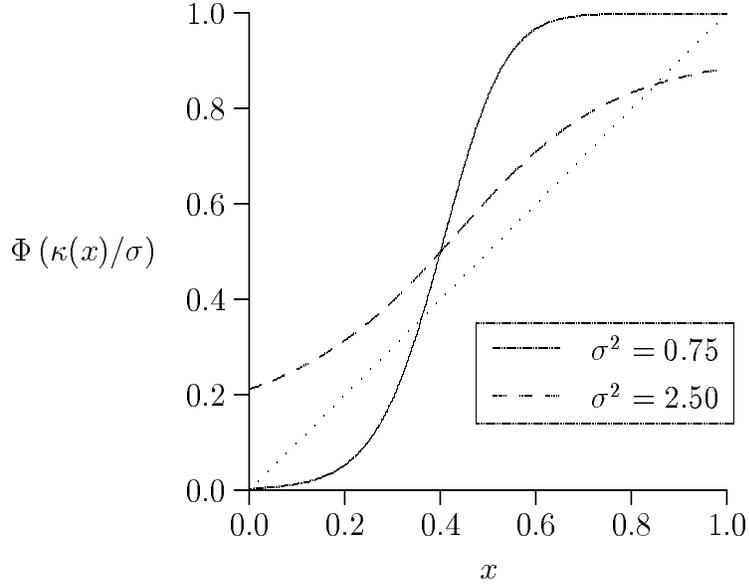
FIGURE 1. Bayesian Nash Equilibria for $\mathcal{G}_{PC}$ with $\Psi = \Psi_B$

with probability x_0. The probability strategy 1 is played as a best response to this is $x_1 = \rho(x_0, \sigma)$ by Lemma 1. The sophisticated play paradigm suggests an iterative process with $x_t = \rho(x_{t-1}, \sigma)$. Consider now the limiting behaviour of such a procedure. Further suppose that σ is small enough such that there are three Bayesian Nash equilibria of $\mathcal{G}$ local to 0, x^* and 1 by Lemma 2. Label these equilibria x_L, x_M and x_H respectively.

Lemma 3. *The limiting behaviour of the best response process satisfies:*

$$x_t \to \begin{cases} x_L & x_0 < x_M \\ x_M & x_0 = x_M \\ x_H & x_0 > x_M \end{cases}$$

The process is illustrated in Figure 2. For generic starting points the process converges to one of the two stable Bayesian Nash equilibria. So far the starting point has not been discussed. Again, due to the sophisticated reasoning process employed by the agents, observations of the current population state provide the initial conjecture; as argued in Section 2.3. With full sampling this entails an identical observation of $s = n$ individuals' actions.
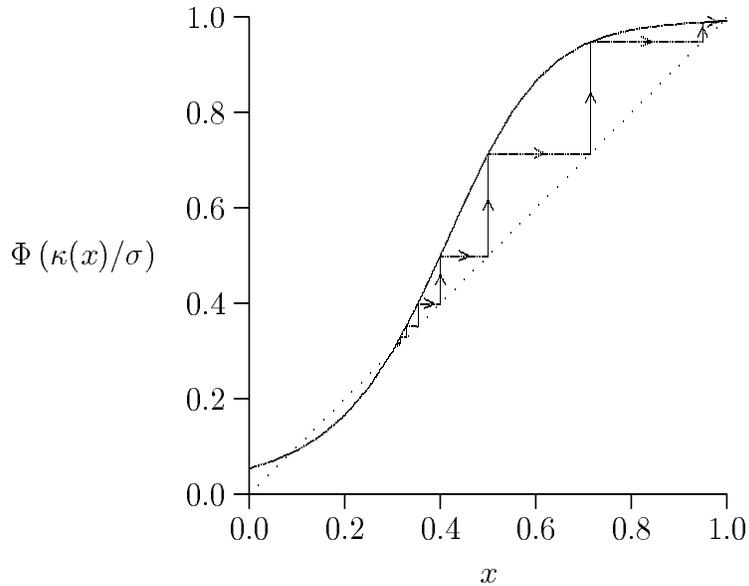
FIGURE 2. Convergence to a Stable BNE

Lemma 4. *Suppose that agents observe i individuals out of n playing strategy 1. Then for sufficiently small σ, the sophisticated reasoning process converges to x_L for $i < \lceil nx^* \rceil$ and x_H for $i \geq \lceil nx^* \rceil$.*

Proof. For σ small enough, x_M is sufficiently close to x^* to avoid discretisation problem. Otherwise follows from Lemma 3. $\square$

3.3. **The Ergodic Distribution and Equilibrium Selection.** Continuing under the assumption that there is no private information to a player aside from their payoffs ($s = n$), Lemma 4 shows that the sophisticated reasoning process induces each agent to play the same Bayesian Nash equilibria of $\mathcal{G}$. The particular equilibrium played is contingent upon the (common) history observed. Furthermore, the probability distribution over an agent's actions is governed by the strategy frequencies of the equilibrium. Hence the population state in any period determines which of the equilibria is to be played. This determines the population state in the following period via the probability distribution over strategies implied by the equilibrium. Throughout this section, σ is assumed small enough for the lemmas of Sections 3.1 and 3.2 to apply.

3.3.1. *The Markov Process.* In a 2×2 coordination game the state space can therefore be reduced to a two-state model. Recall the original state space $Z = \{0, 1, ..., n\}$ where $z \in Z$ is the number of agents playing strategy 1. Consider:

Definition 7. *For the simultaneous updating dynamic define the Markov state space as* $Z^* = \{L, H\}$ *with generic element* z^*. *The process has transition probabilities* $p_{ij} = \Pr\left[z_{t+1}^* = j \mid z_t^* = i\right]$. *The associated Markov transition matrix is:*

$$P = \begin{bmatrix} p_{LL} & p_{LH} \\ p_{HL} & p_{HH} \end{bmatrix}$$

This reduced form is equivalent to $L = \{0, 1, ..., \lceil nx^* \rceil - 1\} \subset Z$ and $H = Z - L$. Further:

$$z^* = \begin{cases} L & z < \lceil nx^* \rceil \text{ and } z \in Z \\ H & z \geq \lceil nx^* \rceil \text{ and } z \in Z \end{cases}$$

Given $z_t^* = H$, $z_t \sim \text{Bin}(x_H, n)$. It follows that:

Lemma 5. *The reduced form Markov transition probabilities satisfy:*

$$p_{LH} = \sum_{i=\lceil nx^* \rceil}^{n} \binom{n}{i} x_L^i (1 - x_L)^{n-i} \qquad p_{LL} = 1 - p_{LH}$$

$$p_{HL} = \sum_{i=0}^{\lceil nx^* \rceil - 1} \binom{n}{i} x_H^i (1 - x_H)^{n-i} \qquad p_{HH} = 1 - p_{HL}$$

Proof. State H is reached if $z_{t+1} \geq \lceil nx^* \rceil$. Strategy 1 is played with probability x_L in state L. Applying the binomial distribution formula, obtain p_{LH}. Calculate the remaining terms similarly. $\square$

3.3.2. *Characterisation of the Ergodic Distribution.* This is a particularly simple Markov process with only two states and $p_{ij} > 0 \ \forall i, j$, recall that in such cases the ergodic distribution satisfies:

$$\mu_L = \frac{p_{HL}}{p_{HL} + p_{LH}} \text{ and } \mu_H = \frac{p_{LH}}{p_{HL} + p_{LH}}$$

μ_i is the probability that the Bayesian Nash equilibria involving x_i is played in the long run. Interest lies in the relative frequencies of these two equilibria. Hence:

Lemma 6. *The relative frequency of states H and L in the ergodic distribution satisfy:*

$$\frac{\mu_H}{\mu_L} = \frac{\sum_{i=\lceil nx^* \rceil}^{n} \binom{n}{i} x_L^i \left(1 - x_L\right)^{n-i}}{\sum_{i=0}^{\lceil nx^* \rceil - 1} \binom{n}{i} x_H^i \left(1 - x_H\right)^{n-i}} \tag{5}$$

Proof. Apply Lemma 5.   □

Note that both x_L and x_H depend upon σ. The next section examines the relative ergodic frequency as idiosyncrasy vanishes, *i.e.* $\sigma \to 0$, thus enabling selection between the two available equilibria.

3.3.3. *Equilibrium Selection.* From Lemma 2, $x_L \to 0$ and $x_H \to 1$ as $\sigma \to 0$. Thus both numerator and denominator in Equation (5) tend to zero as σ vanishes. Investigation of the limiting behaviour begins with the following lemma.

Lemma 7. *The ratio in Equation (5) as $\sigma \to 0$ becomes:*

$$\lim_{\sigma \to 0} \frac{\mu_H}{\mu_L} = \frac{\lceil nx^* \rceil + 1}{n - \lceil nx^* \rceil} \times \lim_{\sigma \to 0} \frac{x_L^{\lceil nx^* \rceil}}{\left(1 - x_H\right)^{n - \lceil nx^* \rceil + 1}} \tag{6}$$

Proof. Multiplying and dividing the numerator of Equation (5) by $x_L^{\lceil nx^* \rceil}$:

$$\sum_{i=\lceil nx^* \rceil}^{n} \binom{n}{i} x_L^i \left(1 - x_L\right)^{n-i} = x_L^{\lceil nx^* \rceil} \sum_{i=\lceil nx^* \rceil}^{n} \binom{n}{i} x_L^{i - \lceil nx^* \rceil} \left(1 - x_L\right)^{n-i}$$

$$= x_L^{\lceil nx^* \rceil} \left\{ \binom{n}{\lceil nx^* \rceil} \left(1 - x_L\right)^{n - \lceil nx^* \rceil} + \sum_{i=\lceil nx^* \rceil + 1}^{n} \binom{n}{i} x_L^{i - \lceil nx^* \rceil} \left(1 - x_L\right)^{n-i} \right\}$$

As $\sigma \to 0$, $x_L \to 0$ and hence:

$$\sum_{i=\lceil nx^* \rceil + 1}^{n} \binom{n}{i} x_L^{i - \lceil nx^* \rceil} \left(1 - x_L\right)^{n-i} \to 0 \text{ and } \binom{n}{\lceil nx^* \rceil} \left(1 - x_L\right)^{n - \lceil nx^* \rceil} \to \binom{n}{\lceil nx^* \rceil}$$

Performing a similar operation on the denominator to obtain:

$$\lim_{\sigma \to 0} \frac{\mu_H}{\mu_L} = \frac{n!}{(n - \lceil nx^* \rceil)! \lceil nx^* \rceil!} \frac{(n - \lceil nx^* \rceil - 1)! (\lceil nx^* \rceil + 1)!}{n!} \times \lim_{\sigma \to 0} \frac{x_L^{\lceil nx^* \rceil}}{(1 - x_H)^{n - \lceil nx^* \rceil + 1}}$$

$$= \frac{\lceil nx^* \rceil + 1}{n - \lceil nx^* \rceil} \times \lim_{\sigma \to 0} \frac{x_L^{\lceil nx^* \rceil}}{(1 - x_H)^{n - \lceil nx^* \rceil + 1}}$$

Which is the desired result. $\square$

Some terminology is useful before proceeding to the calculation of the ergodic distribution for vanishing heterogeneity.

Definition 8. *Strategy* 1 dominates for vanishing heterogeneity *if* $\lim_{\sigma \to 0} \mu_H / \mu_L = +\infty$. *Strategy* 2 dominates for vanishing heterogeneity *if* $\lim_{\sigma \to 0} \mu_H / \mu_L = 0$.

When μ_H / μ_L is large, the Bayesian Nash equilibrium involving probabilities x_H and $1 - x_H$ is played most of the time. For vanishing σ this equilibrium entails agents playing strategy 1 with high probability.

Proposition 1. *If* $(n - \lceil nx^* \rceil + 1) \kappa (1)^2 > \lceil nx^* \rceil \kappa (0)^2$ *then strategy* 1 *dominates for vanishing idiosyncrasy. If the reverse holds then strategy* 2 *dominates.*

Proof. Recall that x_i satisfies the equality $x_i = \rho (x_i, \sigma) = \Phi (\kappa (x_i) / \sigma)$. Substituting into the second term on the right hand side of Equation (6):

$$\lim_{\sigma \to 0} \frac{x_L^{\lceil nx^* \rceil}}{(1 - x_H)^{n - \lceil nx^* \rceil + 1}} = \lim_{\sigma \to 0} \frac{\Phi (\kappa (x_L) / \sigma)^{\lceil nx^* \rceil}}{(1 - \Phi (\kappa (x_H) / \sigma))^{n - \lceil nx^* \rceil + 1}}$$

The first term is just a constant and hence is irrelevant. Separating the above into normal densities and hazards rates the following obtains:

$$\lim_{\sigma \to 0} \left(\frac{\phi (\kappa (x_H) / \sigma)}{1 - \Phi (\kappa (x_H) / \sigma)} \right)^{n - \lceil nx^* \rceil + 1} \left(\frac{\Phi (\kappa (x_L) / \sigma)}{\phi (\kappa (x_L) / \sigma)} \right)^{\lceil nx^* \rceil} \frac{\phi (\kappa (x_L) / \sigma)^{\lceil nx^* \rceil}}{\phi (\kappa (x_H) / \sigma)^{n - \lceil nx^* \rceil + 1}}$$

Now as $\sigma \to 0$, $\kappa (x_L) \to \kappa (0) < 0$ and $\kappa (x_H) \to \kappa (1) > 0$. The limit becomes:

$$\lim_{\sigma \to 0} \left(\frac{\phi (\kappa (1) / \sigma)}{1 - \Phi (\kappa (1) / \sigma)} \right)^{n - \lceil nx^* \rceil + 1} \left(\frac{\Phi (\kappa (0) / \sigma)}{\phi (\kappa (0) / \sigma)} \right)^{\lceil nx^* \rceil} \frac{\phi (\kappa (0) / \sigma)^{\lceil nx^* \rceil}}{\phi (\kappa (1) / \sigma)^{n - \lceil nx^* \rceil + 1}}$$

Note that $\kappa\left(0\right)/\sigma \to -\infty$ and $\kappa\left(1\right)/\sigma \to +\infty$ as $\sigma \to 0$. The hazard rate of the normal is asymptotically linear. Therefore the first two terms in the above expression are asymptotically polynomial. The third term is exponential however and dominates in the limit. Examining this crucial term:

$$\frac{\phi\left(\kappa\left(0\right)/\sigma\right)^{\lceil nx^* \rceil}}{\phi\left(\kappa\left(1\right)/\sigma\right)^{n-\lceil nx^* \rceil+1}} \propto \exp\left\{-\frac{\lceil nx^* \rceil \kappa\left(0\right)^2 - \left(n - \lceil nx^* \rceil + 1\right)\kappa\left(1\right)^2}{2\sigma^2}\right\}$$

It follows that:

$$\lim_{\sigma \to 0} \frac{\phi\left(\kappa\left(0\right)/\sigma\right)^{\lceil nx^* \rceil}}{\phi\left(\kappa\left(1\right)/\sigma\right)^{n-\lceil nx^* \rceil+1}} = +\infty \Leftrightarrow \left(n - \lceil nx^* \rceil + 1\right)\kappa\left(1\right)^2 > \lceil nx^* \rceil \kappa\left(0\right)^2$$

i.e. strategy 1 dominates for vanishing idiosyncrasy; the desired result.   □

It is then straightforward to establish the following key result:

Proposition 2. *If a strategy is risk-dominant and generalised risk-dominant then it dominates for vanishing idiosyncrasy.*

Proof. Assume strategy 1 is risk-dominant and generalised risk-dominant (the proof for strategy 2 is symmetric). Using Definition 6:

$$\kappa\left(0\right)^2 = \frac{\left(b - d\right)^2}{\sigma_b^2 + \sigma_d^2} \text{ and } \kappa\left(1\right)^2 = \frac{\left(a - c\right)^2}{\sigma_a^2 + \sigma_c^2}$$

If strategy 1 is generalised risk-dominant (see Definition 4) then $\kappa\left(1\right)^2 > \kappa\left(0\right)^2$. If strategy 1 is risk-dominant then $x^* < \frac{1}{2}$, so $n\left(1 - x^*\right) > nx^*$ and hence $\left(n - \lceil nx^* \rceil + 1\right) > \lceil nx^* \rceil$. Therefore by Proposition 1 strategy 1 dominates for vanishing idiosyncrasy.   □

3.3.4. *State Dependent Mutations.* Mutations do not arise in this model, in the sense that no mistakes are actually made by the players. However, agents have idiosyncratic preferences and so it may appear to the modeller as if mistakes take place. "Mutations" are observed when agents act against the flow of play. That is, an entrant takes a contrarian action relative to a *mean payoff entrant*. Since all agents are playing one of two Bayesian Nash equilibria this will occur with probability x_L in any state $z \in$

$\{0, 1, ..., \lceil nx^* \rceil - 1\}$ and $1 - x_H$ in any state $z \in \{\lceil nx^* \rceil, ..., n\}$, (see Definition 7 and Lemma 5).

Mutations are state dependent when they vanish to zero at different rates. Here, "mutations" arise endogenously and are a function of idiosyncrasy, σ, which vanishes to zero independently of the state. Nevertheless this model generates state dependence as Corollary 1 below illustrates. Bergin and Lipman (1996) have shown that state dependent mutations can result in the selection of *any* Nash equilibrium and hence argued for more convincing explanations for "mutations". The sophisticated play paradigm is precisely such an explanation. The corollary follows from Proposition 1.

Corollary 1. *The model endogenously generates state dependent mutations.*

Proof. Consider the limit of the ratio of "mutation" rates:

$$\lim_{\sigma \to 0} \frac{x_L}{1 - x_H} = \lim_{\sigma \to 0} \frac{\Phi\left(\kappa\left(x_L\right)/\sigma\right)}{1 - \Phi\left(\kappa\left(x_H\right)/\sigma\right)} = \lim_{\sigma \to 0} \exp\left\{-\frac{\kappa\left(0\right)^2 - \kappa\left(1\right)^2}{2\sigma^2}\right\}$$

Where the first equality follows from the fact that x_i satisfies $x_i = \rho\left(x_i, \sigma\right) = \Phi\left(\kappa\left(x_i\right)/\sigma\right)$, and the second from an analogous argument involving normal densities and hazards to that in Proposition 1. Finally note that this limit is either zero or infinity for generic games. $\qquad\square$

Notice that the state dependence in this model is of a much simpler nature than that of MW (1997). In that paper the "mutations" vanish to zero at different rates in every state. Here there are only two different rates: The *depth* of each basin of attraction is constant.

3.4. **The Reasoning Process with Incomplete Sampling.** With incomplete sampling, players may observe different strategy frequencies from the population. All agents obtain samples of size s. Hence there are $s + 1$ possible observations. Index these by $i \in S = \{0, 1, ..., s\}$, where the generic element i is the number of individuals seen to be playing strategy 1. An agent's strategy is a mapping from their payoffs and sample to the probability with which they play strategy 1.

Given their beliefs about an opponent's play, agents are restricted to consider only best responses. In Section 3.1 the optimal response of an agent who believes an opponent plays strategy 1 with probability x was shown to entail playing strategy 1 with probability $\rho(x, \sigma)$.

Since payoffs are independently distributed, beliefs are allowed to depend only on the sample observed.[7] $\rho(x, \sigma)$ is the probability with which strategy 1 is a best response. All that remains is to specify the beliefs of the agents contingent on the sample they observe.

Represent a player's beliefs as a vector $\gamma \in [0, 1]^{s+1}$, where the ith element, $\gamma_i \in [0, 1]$, is the probability with which an opponent is believed to play strategy 1 given the player has observed i out of s individuals playing strategy 1. This is the appropriate state variable for the reasoning process.

Therefore, given a belief vector, a player observing a sample of $i \in S$ will play strategy 1 with probability $\rho(\gamma_i, \sigma)$. The value of γ is determined by the sophisticated play paradigm: Players anticipate best responses in the population and construct their belief profiles iteratively.

Definition 9. *q_{ij} is the probability of an opponent observing a sample of $j \in S$ given the agent has observed a sample of $i \in S$.*

q_{ij} is determined by the type of sampling procedure the agents use and their prior beliefs over the population states. Two important examples of sampling procedure are considered in Section 3.6; uniform sampling with and without replacement. Of course, there are other possible procedures and for this reason the specification of q_{ij} is left open at this stage. The following Lemma is immediate.

Lemma 8. *If agents initially hold belief γ, then after one iteration of best response they will hold updated belief $\widetilde{\gamma}$, where:*

$$\widetilde{\gamma}_i = \sum_{j=0}^{s} q_{ij} \rho(\gamma_j, \sigma)$$

---

[7]Beliefs are not contingent on a players' identity; agents only have an "identity" in as far as they have idiosyncratic payoffs and samples.

Assembling into vector notation, define $\rho\left(\gamma,\sigma\right)\in[0,1]^{s+1}$ as the vector with ith element $\rho\left(\gamma_i,\sigma\right)$ and Q as the matrix with (i,j)th element q_{ij}.

Definition 10. *The iterative reasoning mapping is* $\widetilde{\gamma}:[0,1]^{s+1}\longmapsto[0,1]^{s+1}$, *where:*

$$\widetilde{\gamma}\left(\gamma\right)=Q\rho\left(\gamma,\sigma\right)$$

The iterative reasoning mapping extends the best response process of Section 3.2. The difference is that play may be contingent on an agent's observation, and because of this players take into account the fact that others may have observed samples at odds with their own. The sophisticated play paradigm suggests that (i) the process should be iterated until it converges and (ii) the starting point for the process should be determined by the observations. Formally:

Definition 11. *The sophisticated reasoning process* $\{\gamma^t\}$ *is constructed with* $\gamma^t=\widetilde{\gamma}\left(\gamma^{t-1}\right)$, *where the starting point,* $\gamma^{\mathbf{0}}$, *satisfies:*

$$\gamma_i^{\mathbf{0}}=\rho\left(\frac{i}{s},\sigma\right)$$

3.5. **Equilibrium Selection with Incomplete Sampling.** The limiting behaviour of the sophisticated reasoning process determines the equilibrium to be played. The main proposition in this section gives sufficient conditions on the sampling procedure (and hence Q) for selection of the risk-dominant equilibrium.

If the sophisticated reasoning process converges, so that $\lim_{t\to\infty}\gamma^t=\gamma^*$, then the agents play a particular Bayesian Nash equilibrium of the trembled stage game with sampling. In such an equilibrium, an updating individual plays strategy 1 with probability $\sum_{i=0}^s\Pr\left[i\mid z\right]\gamma_i^*$, where $\Pr\left[i\mid z\right]$ is the probability of observing a sample $i\in S$ given the population state $z\in Z$. Hence the appropriate Markov transition probabilities can be constructed. Notice, however, that if $\gamma_i^*=\gamma_j^*$ for all $i\neq j$, then the updating agents play the same Bayesian Nash equilibrium regardless of their samples, or indeed the *population state* itself. Hence conditions are sought under which this may occur.

Since interest lies in the case of vanishing idiosyncrasy, the analysis below centres upon $\sigma = 0$. With incomplete sampling, equilibrium selection in this sense does not require an examination of the perturbed stage game. The sophisticated reasoning process itself is enough to select a unique equilibrium. Nevertheless, in Section 4.2 the issue of $\sigma > 0$ is discussed as it is of independent interest.

Condition 1. *When a player observes $i < \lceil sx^* \rceil$ their belief that an opponent has observed $j > i$ must be greater than x^*:*

$$x^* < \min_{i < \lceil sx^* \rceil} \left\{ \sum_{j=i+1}^{s} q_{ij} \right\} \tag{7}$$

Condition 2. *The sampling procedure must satisfy a minimal first order stochastic dominance property, or more weakly:*

$$x^* < \min_{i \geq \lceil sx^* \rceil} \left\{ \sum_{j=\lceil sx^* \rceil}^{s} q_{ij} \right\} \tag{8}$$

The first condition is key. The right hand side of Equation (7) is a measure of how "dispersed" the Q matrix is. This in turn reflects the agents' beliefs about opponents' observations given their own. Players must place sufficiently high weight on the possibility of an opponent having observed a sample higher than they received themselves.

Proposition 3. *If Conditions 1 and 2 hold then for all $i \in S$, $\lim_{t \to \infty} \gamma_i^t = 1$.*

Proof. A best response to a belief of $\gamma_i \geq x^*$ is to play strategy 1. So, write $\rho(x, 0) = I(x \geq x^*)$ where I is an indicator function. The proof proceeds by showing that $\gamma_i^t \geq x^*$ for $j \geq \lceil sx^* \rceil - t$ and $t \leq \lceil sx^* \rceil$. Recall that $\gamma_i^0 = \rho\left(\frac{i}{s}, 0\right)$, then $\gamma_i^0 = 1$ for $i \geq \lceil sx^* \rceil$. Hence the hypothesis holds for $t = 0$, yielding an induction basis. Now suppose that the hypothesis holds for some $\tau < \lceil sx^* \rceil$, so that $\gamma_i^\tau \geq x^*$ for $i \geq \lceil sx^* \rceil - \tau$. Notice that:

$$\gamma_i^{\tau+1} = \sum_{j=0}^{s} q_{ij} \rho\left(\gamma_j^\tau, 0\right) = \sum_{j=0}^{s} q_{ij} I\left(\gamma_j^\tau \geq x^*\right) \geq \sum_{j=\lceil sx^* \rceil - \tau}^{s} q_{ij}$$

First consider $\lceil sx^* \rceil - \tau - 1 \leq i < \lceil sx^* \rceil$. Then:

$$\gamma_i^{\tau+1} \geq \sum_{j=\lceil sx^* \rceil - \tau}^{s} q_{ij} \geq \sum_{j=i+1}^{s} q_{ij} \geq \min_{i < \lceil sx^* \rceil} \left\{ \sum_{j=i+1}^{s} q_{ij} \right\} > x^*$$

Where the last inequality holds by Condition 1's Equation (7). For $i \geq \lceil sx^* \rceil$:

$$\gamma_i^{\tau+1} \geq \sum_{j=\lceil sx^* \rceil - \tau}^{s} q_{ij} \geq \sum_{j=\lceil sx^* \rceil}^{s} q_{ij} \geq \min_{i \geq \lceil sx^* \rceil} \left\{ \sum_{j=\lceil sx^* \rceil}^{s} q_{ij} \right\} > x^*$$

By Equation (8) in Condition 2. Thus, $\gamma_i^{\tau+1} > x^*$ for $i \geq \lceil sx^* \rceil - \tau - 1$. By the principle of induction $\gamma_i^t \geq x^*$ for $i \geq \lceil sx^* \rceil - t$ and $t \leq \lceil sx^* \rceil$. In particular, this holds for $\hat{t} = \lceil sx^* \rceil$. Hence, $\gamma_i^{\lceil sx^* \rceil} \geq x^*$. But then, for all $t > \hat{t}$, $\gamma_i^t = 1$ for all i. Therefore, $\lim_{t \to \infty} \gamma_i^t = 1$ for all $i \in S$. $\square$

This proof employs an infection argument analogous to that in Morris, Rob and Shin (1995). This point is returned to in more detail in Section 4.2. Two important remarks follow from this proof.

Remark 1. *If strategy 1 is sufficiently risk-dominant then all agents will play strategy 1 regardless of the samples they observe.*

In other words, for (almost) any sampling procedure, there is always an x^* small enough such that the Conditions 1 and 2 hold and hence Proposition 3 applies. Of course, for reasonable assumptions on the sampling procedure the conditions can hold for any $x^* < \frac{1}{2}$, as will be shown in the next section.

Remark 2. $\lceil sx^* \rceil$ *is an upper bound on the number of levels of reasoning required to converge to the equilibrium.*

As mentioned earlier (see Section 2.3), psychological research has focused on the number of iterations humans are capable of — and it is not many. Here, there is no need to assume

the mathematical nicety of infinite reasoning capacity. In the incomplete sampling case equilibrium is reached in finite time.[8]

3.6. **Sampling Procedures.** If Conditions 1 and 2 are satisfied then all players adopt the same strategy. Notice that no evolutionary considerations are required — reasoning alone enables selection. A sufficiently risk-dominant strategy will always be selected by the sophisticated reasoning process. More formally, fix a sampling procedure and associate matrix Q then for x^* sufficiently small, strategy 1 will be selected. A related question is raised in this section: Fixing x^*, what properties of the sampling procedure are required to result in the selection of the risk-dominant equilibrium? Conditions 1 and 2 are sufficient and hence this section concentrates on whether two reasonable sampling procedures (with and without replacement) satisfy these requirements.

3.6.1. *Sampling with Replacement.* Agents observe a sample of size s, with replacement. The probability of observing an agent playing strategy 1 is $z/n = p$.[9] Therefore the number of agents playing strategy 1 in an individuals sample is distributed binomially with parameters p and s. The probability they observe exactly i strategy 1 players is:

$$\Pr\left[i \mid p\right] = \binom{i}{s} p^i \left(1 - p\right)^{s-i}$$

Players do not know the population state and thus have priors over p (or z) , $g\left(p\right)$, with distribution function $G\left(p\right)$. So:

$$\Pr\left[i\right] = \int_p \Pr\left[i \mid p\right] dG\left(p\right)$$

Calculating the elements of Q, the q_{ij} are given by:

$$q_{ij} = \frac{\Pr\left[i \cap j\right]}{\Pr\left[i\right]} = \frac{\int_p \Pr\left[i \mid p\right]\Pr\left[j \mid p\right] dG\left(p\right)}{\int_p \Pr\left[i \mid p\right] dG\left(p\right)}$$

---

[8]This time could be a very short period indeed. $\lceil sx^* \rceil$ can be very small with reasonable sample sizes, and in any case it is a rather *loose* upper bound.
[9]p will be the standard notation throughout this section, simply for convenience.

For concreteness assume the prior is Beta distributed with parameters β_1 and β_2:[10]

$$g\left(p\right) = \frac{\Gamma\left(\beta_1 + \beta_2\right)}{\Gamma\left(\beta_1\right)\Gamma\left(\beta_2\right)} p^{\beta_1 - 1}\left(1 - p\right)^{\beta_2 - 1}$$

The Beta is technically convenient as well as simplifying nicely to a symmetric prior $(\beta_1 = \beta_2)$ and, as a special case, the uniform $(\beta_1 = \beta_2 = 1)$. It is then easily shown that a closed form can be obtained for the elements of the Q matrix.

Proposition 4. *With a Beta prior (and n sufficiently large) the elements of Q are:*

$$q_{ij} = \frac{\Gamma\left(s + 1\right)}{\Gamma\left(j + 1\right)\Gamma\left(s - j + 1\right)} \frac{\Gamma\left(\beta_1 + \beta_2 + s\right)}{\Gamma\left(\beta_1 + i\right)\Gamma\left(\beta_2 + s - i\right)} \frac{\Gamma\left(\beta_1 + i + j\right)\Gamma\left(\beta_2 + 2s - i - j\right)}{\Gamma\left(\beta_1 + \beta_2 + 2s\right)}$$

Proof. The proof is contained in Appendix A. □

Therefore, with the uniform distribution the elements reduce to:

$$q_{ij} = \frac{s + 1}{2s + 1} \frac{\binom{s}{i}\binom{s}{j}}{\binom{2s}{i+j}} \tag{9}$$

Using this formulation Appendix A reports the results of numerical calculations to find (for a range of sample sizes, s) a lower bound on the maximum value of x^* for which Conditions 1 and 2 hold. This lower bound rapidly approaches $1/2$ as s increases. As x^* is increased a higher sample size s is required to satisfy the conditions. For a given x^*, the larger the sample size the more likely the risk-dominant equilibrium is selected immediately.

In conclusion, this section combined with Appendix A demonstrates that sampling with replacement is one procedure that can satisfy the conditions of Section 3.5.

3.6.2. *Sampling without Replacement.* An alternative to the procedure of Section 3.6.1 is the slightly more intuitive case of sampling without replacement. This is the scenario Young (1993) investigates. Players observe a random sample of size s consisting of different individuals' strategy choices. In a population of size n there are z agents playing

---

[10]The Beta is a continuous distribution and hence n needs to be large enough for a this to be a reasonable approximation. This is assumed throughout, although it is stated in Proposition 4.

strategy 1. The number of strategy 1 observations a player makes is then hypergeomet-
rically distributed with parameters z, n and s. The probability of observing exactly i
agents playing strategy 1 is:

$$\Pr\left[i \mid z\right] = \frac{\binom{z}{i}\binom{n-z}{s-i}}{\binom{n}{s}}$$

Retaining the assumptions of Section 3.6.1 concerning the prior once again yields q_{ij}.
Hypergeometric probabilities are well approximated by Binomial probabilities for suffi-
ciently large n/s. Hence, by Proposition 4 and the numerical results of Appendix A, with
n, n/s and s sufficiently large Conditions 1 and 2 are satisfied for a given $x^* < 1/2$.

The size of n/s is the only additional concern for sampling without replacement. The
condition that this ratio be large is reminiscent of Young (1993). Sample and population
sizes must be large, but the total number of individuals in the population must still dwarf
the number observed by any agent.

4. DISCUSSION

The results are briefly discussed with reference to the example introduced in Section 2.4.

4.1. **Full Sampling.** The model endogenously generates state-dependent mutations (see
Figure 3). From the results of Bergin and Lipman (1996) selection need not focus on the
risk-dominant equilibrium. However, despite this possibility, the emphasis remains upon
risk-dominance. Indeed, the tight selection results described in Section 3.3.3 corroborate
the earlier research of MW (1997) with a continued rôle for both risk-dominance and
generalised risk-dominance. In fact, as proven in Proposition 2, if an equilibrium has
both these properties then it *is* selected for vanishing heterogeneity.

Turning to the example of Section 2.4, $x^* = \frac{2}{5}$ and hence the IBM equilibrium is risk-
dominant. With balanced trembles (Ψ_B in the example) notice that Definition 4 reduces
to standard risk-dominance. Hence an equilibrium which is risk-dominant will always be
generalised risk-dominant and thus will be selected. For unbalanced trembles (Ψ_U) this
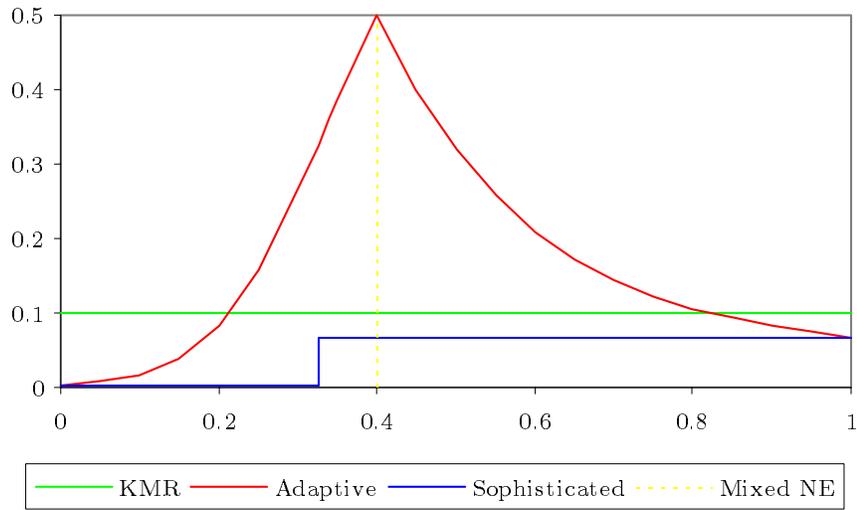
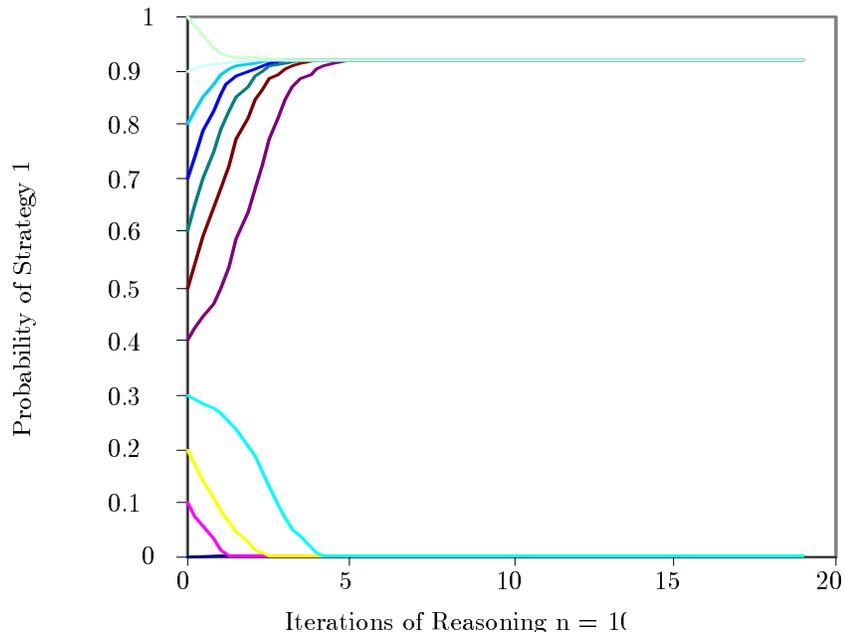FIGURE 3. Endogenously Generated "Mutations"

FIGURE 4. Full Sampling $s = 10$.

need not be the case. The Mac equilibrium will be selected in favour of IBM, as a short inspection of the selection criterion in Proposition 1 will verify.

Given some $\sigma > 0$ players will reason their way to one of the two stable Bayesian Nash equilibria. Figure 4 illustrates this process. Which of the equilibria they play depends upon their starting point — provided by a common observation of history. In the figure, players rather quickly converge given any possible initial conjecture. Hence the two basins of attraction for the equilibria are characterised by a different uniform rate of "mutation", see Figure 3. In the diagram, the first basin — $[0, x_M)$ — is narrower, but has a much smaller "mistake probability", determined by the chance of playing IBM (strategy 1) in the Bayesian Nash equilibrium which involves playing Mac (strategy 2) with high probability. The second basin — $(x_M, 1]$ — is wider and has a relatively high probability of "mistakes".[11] The depth and width of the basins of attraction determine selection. Having played their part in a Bayesian Nash equilibrium their actions become part of history and the whole process repeats itself. As $\sigma \to 0$, the "mutations" vanish *at different rates*, the Bayesian equilibria become the two pure strategy Nash equilibria and selection is achieved.

The similarities between the results here and those of MW (1997) are striking. Firstly, a risk-dominant and generalised risk-dominant equilibrium will be selected for vanishing heterogeneity. Secondly, when $\sigma > 0$, MW (1997) concludes that the modes of the ergodic distribution correspond to the Bayesian Nash equilibria of the underlying stage game. Here, this fact is trivial. Agents reason their way to Bayesian Nash equilibria via the sophisticated deliberation process. The ergodic distribution consists of two atoms, one at each of the equilibria.

4.2. **Incomplete Sampling.** With sufficiently incomplete sampling there is no need to rely upon evolutionary game theoretic arguments. Equilibrium selection is obtained without recourse to limiting results and the introduction of "mutations" is unnecessary. Fixing $\sigma = 0$, the sophisticated thought process alone allows agents to reason their way to the risk-dominant equilibrium. Convergence takes place instantaneously.

---

[11]Figure 3 also illustrates equivalent "mutation" rates for the same σ in the MW (1997) case and for the KMR (1993) dynamic — which of course is not state-dependent.

Conditions 1 and 2 determine precisely *how* incomplete sampling needs to be for any particular game and prior. Agents must place sufficiently high weight on the possibility that others have observed quite different samples from themselves. Given that they have observed a sample of $i < \lceil sx^* \rceil$ agents playing the risk-dominant strategy, they must believe it more than likely that another agent has observed $j > i$.[12] Since this is true for all players, they can infer that each agent will believe it more than likely that the others have observed a $k > j$, and so on. Through this consideration of the deliberations of others, agents will eventually reach a point at which they consider it sufficiently likely that their prospective opponent will play the risk-dominant strategy to make their optimal response the same (with high probability). Once this point is reached both players begin (through the iterative thought process) to believe it more and more likely that their opponent will play the risk-dominant strategy. Condition 2 guarantees this. Eventually (and in finite iterations), all players find it optimal to play the risk-dominant strategy with probability 1. Figure 5 depicts this process.[13]

Notice that initial observations are ignored. As soon as the iterative procedure begins, an infection process takes the posteriors to the belief that the prospective opponent observed all risk-dominant players. Proposition 3 employs this infection argument formally — and is described in the above paragraph. Morris, Rob and Shin (1995) use a similar method which, although subtly different, bears a strong resemblance (at least in form) to that presented above.

In Section 3.6.1 and Appendix A a particular sampling process is shown to abide by Conditions 1 and 2, given a particular prior. Other reasonable sampling procedures will also satisfy the conditions. Section 3.6.2 argues that sampling without replacement (perhaps the most intuitively appealing procedure) presents no problems. The key requirement here is that although s must be large, the population must still dwarf the sample

---

[12]Condition 1 states this formally.

[13]$\sigma > 0$ merely for illustrative purposes. However, this does demonstrate that the results will go through in general when there are "mutations", although this case is of less interest. If σ is sufficiently small it is straightforward that the conclusions hold. For σ sufficiently large recall there is only one Bayesian Nash equilibrium of $\mathcal{G}$ and hence selection is of no importance.
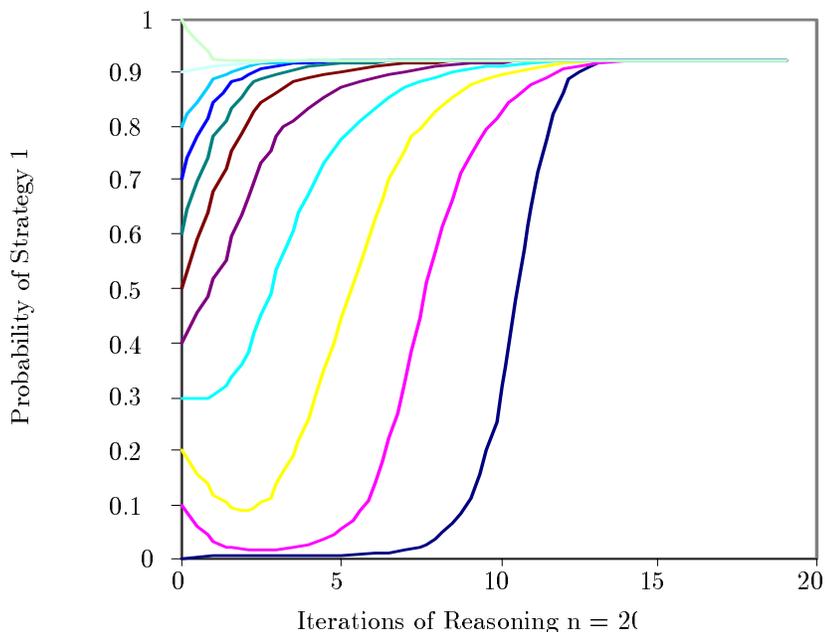
FIGURE 5. Effects of Sampling $s = 10$.

size. Again, sampling needs to be sufficiently incomplete. This constraint is reminiscent of Young (1993) — of course, here, when this proviso is met, evolution and mutation are not required to select the risk-dominant equilibrium. Naturally it will be possible to generate sampling procedures and parameters for which the conditions fail. However, both the examples presented are common and easily justified, and the restrictions on the matrix Q are themselves not unreasonable.

5. CONCLUSION

Focal points often provide a way to select between strict equilibria. However, they can be arbitrary and game specific. The focal point here is provided by history — the *context* in which the game is played. Players do not imitate history or play a naïve best response. Rather they make an initial conjecture — that their opponent will play a best response — and consider what to do in such an event. Nor does their deliberative process stop there, they continue to make iterative conjectures in a *sophisticated* manner. This procedure leads them to an action which then becomes part of history. Context evolves.

With such a framework in mind, the conclusions unsurprisingly rest upon the completeness of information available to agents. When each individual observes the entire population in their sample, an exact equilibrium selection criterion is found. Using the sophisticated play paradigm expounded above, agents reason their way to Bayesian Nash equilibria. In the classic Stag-Hunt game, which equilibrium strategy is played depends on their initial observation of history. As the idiosyncratic nature of the population is reduced to zero, Bayesian equilibria become Nash and selection takes place. If a strategy is both *risk-dominant* and *generalised risk-dominant* then it is selected.

Agents are more likely to obtain only incomplete information concerning their environment. A complete characterisation is not available in such a scenario. Nevertheless, much can be deduced. There is no need for the introduction of artificial "mutations" to enable selection. In fact, with sufficiently incomplete sampling — and hence only partial information — the risk-dominant equilibrium is selected immediately. Players ignore history altogether in an effort to coordinate with their opponents. Two conditions on the game and sampling procedure suffice to ensure this result. For two common sampling procedures with reasonable constraints on the population and sample sizes, these conditions hold. Further, if a strategy is sufficiently risk-dominant it is selected. Finally, it only takes an agent a finite (and small) number of iterations of the form "I believe that you believe that I believe..." to find it optimal to play the risk-dominant strategy.[14] Or perhaps, at this point, the reader shares the sentiments of Dmitri Nekhlyudov:

> "[...] Now suppose that they suppose that you suppose that... In a word," he concluded, feeling that his argument was getting tangled up, "it's much better not to suppose anything at all."

<div align="right">Tolstoy: *Youth*</div>

---

[14]Moreover thought processes of this type are *all* that is required for selection to take place — "mutations", evolution and observations of history are rendered irrelevant.

REFERENCES

AUMANN, R., AND A. M. BRANDENBURGER (1995): "Epistemic Conditions for Nash Equilibrium," *Econometrica*, 63(4), 1161–80.

BERGIN, J., AND B. LIPMAN (1996): "Evolution with State-Dependent Mutations," *Econometrica*, 64(4), 943–956.

BINMORE, K. (1994): *Game Theory and the Social Contract Volume 1: Playing Fair*. MIT Press, Cambridge, MA.

DUNBAR, R. (1996): *Grooming, Gossip and the Evolution of Language*. Faber and Faber, London.

HARSANYI, J., AND R. SELTEN (1988): *A General Theory of Equilibrium Selection in Games*. MIT Press, Cambridge MA.

HUME, D. (1764): "An Enquiry Concerning Human Understanding," in *Essays and Treatises on Several Subjects*. Millar, Kincaid and Donaldson, London and Edinburgh.

KANDORI, M., G. J. MAILATH, AND R. ROB (1993): "Learning, Mutation and Long-Run Equilibria in Games," *Econometrica*, 61(1), 29–56.

KANDORI, M., AND R. ROB (1995): "Evolution of Equilibria in the Long Run: A General Theory and Applications," *Journal of Economic Theory*, 65, 383–414.

KINDERMAN, P., R. DUNBAR, AND R. BENTALL (1998): "Theory of Mind Deficits and Causal Attributions," *British Journal of Psychology*, 89, 191–204.

MORRIS, S., R. ROB, AND H. S. SHIN (1995): "p-Dominance and Belief Potential," *Econometrica*, 63(1), 145–157.

MYATT, D. P., AND C. WALLACE (1997): "Adaptive Dynamics with Payoff Heterogeneity," Discussion paper, Oxford, Nuffield College Discussion Paper No. 130.

POPE, A. (1733): *An Essay on Man: Epistle I*. Printed for J. Wilford, London.

ROUSSEAU, J.-J. (1755): *Discours sur L'origine et Les Fondemens de L'inégalité Parmi Les Hommes*. Chez Marc Michel Rey, Amsterdam.

SCHELLING, T. C. (1960): *The Strategy of Conflict*. Harvard University Press, Cambridge MA.

TOLSTOY, L. (1857): "Youth," in *Childhood, Boyhood, Youth*, ed. by R. Edwards, pp. 179–319. Penguin Classics, London, 1988.

YOUNG, H. P. (1993): "The Evolution of Conventions," *Econometrica*, 61(1), 57–84.

APPENDIX A. OMITTED PROOFS AND NUMERICAL CALCULATIONS

Section A.1 provides the proof of Proposition 4. Section A.2 illustrates and describes the numerical calculations referred to throughout Section 3.6.

A.1. **Proof of Proposition 4.** Allowing n sufficiently large for the prior to be well approximated by a Beta distribution, note that:

$$\Pr(p \mid i) = \frac{g(p) \Pr(i \mid p)}{\int_0^1 \Pr(i \mid x) g(x) \, dx}$$

$$\propto g(p) \Pr(i \mid p)$$

$$\propto p^{\beta_1 - 1} (1-p)^{\beta_2 - 1} p^i (1-p)^{s-i}$$

$$\propto p^{\beta_1 + i - 1} (1-p)^{\beta_2 + s - i - 1}$$

Which is a Beta distribution with parameters $\beta_1 + i$ and $\beta_2 + s - i$. The multiplicative factor to ensure this probability integrates to one can be inserted to give an exact form for $\Pr(p \mid i)$. Hence:

$$\Pr(p \mid i) = \frac{\Gamma(\beta_1 + \beta_2 + s)}{\Gamma(\beta_1 + i) \Gamma(\beta_2 + s - i)} p^{\beta_1 + i - 1} (1-p)^{\beta_2 + s - i - 1}$$

Now the elements of Q are:

$$q_{ij} = \Pr(j \mid i) = \int_0^1 \Pr(j \mid p) \Pr(p \mid i) \, dp$$

$$= \frac{s!}{j!(s-j)!} \frac{\Gamma(\beta_1 + \beta_2 + s)}{\Gamma(\beta_1 + i) \Gamma(\beta_2 + s - i)} \int_0^1 p^{\beta_1 + i + j - 1} (1-p)^{\beta_2 + 2s - i - j - 1} \, dp$$

$$= \frac{s!}{j!(s-j)!} \frac{\Gamma(\beta_1 + \beta_2 + s)}{\Gamma(\beta_1 + i) \Gamma(\beta_2 + s - i)} \frac{\Gamma(\beta_1 + i + j) \Gamma(\beta_2 + 2s - i - j)}{\Gamma(\beta_1 + \beta_2 + 2s)}$$

$$\times \int_0^1 \frac{\Gamma(\beta_1 + \beta_2 + 2s)}{\Gamma(\beta_1 + i + j) \Gamma(\beta_2 + 2s - i - j)} p^{\beta_1 + i + j - 1} (1-p)^{\beta_2 + 2s - i - j - 1} \, dp$$

The second term in the last line is a Beta distribution with parameters $\beta_1 + i + j$ and $\beta_2 + 2s - i - j$ which integrates to one. Hence, as required:

$$q_{ij} = \frac{\Gamma(s+1)}{\Gamma(j+1)\Gamma(s-j+1)} \frac{\Gamma(\beta_1 + \beta_2 + s)}{\Gamma(\beta_1 + i)\Gamma(\beta_2 + s - i)} \frac{\Gamma(\beta_1 + i + j)\Gamma(\beta_2 + 2s - i - j)}{\Gamma(\beta_1 + \beta_2 + 2s)}$$

A.2. **Numerical Calculations.** Setting $\beta_1 = \beta_2 = 1$ obtains the uniform distribution for the prior. Then q_{ij} simplifies to the expression in Equation 9. Condition 1 can provide a lower bound for maximum values of x^* for which it is satisfied. Since $x^* < 1/2$:

$$\min_{i < \lceil sx^* \rceil} \left\{ \sum_{j=i+1}^{s} q_{ij} \right\} \geq \min_{i < \lceil s/2 \rceil} \left\{ \sum_{j=i+1}^{s} q_{ij} \right\}$$

The second expression can be numerically calculated for a range of sample sizes s, using Equation 9. If x^* is less than this value, then Condition 1 is certainly satisfied. Such lower bounds are illustrated in Figure 6.
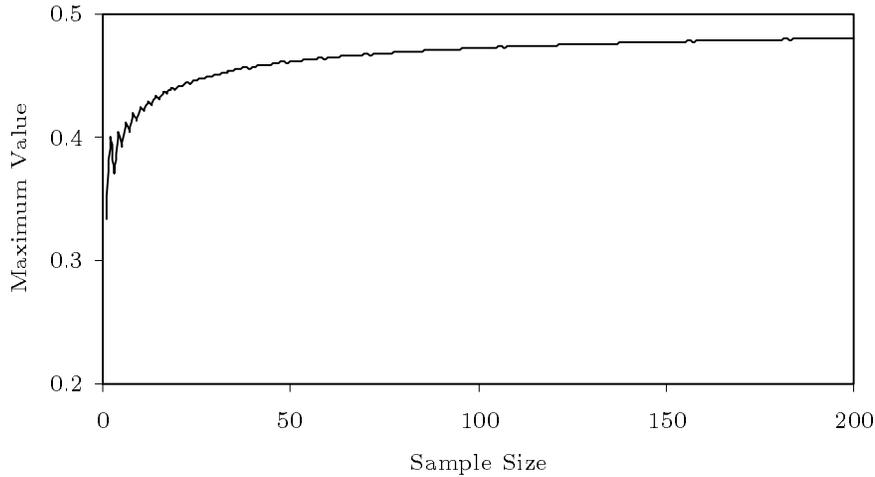

FIGURE 6. Lower Bounds on the maximum value of x^* for Condition 1 to hold.

Finally, Condition 1 implies Condition 2 for such q_{ij} as shown in the following lemmas:

Lemma 9. *If $j > i$ then $q_{ij} < q_{i+1j}$ and if $j \leq i$ then $q_{ij} \geq q_{i+1j}$, where:*

$$q_{ij} = \frac{s+1}{2s+1} \frac{\binom{s}{i}\binom{s}{j}}{\binom{2s}{i+j}}$$

Proof. Consider the ratio q_{ij}/q_{i+1j}:

$$\frac{q_{ij}}{q_{i+1j}} = \frac{\frac{s+1}{2s+1}\binom{s}{i}\binom{s}{j}\binom{2s}{i+1+j}}{\frac{s+1}{2s+1}\binom{2s}{i+j}\binom{s}{i+1}\binom{s}{j}} = \frac{\binom{s}{i}\binom{2s}{i+1+j}}{\binom{2s}{i+j}\binom{s}{i+1}}$$

Expanding the terms of this fraction leaves:

$$\frac{q_{ij}}{q_{i+1j}} = \frac{(2s-i-j)!\,(i+j)!}{(s-i)!i!}\frac{(s-i-1)!\,(i+1)!}{(2s-i-1-j)!\,(i+1+j)!}$$

$$= \frac{(i+1)\,(2s-i-j)}{(s-i)\,(i+1+j)} = \frac{(i+1)\,(s-i)+(i+1)\,(s-j)}{(i+1)\,(s-i)+j\,(s-i)}$$

If $j > i$ then (since both i and j are integers and lie between 0 and s) $s\,(j-i) > s-j$ and hence $j\,(s-i) > (i+1)\,(s-j)$, so $q_{ij}/q_{i+1j} < 1$ and the result holds. Likewise for $j \leq i$. □

Lemma 10. *Condition 1 implies Condition 2 for such q_{ij}.*

Proof. Consider the bound in Condition 2:

$$\min_{i \geq \lceil sx^* \rceil}\left\{ \sum_{j=\lceil sx^* \rceil}^{s} q_{ij} \right\} = \min_{i \geq \lceil sx^* \rceil}\left\{ 1 - \sum_{j=0}^{\lceil sx^* \rceil-1} q_{ij} \right\} = 1 - \max_{i \geq \lceil sx^* \rceil}\left\{ \sum_{j=0}^{\lceil sx^* \rceil-1} q_{ij} \right\}$$

Notice that in the last term $j < i$ throughout the sum. Bringing the results of Lemma 9 to bear, $q_{ij} \geq q_{i+1j}$ for all such $j < i$. Hence, the maximum takes place at the minimum i available, *i.e.* $i = \lceil sx^* \rceil$. So, the equality becomes:

$$\min_{i \geq \lceil sx^* \rceil}\left\{ \sum_{j=\lceil sx^* \rceil}^{s} q_{ij} \right\} = 1 - \left\{ \sum_{j=0}^{\lceil sx^* \rceil-1} q_{\lceil sx^* \rceil j} \right\} = \sum_{j=\lceil sx^* \rceil}^{s} q_{\lceil sx^* \rceil j}$$

But if Condition 1 holds then:

$$x^* < \min_{i < \lceil sx^* \rceil}\left\{ \sum_{j=i+1}^{s} q_{ij} \right\} \leq \sum_{j=\lceil sx^* \rceil}^{s} q_{\lceil sx^* \rceil-1j} < \sum_{j=\lceil sx^* \rceil}^{s} q_{\lceil sx^* \rceil j}$$

Where the first inequality follows simply because the minimum is being taken over all values of $i < \lceil sx^* \rceil$, which includes $i = \lceil sx^* \rceil - 1$. The second inequality again follows

from Lemma 9. $i = \lceil sx^* \rceil - 1 < j$ and hence $q_{ij} < q_{i+1j}$. That is, $q_{\lceil sx^* \rceil - 1j} < q_{\lceil sx^* \rceil j}$ for all such j. The bound in Condition 2 is met once Condition 1 is satisfied. $\square$

Figure 6 therefore illustrates lower bounds on the maximum value of x^* for *both* Conditions to hold. It seems, from the numerical calculations, that any value of $x^* < 1/2$ can satisfy the Conditions given a large enough s. Even for relatively small sample sizes, x^* can be very close to $1/2$.

NUFFIELD COLLEGE, OXFORD OX1 1NF, UNITED KINGDOM

E-mail address: david.myatt@nuffield.oxford.ac.uk

URL: http://www.nuff.ox.ac.uk

NUFFIELD COLLEGE, OXFORD OX1 1NF, UNITED KINGDOM

E-mail address: christopher.wallace@nuffield.oxford.ac.uk

URL: http://www.nuff.ox.ac.uk