SEMINARIO DE ECONOMÍA

Xoves, 20 de Outubro

Título:

"Learning Dynamics Based on Social Comparisons"

Ponente:

Juan I. Block
(*University of Cambridge*)

Lugar: Aula-Seminario 6

Hora: 13:00 h

Organiza:

ECOBAS
Economics and Business Administration for Society
www.ecobas.es

Learning Dynamics Based on Social Comparisons*

Juan I. Block† Drew Fudenberg‡ David K. Levine§

June 30, 2016

Abstract

We study models of learning in games where agents with limited memory use social information to decide when and how to change their play. We demonstrate that agents come close to Nash equilibrium behavior for (generic) games. Transitions between equilibria are governed by trembles and we characterize the stochastically stable equilibria as these perturbations vanish. When agents only observe the aggregate distribution of payoffs and only recall information from the last period we show that pure equilibria are generally more stable than mixed equilibria. When agents observe which actions have given high payoffs and remember those actions for a number of periods, mixed equilibria can be favored over pure ones provided that the agent's memory is sufficiently long.

KEYWORDS: Evolution, social learning, strict equilibria, best response dynamics, equilibrium selection.

*We thank Glenn Ellison, Salvatore Modica, and Larry Samuelson for helpful comments.

†Faculty of Economics, University of Cambridge. Email: jb2002@cam.ac.uk

‡Department of Economics, MIT. Email: drew.fudenberg@gmail.com

§Departments of Economics, EUI and WUSTL. Email: david@dklevine.com

1 Introduction

This paper develops and analyzes two models of learning in games based on social comparisons, where agents have limited information and memory. In our *low-information* model, agents observe the highest utility realized in their own population without observing the corresponding actions. If they are getting close to the highest payoff in their population, then they are content, and continue to play the same action. Otherwise, they become discontent and experiment at random with different actions in hopes of doing better. Memory is limited in sense that agents do not remember all the things that happened in the past, just whether they are content, and if so, what they did last period.[1] In addition, the behavior we specify implicitly supposes that agents do not try influence the future play of others; this "strategic myopia" makes the most sense when the population is relatively large.

When people can observe strategies that worked well for other people the most natural thing to do is mimic those strategies. When they do not observe each others' behavior but can observe their payoffs, people may experiment if they find they are doing less well than others. We are motivated by the fact that, for example, individuals may learn from reading newspaper or watching television which report aggregate data on the economy payoffs (stock index, average wages per industry, and income distribution). This restricted form of information seems plausible in many real world social interactions in which it is difficult for people to obtain detailed information about other people's behavior. It is also of relevance to laboratory experiments on games with large extensive forms, such as indefinitely repeated games: Here it is feasible to tell participants the payoffs that other participants received in past plays of the repeated game, but not to tell them the exact strategies used by the participants who obtained high payoffs.[2]

In addition to the random play of discontent agents, our model has three other stochastic components. First, with a probability that we send to 0, agents tremble and become discontent, which triggers a wide search on the strategy space. Second, agents only reassess their play with a probability that is bounded away from 1. Finally, we assume there is a small number of *committed* agents who are always content and so play specific actions regardless of what they observe. We are interested in social norms that emerge and last in the long run so they must satisfy a notion of robustness, in particular, they must not be sensible to a small number of agents that make choices outside the social convention. These assumptions make the resulting system ergodic,[3] and

[1] In a recent paper, Fudenberg and Peysakhovich (2014) find experimentally that last period experiences have a larger impact on individuals' behavior than do earlier observations, and that individuals approach optimal strategies when provided with summary statistics. For a discussion of recency effects in decision making experiments, see Erev and Haruvy (2013). Recency effects have also been found in the field, for example, in the credit card market as in Agarwal et al. (2008), in the stock market as in Malmendier and Nagel (2011), or in consumers' choices made from a list as in Feenberg et al. (2015).

[2] See for example the repeated prisoner's dilemma experiments surveyed in Dal Bó and Fréchette (2016). In many of these, a substantial minority of participants defects most or all of the time, and receive a much lower overall payoff than subjects who appear to be "conditionally cooperative," which raises the question of what would happened if participants were told something about the payoffs that others have received in previous plays of the repeated game.

[3] This rules out the effect of history or initial conditions epitomized in Schelling's (1960) focal points, which allows us to make predictions based solely on the payoff matrix of the game; we view this as an approximation of social

in the limit as the probability of trembling goes to zero, the system spends almost all of its time at states where all but the committed agents are getting about the same payoff. Moreover, the presence of the committed agents means that every possible action has positive probability, so in generic games these limit states must be approximate Nash equilibria. The stochastically stable states, those that have non-vanishing frequency in the limit as the probability of a "tremble shock" goes to 0, are those where the largest number of shocks is required to lead the system to another equilibrium state; these numbers are called the "radii" of the equilibria (Ellison (2000)). We find that while the radius of pure equilibria is generally large (in particular growing linearly with the size of the population) the radius of all mixed equilibria is fixed at 1. We use this to show that in large populations mixed equilibria are significantly less stable than any of the pure equilibria, even those that are not stochastically stable, and even when the mixed equilibrium gives the players a higher payoff in line with experimental evidence (see, for example, Van Huyck et al. (1990)).[4]

There are environments in which people can access aggregate information about behavior, for example, petitions, public protests and other kinds of activism. We are interested in limited memory of bounded length because it is common practice in some institutions to delete all records after a fixed period of time (due to storage costs or law), because recency effects can exclude more distant observations, and record-keeping devices depreciate. We then explore the effect of allowing the agents to use more information and memory on long-run dynamics while still relying basing their decisions mainly on social information. Our *high-information* model supposes that agents observe the highest payoff realized in their own population together with the corresponding action, and moreover that they recall the actions that were best responses in the last finite T periods.[5] Since agents generally experiment less when they are more experienced we assume that discontent agents randomize over the set of remembered best responses and last period action instead of over all actions. If agents only recall best responses in the last period, we show that our social learning process and the standard best response with inertia dynamic (Samuelson (1994)) predict the same stochastically stable set. In particular both models can have stochastically stable cycles, and we believe that it is more likely that that the system would be bogged down in a best response cycle rather than moving to a mixed equilibrium in generic games.[6] However, having sufficiently long memory in our learning process leads to global convergence to approximate Nash equilibria in generic games, even games that have only mixed Nash equilibria; unlike the best response with inertia dynamic for which convergence is obtained by assuming acyclic games (Young (1993)). We highlight the role of memory by showing play always converges to a Nash equilibrium if agents have

norms or conventions where payoff considerations are the most important forces.

[4]This is the first such result we know of for this sort of process. Fudenberg and Imhof (2006) characterize the relative frequencies of various homogeneous steady states in a family of imitation processes, but the processes they study can in some games spend most of their time near non-Nash states. Levine and Modica (2013) like us examine the relative amount of time spent at different steady states corresponding to Nash equilibria but examine a dynamic based on group conflict rather than driven by learning errors.

[5]In contrast, Young's (1993) adaptive learning rule has one agent revising at a time that observes a sample of size K from the last T periods and chooses among those actions that are best response to the empirical distribution of actions in the sample.

[6]We are not aware of a general characterization for best response with inertia dynamics.

a memory T that is at least $k \times l$, when the game is $k \times l$ *acyclic*, meaning that from any strategy profile there is a best response path to a $k \times l$ curb block (Basu and Weibull (1991)). Because every game is acyclic for k and l at least as large as the action spaces, this means that global convergence is guaranteed in any game when memory is sufficiently long.

The main methodological contribution of the paper is to characterize the learning dynamics combining the standard theory of perturbed Markov chains and the method of circuits (see Levine and Modica (2016)), adapting their Theorem 9 to the case in which there is a single circuit. To illustrate the complementarity between this approach and past work, we show how to find the stochastically stable set constructing circuits of circuits and alternatively by using Ellison's (2000) radius-coradius theorem. Our results also contribute to the long-standing debate about pure versus mixed equilibria, providing a clear connection between what players observe and equilibrium selection. We show that, in large populations, pure equilibria are more stable in environments where agents only know that there is a better response whereas mixed equilibria are sometimes more stable in environments where agents have enough information that they know the best response.

In addition to its focus on learning from summary statistics based on social information, this paper contributes to the larger literature that uses non-equilibrium adaptive processes to understand and predict which Nash equilibria are most likely to be observed. The literature on belief-based learning models such as stochastic fictitious play (Fudenberg and Kreps (1993), Fudenberg and Levine (1998), Benaïm and Hirsch (1999), Hofbauer and Sandholm (2002)) concludes that stable equilibria can be observed while unstable equilibria cannot be, but also concludes there can be stable cycles. The same conclusion applies in the literature that studies deterministic best-response-like procedures perturbed with small random shocks (Kandori et al. (1993), Young (1993), others), although that literature, unlike the one on stochastic fictitious play, does sometimes provide a way of selecting between strict equilibria; for example it selects the risk-dominant equilibrium in 2×2 coordination games, as does our social comparison dynamic. In larger coordination games, the two dynamics can make different selections; we discuss this further in Section 5. The idea that players observe outcomes and update play with probability less than 1 appears in the Nöldeke and Samuelson (1993) analysis of evolution in games of perfect information, unlike their learning procedure we assume that agents are able to observe the average payoff and/or action distribution not the outcomes of all matches for the current round of play.[7] The ideas that agents only change their actions if they are "dissatisfied" and/or that they have information about the distribution of payoffs have also been explored in the literature, but these papers (for example Björnerstedt and Weibull (1996), Binmore and Samuelson (1997)) have assumed that agents receive information about the actions or strategies used by agents they have not themselves played as we do in our high information model. Our committed agents resemble "non-conventional" agents proposed by Myerson and Weibull (2015) in that committed agents consider a (strict) subset of actions, however, we focus on committed agents with singleton action sets. We restrict attention to populations in

[7] As in our model, this stochastic observation technology means that every sequence of one-move-at-a-time intentional adjustments has positive probability; they use this to show that if a single state is selected as noise goes to 0, it must be a self-confirming equilibrium (Fudenberg and Levine (1993)).

which the prevalent agents are learners (i.e. non-committed agents) similar to the player type distribution they assume.

A more recent literature has considered learning procedures that involve a substantial amount of randomization when players are "dissatisfied." These papers are oriented at showing the possibility of global convergence in a setting with long memory. By contrast we are focused on long-run comparative statics: we compare a range of different learning procedures to characterize which ones have global convergence in which types of games and on the relative time spent at different steady states, for example, mixed versus pure. Some recent papers, such as Hart and Mas-Colell (2006), Fudenberg and Levine (2014) show that there are classes of such procedures that converge with probability one to Nash equilibrium, while others such as Foster and Young (2003; 2006), Young (2009), Pradelski and Young (2012), Foster and Hart (2015) show that there are other classes of procedures that remain at Nash equilibrium a very large fraction of the time. Building on Young (2009), Pradelski and Young (2012) show that an efficient equilibrium is selected in games with generic payoffs for which a pure Nash equilibrium exists. In contrast, we also consider generic games with only mixed equilibria, and our procedure selects the risk dominant equilibrium in 2×2 coordination games as is standard in the literature whereas their procedure does not. However for these stochastic procedures with good global convergence properties, little is known about which Nash equilibria are likely to be observed in generic games.

2 The Model

2.1 The Stage Game

We consider play between two populations. There are N identical agents in each population, indexed by i. In the stage game, agent i of each population j chooses an action a^j from the finite set A^j. Agents are matched round robin consecutively against agents of the opposing population. We refer to *uniform play* by agent i of player j to indicate that the agent chooses randomly and uniformly a^i over all actions in A^j, where the choice is held fixed throughout the round robin. Agent i in population j who plays a^i against an opponent playing a^{-j} receives utility $u^j(a^i, a^{-j})$. For any finite set X, we let $\Delta(X)$ denote the space of probability distributions over X. Play in population j can be represented by the mixed profile $\alpha^j \in \Delta(A^j)$, and $\alpha^j(a^j)$ can be interpreted as the fraction of agents i playing $a^i = a^j$. The overall utility of agent i is then $u^j(a^i, \alpha^{-j})$ since he plays each opponent in the opposing population in turn.[8]

We define the payoff frequencies based on the population play during the round robin. Let $U^j(\alpha^{-j})$ denote the finite vector of utilities corresponding to $u^j(a^i, \alpha^{-j})$ for $a^i \in A^j$, and let $\phi^j(\alpha) \in \Delta(U^j(\alpha^{-j}))$ be the frequency distribution of utilities of population j corresponding to the mixed strategy profile α. Let $\mathbf{A}^j(u^j, \alpha^{-j}) \subseteq A^j$ be the possibly empty subset of actions a^i for which

[8]This can be thought of as an approximation to a situation where each agent is randomly matched against the opposing population a substantial number of times. See Ellison et al. (2009) for conditions under which this approximation is valid.

$u^j(a^i, \alpha^{-j}) = u^j$. Then the frequency of utility u^j is $\phi^j(\alpha)[u^j] = \sum_{a^i \in \mathbf{A}^j(u^j, \alpha^{-j})} \alpha^j(a^i)$.

Finally, we introduce a notion of (an approximate) best response that captures the fact that agents only choose pure actions: for $\nu \geq 0$ we say that $\hat{a}^i$ is a *ν-best response to* $\alpha^{-j} \in \Delta(A^{-j})$ if $u^j(\hat{a}^i, \alpha^{-j}) + \nu \geq u^j(a^i, \alpha^{-j})$ for all $a^i \in A^j$.

2.2 Low Information Social Learning

We propose a learning procedure in which agents have no direct information about the behavior of their own and or the opposing population. More precisely, we assume that all agents in population j observe only the frequency of utilities ϕ^j in their own population, and that they have only partial ability to keep track of that information over time due to limited memory.

The environment in which learning takes place is the stage game played in every period $t = 0, 1, 2, \ldots$. A fixed number are *committed* agents. Committed agents always play the action they are committed to; we assume there is at least one committed agent of type ξ^j for each action $a^j \in A^j$, and denote the set of committed agents by Ξ^j. We refer to the other $N - \#\Xi^j$ agents in each population as *learners.* The state of an agent's limited memory at the start of period t is $\theta_t^j \in \Theta^j \equiv A^j \cup \{0\} \cup \Xi^j$, we call this the agent's *type.* For learners, if the type $\theta_t^i \in A^j$ the agent is *content* with the action θ_t^i, and if the type $\theta_t^i = 0$ the agent is *discontent.* The process begins with an exogenous initial distribution of these types.

The play of agents is determined by their type. We assume that learners may tremble when choosing their action, independently across agents and time.[9] Each learner i of player j *trembles* with independent probability ϵ, and engages in uniform play, meaning that the agent chooses a_t^i with a uniform distribution over A^j. If the agent does not tremble, his behavior depends on his type θ_t^i. If agent i is content in period t or is committed he plays $a_t^i = \theta_t^i$. A discontent agent i engages in uniform play.[10]

In period $t+1$, the type θ_{t+1}^i of a learner in population j is determined by the action a_t^i they played in period t, their previous type θ_t^i, the *social comparison* parameter $\nu \geq 0$, and the aggregate statistic $\phi^j(\alpha_t)$ as follows. A learner who trembled is discontent, thus $\theta_{t+1}^i = 0$. Otherwise each agent i has an independent probability $1 > p > 0$ of being *active* and complementary probability $1 - p$ of being *inactive.* Inactive agents remain content or discontent as they were in the previous period so that $\theta_{t+1}^i = \theta_t^i$. Active agents may change contentment depending upon $\phi^j(\alpha_t)$ and ν; we now explain how this updating process takes place. Let $\overline{u}^j(\phi^j(\alpha_t))$ denote the highest time-t utility received in population j.[11] If $u^j(a_t^i, \alpha_t^{-j}) > \overline{u}^j(\phi^j(\alpha_t)) - \nu$ and agent i is active he becomes or remains content, so $\theta_{t+1}^i = a_t^i$. Otherwise he becomes or remains discontent, so $\theta_{t+1}^i = 0$. Note that this social comparison will indicate whether the agent is playing an approximate best response,

[9] Notice that we allow discontents to tremble although since they play the same way if they tremble as if they do not it does not matter.

[10] In place of uniform play we can allow state dependent probability distributions that may have a bias towards certain actions. As long as these probabilities are bounded away from zero independent of ϵ our results are robust. However, it is not clear what the probability of the last action played should be.

[11] Agents observe the average payoff distribution of actions played, not the payoff distribution across matches.

since there is always a committed agent playing a ν-best response.

In summary, the play of the learners is governed by three parameters: the probability ϵ of trembling, the probability p of being active and ν, the tolerance for getting less than the current highest possible payoff.[12]

2.3 Aggregate Dynamics

The behavior of individual agents gives rise to an aggregate dynamic. This can be described either statistically in terms of states describing the population shares of the different types or in terms of agent-states describing the specific states of individual agents. It is convenient for descriptive purposes to take the state z to be described by population shares. However, to derive the transition probabilities $P_\epsilon(z_{t+1}|z_t)$ it is useful to use the greater detail of agent-states.

For any integer K and any set X let $\Delta^K(X)$ be the subset of $\Delta(X)$ where each coordinate is an integer multiple of $1/K$. Let $\Phi_t^j \in \Delta^N(\Theta^j)$ be a vector of population shares of the player j types in period t. Define the (finite) state space $Z = \Delta^N(\Theta^1) \times \Delta^N(\Theta^2)$ to be the set of vectors $z = (\Phi^1, \Phi^2)$. We will also want to deal with the population fractions playing different actions. We call $\Delta^N(A^j)$ the *grid for population* j; the *grid* is the product space $\Delta^N(A) = \Delta^N(A^1) \times \Delta^N(A^2)$. We can approximate any $\alpha \in \Delta(A)$ by means of a population play in $\Delta^N(A)$ when N is large. We will also make use of the grids for subsets of the population.

2.3.1 Aggregate Transition Probabilities

In this subsection we derive the transition probabilities of aggregate statistical states $P_\epsilon(z_{t+1}|z_t)$. We first introduce the notion of an *agent state* $x = (x^1, x^2)$ as an assignment of types to agents $x^j \in N^{\Theta^j}$. An agent state x induces population shares of player types (Φ^1, Φ^2); it is *consistent* with a state z if the shares match those in z, in which case we write $x \in X(z)$. To determine the aggregate transition probability $P_\epsilon(z_{t+1}|z_t)$ from z_t to z_{t+1} start by choosing an agent state $x_t \in X(z_t)$, that is, consistent with z_t. For any $x_{t+1} \in X(z_{t+1})$ we will define the *agent-state transition probability* $P_\epsilon(x_{t+1}|x_t)$ and we then compute $P_\epsilon(z_{t+1}|z_t) \equiv \sum_{x_{t+1} \in X(z_{t+1})} P_\epsilon(x_{t+1}|x_t)$. This is well defined since while $P_\epsilon(x_{t+1}|x_t)$ depends on which $x_t \in X(z_t)$ is chosen the sum does not.[13] Let $\mathcal{T}^j$ denote the learners of player j that tremble and let $\mathcal{N}^j$ be the non-trembling learners. Define any assignment of actions to all agents for each j by $\sigma^j \in N^{A^j}$. Let each subset $\mathcal{R}^j \subseteq \mathcal{N}^j$ be the agents who are active conditional on $\mathcal{T}^j$. Define $D^j(x_t)$ to be the number of discontent types in x_t. Let $\mathcal{T}_C^j(x_t) \subseteq \mathcal{T}^j$ and $\mathcal{N}_C^j(x_t) \subseteq \mathcal{N}^j$ be the subsets corresponding to content agents in x_t.

Lemma 1. *The aggregate transition probabilities are given by*

[12] We assume that the learning model parameters are common to all players, and that the actions of the discontent players are drawn from a uniform distribution. As long as all errors and actions have positive probability and the order of magnitude of the error rates is common to all players these assumptions do not change our conclusions.

[13] If we permute the names in x_t and the names in x_{t+1} the same way then the agent-state transition probability is unchanged.

$$P_\epsilon(z_{t+1}|z_t) = \sum_{x_{t+1}\in X(z_{t+1})} \sum_{\mathcal{T},\sigma,\mathcal{R}} \prod_{j=1,2} \epsilon^{\#\mathcal{T}^j}(1-\epsilon)^{\#\mathcal{N}^j} \left(\frac{1}{\#A^j}\right)^{D^j(x_t)+\#\mathcal{T}^j_C(x_t)} p^{\#\mathcal{R}^j}(1-p)^{\#\mathcal{N}^j-\#\mathcal{R}^j}$$
$$= \sum_{x_{t+1}\in X(z_{t+1})} \sum_{\mathcal{T},\sigma,\mathcal{R}} P(\mathcal{T},\sigma,\mathcal{R},x_{t+1}|x_t),$$

if σ^j is feasible given $\mathcal{T}^j$ and x_t, or if $x_{t+1} \in X(z_{t+1})$, otherwise $P_\epsilon(z_{t+1}|z_t) = 0$.

Next we note that an agent who is doing well will never get a signal that suggests he is doing poorly, so these agents only become discontent when they tremble.

Lemma 2 (ν**-Best Responses Stick**). *If σ^j is feasible with respect to $\mathcal{T}^j$ and some content agent $i \in \mathcal{R}^j$ is playing an a^i_t which is a ν-best response to α_t^{-j}, and $\theta^i_{t+1} \neq a^i_t$ in x_{t+1}, then $P(\mathcal{T},\sigma,\mathcal{R},x_{t+1}|x_t) \leq \epsilon$.*

Proof. Since $i \in \mathcal{R}^j$ is content and playing a ν-best response to α_t^{-j} it cannot be that $u^j(a^i_t,\alpha_t^{-j}) \leq \overline{u}^j(\phi^j(\alpha_t)) - \nu$. Hence agent i must either remain content with a^i_t or must have trembled: in the latter case the whole transition has probability at most ϵ. □

2.3.2 ν-Robust States

We have defined an *aggregate transition probability* $P_\epsilon(z_{t+1}|z_t)$. This then gives rise to a (time homogeneous) Markov process on the state space Z that captures the dynamics of learning. The resulting stochastic process is governed by a Markov transition kernel $P_\epsilon(\cdot|z) \in \Delta(Z)$ which takes Z into a probability distribution on Z. Our interest is in studying this Markov process and how it depends upon ϵ, the tremble probability of each learner. Let $D^j(z)$ be the number of discontent agents of player j in state z. For example, when all learners from population j have $\theta^i \in A^j$, $D^j(z) = 0$. We continue to let $\overline{\alpha}^j(z) \in \Delta^{N-D^j(z)}(A^j)$ be the action profile corresponding to the content and committed types in z and to denote by $\mathcal{A}^j(z)$ to be the set of feasible $\alpha^j \in \Delta^N(A^j)$ such that $N\alpha^j = (N - D^j(z))\overline{\alpha}^j(z) + D^j(z)\tilde{\alpha}^j$ for some action profile $\tilde{\alpha}^j \in \Delta^{D^j(z)}(A^j)$.

Definition 1. A state z is *ν-robust* if $D^j(z) = 0$ for all j, and all the learners i from either population j are playing a ν-best response to the unique $\alpha^{-j}(z) \in \mathcal{A}^{-j}(z)$.

Note that a *ν-robust* state is automatically ν'-robust for any $\nu' > \nu$. We say that a state z is *pure for population j* if all learners in population j are in the same state, and that the state is *pure* if it is pure for both populations. Otherwise, we refer to as a *mixed* state.

2.4 Assumptions

We make the following generic assumption.

Assumption 1. *Every pure action in the stage game has a unique best response.*

Since a unique best response must be strict, we may define $g > 0$ as the smallest difference between the utility of a best response and a second best response.

Assumption 2. $\nu < g$.

We previously noted that the assumption of myopic play makes the most sense when N is relatively large. For ν-robust states to be interesting we also need ν sufficiently small so that not all possible actions are consistent with approximate best response, and $M \equiv \max\{\#\Xi^1, \#\Xi^2\}$ is not too large relatively to N, otherwise, given the fixed play of the M committed agents, there might be no ν-robust state. Specifically we are interested in the following result.

Lemma 3. *Under Assumption 2, there is an η such that if $N/M > \eta$ then*

(1) if a^j is a ν-best response to $a^{-j} \in A^{-j}$ then a^j is a strict 0-best response to all $\alpha^{-j} \in \Delta^N(A^{-j})$ such that $\alpha^{-j}(a^{-j}) > 1 - M/N$;

(2) for $\nu > 0$ a ν-robust state exists; if the game has a pure strategy Nash equilibrium a 0-robust state exists.

Assumption 3. *$N/M \geq \eta$ where η is large enough that Lemma 3 holds.*

Note that the η required by Assumption 3 depends on the value of ν, and that when N/M is sufficiently large, strict best responses to pure strategies are also strict best responses regardless of the play of committed agents.

This gives the following useful result.

Lemma 4. *Under Assumption 3 if there is j and an $\hat{a}^j \in A^j$ such that for all $\alpha^{-j} \in \Delta^N(A^{-j})$ we have $u^j(\hat{a}^j, \alpha^{-j}) + \nu \geq u^j(a^j, \alpha^{-j})$ for all $a^j \in A^j$, then for $N/M \geq \eta$ there is a unique ν-robust state.*

For the rest of the paper we will maintain Assumptions 1-3.

3 Preliminary Results

Our analysis is divided into two cases: exact best response behavior and pure strategy equilibria, and approximate equilibria. In this section, we develop a battery of preliminary results that can be applied to both cases.

3.1 Absorbing States

We start by analyzing some of the dynamics of the learning procedure in the absence of trembles. We will study these dynamics by characterizing the induced Markov process with transition probabilities that obey P_0. A state $z \in Z$ is *absorbing* if the process stays in the state z forever once it has visited that state, and a collection of states is a *recurrent class* or *limit set* if it has positive probability of being reached, and once it is reached every state in the class recurs infinitely often. Our first result establishes that ν-robust states are absorbing states of the process.

Theorem 1. *If $\epsilon = 0$ a ν-robust state z is absorbing.*

Proof. Starting in a ν-robust z_t, take $x_t \in X(z_t)$. Since $\epsilon = 0$ nobody trembles so $\mathcal{T}^j = \emptyset$. Consider any feasible σ^j and any specification of which learners are active. By assumption all learners are content so by Lemma 2 they all remain at $\theta^i_{t+1} = a^i_t$. The committed agents never change state by assumption, so $x_{t+1} = x_t$ with probability 1. This implies that $z_{t+1} = z_t$ with probability 1. □

We will subsequently show that all other states are transient when $\epsilon = 0$ and use this to characterize the ergodic distribution of states when ϵ is small but not zero.

3.2 Learning Dynamics When $\epsilon > 0$

We now examine the aggregate dynamics when ϵ is small but positive. Our goal is to characterize how much time the system spends in such states as the tremble probability becomes small (that is, as $\epsilon \to 0$) in the long run. We note that the Markov process P_ϵ is irreducible and aperiodic (see Lemma 15 in the Appendix).

From Young (1993) the fact that P_ϵ is irreducible and aperiodic implies that there is a unique invariant distribution $\mu^\epsilon \in \Delta(Z)$ for every small $\epsilon > 0$ satisfying $\mu^\epsilon P_\epsilon = \mu^\epsilon$ and we denote by μ^ϵ_z for each $z \in Z$ the (ergodic) probability assigned to state z. To characterize the support of the ergodic distribution on states as $\epsilon \to 0$ we will make use of the concept of the *resistance* of the various state transitions. Because $P_\epsilon(z'|z)$ is a finite polynomial in ϵ for any $z, z' \in Z$, it is *regular,* meaning that $\lim_{\epsilon\to 0} P_\epsilon = P_0$ exists, and if $P_\epsilon(z'|z) > 0$ for $\epsilon > 0$ then for some non-negative number $r(z, z')$ we have $\lim_{\epsilon\to 0} P_\epsilon(z'|z)\epsilon^{-r(z,z')}$ exists and is strictly positive. We then write $P_\epsilon(z'|z) \sim \epsilon^{r(z,z')}$; let $r(z, z') \in [0, \infty]$ denote the resistance of the transition from z to z'. Moreover if $P_\epsilon(z'|z) = 0$ then this transition is not possible and we set $r(z, z') = \infty$, while if $P_0(z'|z) > 0$ we have $r(z, z') = 0$.

Note that some agent state transition probabilities are *appreciable* in the sense that they are bounded away from zero independent of ϵ, and the same is true of some aggregate transition probabilities. In analyzing resistance $r(z, z')$ it is useful to observe that $P_\epsilon(z'|z)$ is defined as a sum. The resistance of a sum is given by the least resistance of any term in the sum.

Since the terms in the sum are $P(\mathcal{T}, \sigma, \mathcal{R}, x_{t+1}|x_t)$ it is sufficient when analyzing resistance to look for the target x_{t+1} and realizations $\mathcal{T}, \sigma, \mathcal{R}$ for which the probability has the least resistance. Denote this resistance as $r(x_t, x_{t+1})$. For this to have finite resistance it must be that σ^j is feasible given $\mathcal{T}^j$ for $j = 1, 2$ and that $x_{t+1} \in X(z_{t+1})$. In that case the resistance is equal to number of trembles, $r(x_t, x_{t+1}) = \#\mathcal{T}^1 + \#\mathcal{T}^2$. In particular to show that the aggregate resistance is zero it is sufficient to find an agent state resistance for the transition that has resistance zero. We also write $z \to z'$ for the transition (z, z').

It is convenient to define transitions between more than two states since the Markov process may pass through various intermediate states when going from one state to a target state. We say a path $\mathbf{z}$ is a finite sequence of at least two not necessarily distinct states $(z_0, z_1, \ldots, z_t)$ and its resistance is defined as $r(\mathbf{z}) = \sum_{k=0}^{t-1} r(z_k, z_{k+1})$. Notice that we allow for *loops* where some states are revisited along the path.

Lemma 5. *If $\mathbf{z} = (z_0, z_1, \ldots, z_t)$ is a path then there exists a path $\tilde{\mathbf{z}} = (\tilde{z}_0, \tilde{z}_1, \ldots, \tilde{z}_{\tilde{t}})$ with $\tilde{z}_0 = z_0$ and $\tilde{z}_{\tilde{t}} = z_t$ with $r(\tilde{\mathbf{z}}) \leq r(\mathbf{z})$, and an agent state $\tilde{x}_\tau \in X(\tilde{z}_\tau)$ for $\tau = 0, 1, \ldots, \tilde{t}$ in which no discontent agent trembles and every content agent, including those who tremble, plays the action with which they are content.*

Proof. First observe that we can replace the discontent agents who tremble with discontent agents who play the same way and who are inactive and strictly lower the resistance, so there is a path to the target with no greater resistance if no discontent agent ever trembles. To show that we can have every content agent playing the same action, we replace each transition $z_\tau, z_{\tau+1}$ with two transitions $z_\tau, \tilde{z}_{2\tau+1}, z_{\tau+1}$. Let $x_\tau \in X(z_\tau)$ together with $\mathcal{T}_\tau, \sigma_\tau, \mathcal{R}_\tau, x_{\tau+1} \in X(z_{\tau+1})$ have resistance $r(z_\tau, z_{\tau+1})$. For the transition $z_\tau, \tilde{z}_{2\tau+1}$ choose the same x_τ, set $\tilde{\mathcal{T}}_\tau = \mathcal{T}_\tau$, and $\tilde{\sigma}_\tau$ such that all content agents play the action with which they are content, $\tilde{\sigma}_t^j$ is consistent with $\overline{\alpha}^j(x_\tau, \emptyset)$, and all agents are inactive. Then $r(x_\tau, \tilde{x}_{2\tau+1}) = r(x_\tau, x_{\tau+1})$ so that $r(z_\tau, \tilde{z}_{2\tau+1}) \leq r(x_\tau, x_{\tau+1}) = r(z_\tau, z_{\tau+1})$. For the transition $\tilde{z}_{2\tau+1}, z_{\tau+1}$ take $\tilde{\mathcal{T}}_{2\tau+1}^j = \emptyset$, $\tilde{\sigma}_{2\tau+1} = \sigma_\tau$ and $\tilde{\mathcal{R}}_{2\tau+1} = \mathcal{R}_\tau$ so that the terminal state is $x_{\tau+1} \in X(z_{\tau+1})$ and $r(\tilde{x}_{2\tau+1}, x_{\tau+1}) = 0$ implying $r(\tilde{z}_{2\tau+1}, z_{\tau+1}) = 0$ and concluding that $r(z_\tau, \tilde{z}_{2\tau+1}) + r(\tilde{z}_{2\tau+1}, z_{\tau+1}) \leq r(z_\tau, z_{\tau+1})$. □

The next lemma establishes that if all the learners are currently playing a ν-best response then there is a zero resistance path to a ν-robust state in which they play the same way. To state this precisely define a partial ordering $\succeq$ over states. We say that the state z is *as least as large* as state z', written $z \succeq z'$, if for $j = 1, 2$ $D^j(z) \geq D^j(z')$, and $\overline{\alpha}^j(z)$ is consistent with $\overline{\alpha}^j(z')$ in the sense that $(N - D^j(z'))\overline{\alpha}^j(z') = (N - D^j(z))\overline{\alpha}^j(z) + (D^j(z) - D^j(z'))\tilde{\alpha}^j$ for some action profile $\tilde{\alpha}^j \in \Delta^{D^j(z) - D^j(z')}(A^j)$. This says that we can get from z' to z by making some agents discontent.

Lemma 6. *If $z \succeq \hat{z}$ and $\hat{z}$ is ν-robust then there exists a zero resistance path (of length 1) $\mathbf{z}$ from z to $\hat{z}$.*

The next lemma says that in calculating least resistance paths we may assume that discontent agents remain discontent. We refer to it as the *no cost to staying discontent* principle. Formally we have:

Lemma 7. *For any path $\mathbf{z} = (z_0, z_1, \ldots, z_t)$ starting at any z_0 then there is a path $\tilde{\mathbf{z}} = (\tilde{z}_0, \tilde{z}_1, \ldots, \tilde{z}_{\tilde{t}})$ with $\tilde{z}_0 = z_0$ and $r(\tilde{\mathbf{z}}) \leq r(\mathbf{z})$ satisfying the property that $\tilde{z}_\tau \succeq \tilde{z}_{\tau-1}$ and $\tilde{z}_{\tilde{t}} \succeq z_t$ for all $1 \leq \tau \leq t$.*

Proof. If $r(\mathbf{z}) = \infty$, for any $\tilde{x}_0 \in X(z_0)$ and any $\tilde{t} = 1$, take $\tilde{x}_1$ to have all learners discontent $D^j(\tilde{z}_\tau) = N - \#\Xi^j$ for both j and note that $r(\tilde{\mathbf{z}}) < \infty$ since we may have all learners tremble. It follows that $\tilde{z}_1 \succeq \tilde{z}_0, z_t$.

Next, suppose that $r(\mathbf{z}) < \infty$. We may assume from Lemma 5 that in $\mathbf{z}$ the least resistance transitions have agent transitions in which no discontent trembles and every content plays the action with which they are content. We will now find a path with $\tilde{t} = t$ and prove that if $\tilde{z}_\tau \succeq z_\tau$ we can find a state satisfying $\tilde{z}_\tau \succeq z_\tau, \tilde{z}_{\tau-1}$ and $r(\tilde{z}_\tau, \tilde{z}_{\tau+1}) \leq r(z_\tau, z_{\tau+1})$. To do this use the fact that $\tilde{z}_\tau \succeq z_\tau$ to order the agents of each player j so that the first $N - D^j(\tilde{z}_\tau) - \#\Xi^j$ agents in

$\tilde{x}_\tau \in X(\tilde{z}_\tau)$ have exactly the same type as the first $N - D^j(\tilde{z}_\tau) - \#\Xi^j$ agents in $x_\tau \in X(z_\tau)$. Observe that $r(z_\tau, z_{\tau+1})$ is determined by a particular target $x_{\tau+1} \in X(z_{\tau+1})$ and realizations $\mathcal{T}, \sigma, \mathcal{R}$, and that $r(z_\tau, z_{\tau+1}) = \#\mathcal{T}^1 + \#\mathcal{T}^2$ since σ_τ is feasible as we have assumed a finite resistance path. Also because $\tilde{z}_\tau \succeq z_\tau$ we have $\mathcal{A}^j(z_\tau) \subseteq \mathcal{A}^j(\tilde{z}_\tau)$ and the realization σ_τ is feasible for $\tilde{x}_\tau$ so we set $\tilde{\sigma} = \sigma$. We also define $\tilde{\mathcal{R}}$ to be $\mathcal{R}$ applied only to those agents who are content in $\tilde{x}_\tau$, that is, discontent agents are inactive, but content agents are active if and only if the corresponding agent did in $\mathcal{R}$. Now let $\tilde{\mathcal{T}}$ be $\mathcal{T}$ applied to those learners who are content in $\tilde{x}_\tau$. Given $\tilde{\sigma}, \tilde{\mathcal{T}}$ and $\tilde{\mathcal{R}}$, take $\tilde{x}_{\tau+1} \in X(\tilde{z}_{\tau+1})$ to be the corresponding agent state. Then $r(\tilde{z}_\tau, \tilde{z}_{\tau+1}) = \#\tilde{\mathcal{T}}^1 + \#\tilde{\mathcal{T}}^2 \leq \#\mathcal{T}^1 + \#\mathcal{T}^2 = r(z_\tau, z_{\tau+1})$ since $\mathcal{T}$ applies to every agent to which $\tilde{\mathcal{T}}$ applied. By construction no agent is content in $\tilde{x}_{\tau+1}$ unless she has the same type as in $\tilde{x}_\tau$ so certainly $\tilde{z}_{\tau+1} \succeq \tilde{z}_\tau$. Also by construction every agent who is content in $\tilde{x}_{\tau+1}$ has the same type as the corresponding agent in $x_{\tau+1}$ so indeed $\tilde{z}_{\tau+1} \succeq z_\tau$. □

We introduce a concept that captures the support of mixed strategy profiles that correspond to content agents play $\Delta^{N-\#\Xi^j}(A^j)$. More precisely, the *j-width* of a state z denoted $w^j(z) \in \mathbb{Z}_+$ is the number of distinct types for content learners of player j. The *width* of a state z is $w(z) = w^1(z) + w^2(z)$. Observe that pure ν-robust states z have $w(z) = 2$.

We now introduce the idea of a *proto ν-robust* state z, which is a state in which all content agents from either population j are playing a ν-best response to any $\alpha^{-j} \in \mathcal{A}^{-j}(z)$. We divide these into three types: a *totally discontent* state is one in which $w(z) = 0$ so all learners of both players are discontent; a *semi-discontent* state in which all learners of one player are discontent but $w(z) > 0$ so at least one learner of the other player is content, and a *standard* state in which at least one learner of each population is content. The next result characterizes transitions between states that involve proto ν-robust states with the property that paths have no resistance.

Lemma 8. *The following statements hold for $\nu = 0$ if the game has a pure strategy Nash equilibrium.*

(1) If z is totally discontent there is a zero resistance path to every ν-robust state.

(2) If z is proto ν-robust for $\nu > 0$ there is a zero resistance path to a ν-robust state $\hat{z}$; and if z is standard we can choose $\hat{z}$ so that $w(z) \geq w(\hat{z})$.

(3) If z is not proto ν-robust there exists a zero resistance path to a state $\tilde{z}$ with $w(z) > w(\tilde{z})$.

3.3 Learning Dynamics When $\epsilon = 0$

We proceed to establish that, under the stated assumptions, all non-ν-robust states are transient states of the process without trembling (that is, when $\epsilon = 0$). It readily follows that Lemma 8 holds in this case. A state $z \in Z$ is transient if there is a positive probability that the process will never return to state z at any point in time. Combined with Theorem 1 the next lemma establishes that ν-robust states are the only absorbing states of the process.

Lemma 9. *States that are not ν-robust are transient when $\epsilon = 0$.*

Proof. If states are proto ν-robust there is a zero resistance path to a ν-robust state by Lemma 8 part (2), otherwise, by Lemma 8 part (3), there is a zero resistance path to a state with strictly less width. As long as the system does not reach a proto ν-robust state, it has positive probability of moving along zero resistance paths to states with strictly lower width, applying part (2) and (3) of Lemma 8, until it visits a proto ν-robust state with $w > 0$ or reaches a totally discontent state, from which it has a positive probability of being absorbed at a ν-robust state as established in Lemma 8 part (1). □

Since there are a finite number of states, every state is either recurrent or transient, so when $\epsilon = 0$ the system will eventually be absorbed at a ν-robust state and thus at an approximate equilibrium. We are not restricted to consider pure equilibria (as in Young (2009), Pradelski and Young (2012)), and we also obtain convergence to mixed approximate equilibria (similar to Foster and Young (2006), Hart and Mas-Colell (2006), and Fudenberg and Levine (2014)). Notice that we do not impose acyclicity in order to show that the system converges to equilibrium states.

Because the transition kernel P_ϵ is regular, Young (1993, Theorem 4) implies that as $\epsilon \to 0$ the unique ergodic distributions μ^ϵ have a unique limit distribution μ which is one of possibly many ergodic distribution for P_0. It follows from Theorem 1 and Lemma 9 that the limit distribution μ can place weight only on ν-robust states, that is, if $\overline{z} \in \text{supp}(\mu) = \{z \in Z | \mu_z > 0\}$ then $\overline{z}$ is a ν-robust state.

From this point on we may assume that there is no semi-discontent proto ν-robust state. For if there was such a state then by Lemma 4 there is a unique ν-robust state, and by Lemma 9 the unique limit distribution μ places all weight on the unique ν-robust state.

Assumption 4. *ν is small enough that there is no $\hat{a}^j \in A^j$ such that for all $\alpha^{-j} \in \Delta^N(A^{-j})$ we have $u^j(\hat{a}^j, \alpha^{-j}) + \nu \geq \max_{a^j \in A^j} u^j(a^j, \alpha^{-j})$.*

We now maintain Assumption 4 for the rest of the paper.

3.4 Basin, Radius and Circuit

In characterizing the ergodic distribution μ_ϵ for small ϵ, we combine some standard technical tools and the more recent method of circuits developed by Levine and Modica (2016). Specifically, we define the *basin* of the ν-robust state z to be the set of states for which there is a zero resistance path to z, and no zero resistance path to some other ν-robust state z' when $\epsilon = 0$. We can equivalently define the basin of the ν-robust state z as the set of starting states that lead to state z with probability one according to P_0. We let r_z denote the *radius* of the ν-robust state z; this is defined to be the least resistance of paths from z to states out of its basin. More generally, if S is a union of limit sets, then the radius r_S is the least resistance path from S out of the basin of S, that is, to states where there is a positive probability of being absorbed outside of S. We further define the *stochastically stable states* to be those with non-vanishing ergodic probability as $\epsilon \to 0$,

that is the states z such that $\lim_{\epsilon\to 0}\mu_z^\epsilon > 0$. Let $R(z,z')$ denote the least resistance of any path that starts at z and ends at z'. We say set of ν-robust states Ω is a *circuit* if for any pair $z, z' \in \Omega$ there exists a least resistance *chain* - meaning a sequence $\mathbf{z} = (z_0, z_1, \ldots, z_t)$ with $z_0 = z$ to $z_t = z'$ with $z_k \in \Omega$ and $R(z_k, z_{k+1}) = r_{z_k}$ for $k = 0, \ldots, t-1$. That is, one of the most likely (lowest order of ϵ) transitions from z_0 is to z_1, one of the most likely transitions from z_1 is to z_2, and so forth.

Theorem. *If all ν-robust states and are in the same circuit then $\frac{\mu_z^\epsilon}{\mu_{z'}^\epsilon} \sim \epsilon^{r_{z'}-r_z}$, and in particular the set of stochastically stable states is exactly the ν-robust states with the largest radius.*

This is Theorem 9 in Levine and Modica (2016), specialized to the case where the only recurrent classes when $\epsilon = 0$ are singletons, and there is a single circuit. To understand why this is true we sketch two proofs. First we use the method of Ellison (2000) to show that the stochastically stable states are those with the largest radius. For any target z define the *modified co-radius* $c_z = \max_{z'} \min_{\mathbf{z}=(z',z_1,\ldots,z)} r(z',z_1) + r(z_1,z_2) + \ldots + r(z_{t-1},z) - r_{z_1} - r_{z_2} - \ldots - r_z$. Define the *modified co-radius* c_S of a limit set S to be the minimum over $z \in S$ of c_z. Ellison shows that a sufficient condition for a set S of ν-robust states to be stochastically stable is that $r_S > c_S$. If we let $\bar{r}$ denote the largest radius of any ν-robust state then the set S of ν-robust states with radius $\bar{r}$ itself has radius r_S at least equal to $\bar{r}$. By assumption, all ν-robust states are in the same circuit, so we can compute an upper bound on c_S by considering, for each state $z' \notin S$, a least resistance chain from z' to z, meaning a sequence of states for which the resistance $R(z_k, z_{k+1}) = r_{z_k}$. The modified resistance of this chain is $r_{z'} + r_{z_1} + \ldots + r_z - r_{z_1} - r_{z_2} - \ldots - r_z = r_{z'}$ and since $r_{z'} < \bar{r} = r_S$ the conclusion follows.

For the sharper result that $\frac{\mu_z^\epsilon}{\mu_{z'}^\epsilon} \sim \epsilon^{r_{z'}-r_z}$, we use the method of Levine and Modica (2016). For any ν-robust state z we consider *trees with root z,* where the nodes of the tree are all of the ν-robust states and the resistance of the tree is the sum of all the $R(z_k, z_{k+1})$ where z_{k+1} is the successor of z_k. Using the Markov chain tree formula (see for example Bott and Mayberry (1954)) it follows, as noted by Freidlin and Wentzell (1984), that $\log(\mu_z^\epsilon/\mu_{z'}^\epsilon)/\log\epsilon$ converges to the difference in resistance between the least resistance tree with root z and that with root z'. Notice that since each ν-robust state must be in the tree, the resistance of connecting that node is at least r_{z_k}, so that the least resistant tree cannot have less resistance than the sum of the radii of all nodes except the root. We now show there is a tree with exactly that resistance by building it recursively. Place the root z first. There must be some remaining node that can be connected to the tree at resistance equal to the radius because all stable states are in the same circuit. Add that node to the tree with that resistance. Continuing in this way we eventually construct a tree in which the resistance is exactly the sum of radii of all but the root node. It follows that the difference in resistance between the least resistance tree with root z and root z' is exactly the difference in the radii which is what is asserted in the Theorem.

4 Exact Pure Strategy Equilibria

In this section, we assume that pure strategy Nash equilibria exist, and set the social comparison parameter $\nu = 0$.

Assumption 5. *The game has at least one pure strategy Nash equilibrium.*

We also make the following generic assumption about the grid.

Assumption 6. *Player j has a unique best response to every $\alpha^{-j} \in \Delta^N(A^{-j})$.*

In light of Lemma 3 part (2), the existence of at least one 0-robust state is guaranteed. As already mentioned, discontent agents play a fundamental role in determining least resistance paths. On a path that moves away from a 0-robust state, content agents, must tremble, and so the path has positive resistance. In addition to the random mistakes, every agent that is not playing a best response transitions to discontentment with no resistance irrespective of her current individual state. For each 0-robust state z, we define the non-negative integer number $r_z^j \in \mathbb{Z}_+$ for player j to be the least number of learners of player $-j$ that need to deviate for there to be a learner of player j such that the learner is not using a best response. Then in finding least resistance paths out of the basin of a 0-robust state z, we will establish that the critical threshold to be considered is the smaller of r_z^1 and r_z^2. We will use this to characterize the radius of a 0-robust state z, and show that the minimum resistance to any other 0-robust state z' is the same for every z'.

Lemma 10. *The radius of a 0-robust state z is $r_z = \min\{r_z^1, r_z^2\} > 0$. Moreover for any 0-robust state $\overline{z} \neq z$ there is a path from z to $\overline{z}$ that has resistance r_z.*

Proof. Consider a least resistance path $\mathbf{z}$ from a 0-robust state z to any 0-robust state $\overline{z}$. From Lemma 7 we know that there exists a path $\tilde{\mathbf{z}} = (\tilde{z}_0, \tilde{z}_1, \ldots, \tilde{z}_t)$ from $\tilde{z}_0 = z$ with $r(\tilde{\mathbf{z}}) \leq r(\mathbf{z})$ and $\tilde{z}_t \succeq \overline{z}$. Since $\tilde{z}_t \succeq \overline{z}$ and $\overline{z}$ is 0-robust, by Lemma 6 there is a zero resistance path from $\tilde{z}_t$ to $\overline{z}$. Then it is sufficient to compute $r(\tilde{\mathbf{z}})$ in order to obtain the radius of z.

We begin by characterizing the basin of the 0-robust state z. Lemma 7 implies it suffices to consider $D^j(\tilde{z}_\tau)$ for $\tau \leq t$, since discontent learners stay discontent on the path $\tilde{\mathbf{z}}$. If for both players j we have $D^j(\tilde{z}_\tau) < r_z$ we show that $\tilde{z}_\tau$ is in the basin of z. Suppose discontents play the unique best response a^j in each population j, which gives rise to a feasible profile of actions, that they do not tremble, are active and become content. This transition has no resistance. In the resulting state all learners are content and playing a^j the unique best response to any feasible α^{-j}; that is, the state is z. Hence we have a zero resistance path back to z. However to be in the basin there must not be a zero resistance path to some different 0-robust state $\hat{z}$. We show that any such path starting at $\tilde{z}_\tau$ has a resistance of at least one. Moving along any such path requires that for all contents of at least one player j it must be that $\hat{a}^j \neq a^j$ by Assumption 6. Since $D^j(\tilde{z}_\tau) < r_z$ for $j = 1, 2$ all content agents are playing a best response which implies that any transition $\tilde{z}_\tau \to z'$ on the path to $\hat{z}$ we must have that $D^j(z') > D^j(\tilde{z}_\tau)$ for at least one player j. But in this transition

at least one content agent who is playing a best response becomes discontent, by Lemma 2 this transition has resistance at least one.

Next, we establish that any path from z to any other 0-robust state $\hat{z}$ has resistance r_z. We can equivalently show that if we have $D^j(\tilde{z}_\tau) \geq r_z^{-j}$ for either player j then there exists a zero resistance path to any 0-robust state. Suppose that $D^j(\tilde{z}_\tau) \geq r_z^{-j}$ for one player j. Then consider a transition where the profile α^j is such that all contents in $-j$ are active and observe a better response played a committed agent, so become discontent. Whereas learners in j neither are inactive and do not tremble. This transition has zero resistance. The next transition has a profile α^{-j} so that contents in j are active and get a signal about a better response provided by a committed agent, do not tremble, and become discontent while agents in $-j$ do not tremble, are inactive, and continue to be discontent. It follows that this transition has no resistance. By Lemma 8 there is a zero resistance path to any 0-robust state. □

We have shown that computing the radius of a 0-robust state requires us to find two thresholds, one for each player role, that represent the least number of learners that are able to move all learners to discontentment. In words, Lemma 10 establishes that as long as the system remains within the basin of a 0-robust state, not too many discontent agents are experimenting with new strategies and the rest of the learners are content, and playing a best response. Thus from states in this basin the discontent learners are likely to find their way back to equilibrium. Interestingly, we find that once the system leaves the basin of a 0-robust state, there must be lots of agents trying new strategies, which in turn pushes everyone into the state of searching. Since once all learners are discontent the system may transition to any other 0-robust state with no resistance, the result entails that there is a single circuit containing all 0-robust states.

Theorem 2. *All 0-robust states form a single circuit. If z and z' are 0-robust states, then $\frac{\mu_z^\epsilon}{\mu_{z'}^\epsilon} \sim \epsilon^{r_{z'}-r_z}$ and moreover those states with largest radius are stochastically stable.*

Proof. Lemma 10 implies that every 0-robust state belongs to the same circuit. Then the conclusion of the Theorem is immediate from Theorem 9 of Levine and Modica (2016). □

The preceding result implies that the dynamics take a simple form in which for any pair of pure strategy two-player Nash equilibria, which correspond to 0-robust states z and z' respectively, the system would spend approximately $\epsilon^{r_{z'}-r_z}$ times as much time at the pure equilibrium associated to z as at the pure equilibrium associated to z'. It follows from the fact that the probability of leaving the state z is of order ϵ^{r_z}, then the expected length of time spent at z is ϵ^{-r_z}, together with the fact the probability of reaching the other state z' is of order $\epsilon^{r_{z'}}$. Moreover, Theorem 2 provides a characterization for computing all relative ergodic probabilities of equilibria and an important observation: the system spends most of the time at 0-robust states, associated to its corresponding Nash equilibrium, that have big radii as they are hard to leave. This characterization is based on the property that random search is relatively as likely to find one equilibrium as another, meaning that once an equilibrium is left there is no differentiation as to which equilibrium the system is

likely to move next: what matters is the leaving time. Note that our characterization is simple in that it only requires one to compute the radius r_z of each 0-robust state z.

5 Comparison to Best Response Dynamics

Having established our main characterization of exact pure equilibria, we compare our results with those from a two-population version of the best-response-plus-mutation dynamic Kandori et al. (1993) (KMR henceforth).[14] The specific version of the best-response dynamic we study is called *best-response with inertia* and assumes that in each period with some probability $1 > \lambda > 0$ each agent independently continues to play the same action as in the previous period, with probability $1 - \lambda - \epsilon$ they play a best response to the population distribution of opponent's actions, and with probability ϵ they choose randomly over all possible actions. While in the one population case the assumption that $\lambda > 0$ plays little role, as KMR show by example it can lead to better behaved and more sensible dynamics in the two population case with results similar to those with one population.[15] For an analysis of this dynamic see Samuelson (1994).

Our next example illustrates how our low information dynamic can be strictly more selective than the best response with inertia dynamic. Consider the following game G_1:

	A	B	C	D
A	5,5	0,0	0,0	0,0
B	0,0	10,10	0,9	9,0
C	0,0	9,0	10,10	0,9
D	0,0	0,9	9,0	10,10

This game is a special case of what Young (1993) calls a acyclic game by our Assumption 1. Define a *best response path* to be a sequence of action profiles $(a_1, a_2, \ldots, a_t) \in (A^1 \times A^2)^t$ in which for each successive pair of action profiles (a_k, a_{k+1}) only one player changes action, and each time the player who changes chooses a best response to the action the opponent played in the previous period. Under Assumption 1, a game is *acyclic* if for any profile of actions there exists a best response path starting at that profile and ending at some pure Nash equilibrium of the game.[16] In a coordination game a best response path of two steps always does the trick: first we move one player to a best response, then the other. As Young (1993) notes for acyclic games at least one pure Nash equilibrium exists by definition. Following Samuelson (1994), we also observe that in acyclic games with generic payoffs the limit distribution for the best response plus mutation dynamic with inertia contains only singleton pure Nash equilibria. Samuelson (1994) does not provide a proof of this so we give one in the Appendix.

[14] Young (1993) considers a very similar dynamic in which each period only one agent per role acts per period and agents base their decisions on a random subset of the "last few" periods of play.

[15] In the study of Markov chains this sort of inertia is called "laziness," and is used to turn periodic irreducible chains into aperiodic ones; it serves the same purpose here by ruling out limit cycles.

[16] The definition in Young (1993) also requires that the sequence of best responses ends at a Nash equilibrium where each player's best response is unique, which is ensured by Assumption 1.

In the coordination game G_1 there are four pure strategy equilibria which we label A, B, C, D. To escape from A requires about $N/3$ of one population to mutate, say to B, so that is the radius of A. On the other hand to escape from B, C, D requires only about $N/11$ of one population to mutate, from B to C, from C to D and from D to B, so those are the radii of B, C, D. Hence with our dynamic A is stochastically stable as it has the largest radius among pure strategy equilibria. Note that in either dynamic B, C, D have equal ergodic probability by symmetry, that is, $\mu^\epsilon_B = \mu^\epsilon_C = \mu^\epsilon_D$. We now argue that for the best-response-plus-mutations dynamic with inertia the set B, C, D is stochastically stable and A is not despite having the largest radius. Define S to be the union B, C, D. The radius r_S of S is at least $N/2$ since if $1/2$ of one population is playing in B, C, D one of those strategies must earn at least $(1/2)(6 + 1/3)$ while playing A yields no more than $5/2$. On the other hand the co-radius of S is about $N/3$ since A is the only pure Nash equilibrium outside of S and it takes at least that amount to escape from A. Hence by Ellison's theorem the radius of S is bigger than the co-radius so S contains all stochastically stable states: in this case all three states B, C, D have equal probability by symmetry.

One of the reasons that the set S is stochastically stable under the best response with inertia dynamic is that when agents are at the equilibrium B and enough opponents switch to strategy C, agents' behavior adjusts under the assumption that participants can immediately see that choosing C is the optimal strategy. Identifying actions therefore allows them to move to C. In contrast, we focus on situations where agents do not have common knowledge of the structure of the game. We think it is plausible that agents will then base their decisions upon social information. Identifying payoffs, but not actions, allows agents to potentially move from B to A, and once they arrive at A stay there for a long time.

Note that our dynamic can also predict a different equilibrium even when the KMR dynamic with inertia has a singleton stochastically stable set. Suppose that a player obtains $\kappa > 0$ instead of 0 when choosing B against C in the coordination game G_1. To escape from B now about $N/(11 - \kappa)$ of one population needs to mutate so this is the radius of B. Our dynamic selects A as it continues to have the largest radius among pure strategy equilibria. The set S=B, C,D still contains all stochastically stable states. Let $S' = A, B$. The radius of S' is about $N/(11 - \kappa)$ of one population since escaping from S' requires this agents to move to C or D; and the co-radius is about $N/11$. Because the radius of S' is larger than its co-radius the stochastically stable states are in S'. Combining this with the fact that they also lie in S shows that the unique stable state is B although its radius is smaller than the radius of A.

We finally study a smaller coordination game. It is convenient to define a *block* to be any set $W = W^1 \times W^2$ with non-empty subsets of actions $W^j \subseteq A^j$ for $j = 1, 2$ and the associated *block game* is the original game restricting payoffs and actions to the W block. If we consider the block game associated to the upper left 2×2 block of G_1, both dynamics coincide in that B is the only stochastically stable set since B has a radius of about $2N/3$ of one population because it requires about $2N/3$ of one population to change to A in order to leave B. This suggests that there might be games in which participants would play according to some equilibrium in the long run whether

they base their choices on others' behavior or payoffs.

6 Approximate and Mixed Strategy Equilibria

We now turn to the stochastically stability analysis of general finite two player games, where pure strategy equilibria need not exist. Since generically we cannot guarantee that mixed strategy equilibria are attainable by population play represented on the grid $\Delta^N(A)$, we allow agents to evaluate their current actions according to a strictly positive social comparison parameter, that is, implicitly we consider ν-best responses for $\nu > 0$. A mixed profile (α^j, α^{-j}) is a mixed strategy approximate Nash equilibrium of the stage game if for all population j with play α^j it follows that $u^j(\alpha^j, \alpha^{-j}) \geq u^j(\tilde{\alpha}^j, \alpha^{-j}) - \nu$ for all $\tilde{\alpha}^j \in \Delta(A^j)$. In this section, we provide the complete structure of the dynamics between equilibria. We will show that the system starting at a mixed equilibrium either moves with resistance 1 towards mixed equilibria with smaller supports or transitions along resistance 1 paths to every equilibrium. On the other hand we establish that if the system begins at pure equilibria, when the radius is reached it transitions to every equilibrium. Hence, intuitively, the big picture of the dynamics that we obtain is that states that correspond to mixed equilibria travel "downhill" until either all learners become discontent or reach a pure equilibrium which is subject to movements to a state in which all learners become discontent.

The generic assumption about unique best responses, Assumption 6 in Section 4, has no straightforward counterpart for approximate equilibria, so we impose a generic assumption that ensures separability between 0-robust states. Specifically, for each strict Nash equilibrium profile $a = (a^j, a^{-j}) \in A$ of the stage game define $\underline{\rho}_a^j(\nu) \in [0,1]$ for player j to be the minimum probability $\alpha^{-j}(a^{-j})$ such that a^j is not the *only* ν-best response to the corresponding $\alpha^{-j} \in \Delta^N(A^{-j})$.[17] Analogously, let $\rho_a^j(\nu) \in [0,1]$ for player j be the infimum probability $\alpha^{-j}(a^{-j})$ such that a^j is not a ν-best response to the corresponding $\alpha^{-j} \in \Delta^N(A^{-j})$.[18]

Assumption 7. *For each Nash equilibrium a and at least one player j, $\rho_a^j(0) < \underline{\rho}_a^1(0) + \underline{\rho}_a^2(0)$.*

The existence of ν-robust states is guaranteed by Lemma 3 part (2). Before proving the result of separation between ν-robust states we need the following additional definitions. For each pure ν-robust state z with content actions corresponding to a^j, a^{-j}, we define the non-negative integer $\underline{r}_z^j \in \mathbb{Z}_+$ for player j to be the least number of learners of player $-j$ that need to deviate so that a^j is no longer the *only* ν-best response to any feasible play of population $-j$. Similarly, let the non-negative integer $\overline{r}_z^j \in \mathbb{Z}_+$ be the least number of learners of player $-j$ that must deviate for there to be a learner of player j such that the learner is not playing a ν-best response. Observe that $\overline{r}_z^j \geq \underline{r}_z^j$ and $N - \#\Xi^{-j} \geq \overline{r}_z^j, \underline{r}_z^j \geq 0$ for all j. As we work with finite populations, for each $x \in \mathbb{R}$ denote by $\lceil x \rceil$ (resp. $\lfloor x \rfloor$) the smallest (resp. the largest) integer greater than or equal to x (resp. not larger than x).

[17] Note that if a^j is a strictly dominant strategy and ν is sufficiently small then a^j is the only best response regardless of the actions of population $-j$.

[18] By Assumption 1 it must be that $\underline{\rho}_a^j(0) = \rho_a^j(0) > 0$, while it may be that $\underline{\rho}_a^j(\nu) \neq \rho_a^j(\nu)$ if $\nu > 0$.

Lemma 11. *There is a χ and γ with $N/M > \gamma$ and $\nu < \chi$ such that for every pure ν-robust state z we have for at least one j that $\overline{r}_z^j \leq \underline{r}_z^1 + \underline{r}_z^2$ and for both j that $\underline{r}_z^j \geq 1$.*

We assume from now on that the parameters ν and N/M are such that we have separation between ν-robust states.

Assumption 8. *$\nu < \chi$ and $N/M \geq \gamma$ where γ is large enough and χ is small enough that Lemma 11 holds.*

Lemma 12. *The radius of a pure ν-robust state z is $r_z = \min\{\overline{r}_z^1, \overline{r}_z^2\}$, and if $\overline{z}$ is any ν-robust state there is a path from z to $\overline{z}$ with resistance equal to r_z.*

Proof. Let $\mathbf{z}$ be a least resistance path from a pure ν-robust state z to any ν-robust state $\overline{z}$. Lemma 7 implies there is a path $\tilde{\mathbf{z}} = (\tilde{z}_0, \tilde{z}_1, \ldots, \tilde{z}_t)$ from $\tilde{z}_0 = z$ with $r(\tilde{\mathbf{z}}) \leq r(\mathbf{z})$. Moreover, since $\tilde{z}_t \succeq \overline{z}$ and $\overline{z}$ is ν-robust there is a zero resistance path from $\tilde{z}_t$ to $\overline{z}$ by Lemma 6. Hence the radius of z may be computed as the resistance of $\tilde{\mathbf{z}}$. Let a^j, a^{-j} be the profile of content actions corresponding to z.

Suppose for player j $\overline{r}^j \leq \underline{r}_z^1 + \underline{r}_z^2$ and $\overline{r}^j \leq \overline{r}^{-j}$. It suffices to consider the case where $D^j(\tilde{z}_\tau) < \underline{r}^{-j}$ and $\underline{r}^j < D^{-j}(\tilde{z}_\tau) < \overline{r}^j$. In population $-j$ content agents are playing a ν-best response while discontent learners need not be. Consider the transition in which nobody trembles, discontents play a^{-j} and are active. This transition has no resistance. If discontent agents in population j do not play a ν-best response, are active and nobody trembles; we reach this transition with zero resistance. In the former case, since $\tilde{z}_\tau \succeq \tilde{z}_{\tau-1}, z_\tau$ for all τ the number of discontent learners in population j can be increased only if $D^j(\tilde{z}_\tau) < \underline{r}^{-j}$ increases, and since $\underline{r}^{-j} \geq 1$ this requires at least one content agent that is playing a ν-best response to become discontent so this transition has resistance at least one by Lemma 2. This characterizes the basin of z. Next, we show that as long as we leave the basin we can reach any other ν-robust state. Assume $D^{-j}(\tilde{z}_\tau) \geq \overline{r}_z^j$, then player j content agents are not playing a ν-best response to some feasible profile of actions $\alpha^{-j} \in \mathcal{A}^{-j}(\tilde{z}_\tau)$. Let them be active, and no agents tremble. This transition has no resistance. In the following state, suppose that all the discontent agents in j induce a feasible action so that content agents in $-j$ are not playing a ν-best response. Then discontent agents in j and $-j$ are inactive, content agents in $-j$ are active to the fact that they are not playing ν-best response and there are no trembles. This zero resistance transition results in a state where all agents are discontent. By Lemma 8 there is a zero resistance path to any ν-robust state. □

Therefore the least resistance of paths leaving the basin of a pure ν-robust state z is characterized in terms of the thresholds $\overline{r}_z^1$ and $\overline{r}_z^2$, which represent the least number of learners that need to deviate so that a critical mass of agents are no longer playing a ν-best response which results in all learners eventually becoming discontent. Then the system leaves the basin of z, that is the corresponding pure strategy approximate Nash equilibrium, but in this case it may transition to any other ν-robust state, which may correspond to either a mixed or pure approximate Nash

equilibrium. Note that if a is pure Nash equilibrium the corresponding pure ν-robust state has radius $\min_j \lceil (N-M)\rho_a^j(\nu) \rceil$.

We proceed to calculate the least resistance of paths exiting the basin of mixed ν-robust states z, thus with width $w(z) > 2$. In general, there will be multiple approximate mixed strategy equilibria in the neighborhood of mixed strategy equilibrium so one might expect to move between those approximate mixed equilibria through one agent changing play at a time. We next formalize this intuition by showing that the radius of ν-robust states with width $w(\cdot) > 2$ is one, and that once we leave the basin we either move to another ν-robust state with weakly smaller support of the content action distribution and $w(\cdot) > 2$, for example, a state with exactly the same support and slightly different numbers of content agents using each action; or reach any other ν-robust state. We now introduce a notion that captures the largest mass of learners in the support of the current distribution of content actions: for any state z let the *depth* $h(z) \in \mathbb{Z}_+$ be the largest number of learners playing an action in the support of $\overline{\alpha}(z)$ the action profile that corresponds to the aggregate play of contents and committed agents.

Lemma 13. *The radius of any ν-robust state z with $w(z) > 2$ is $r_z = 1$, and there is either a path with resistance 1 to every ν-robust state $\overline{z}$ or to a ν-robust state $\tilde{z}$ with $w(\tilde{z}) \leq w(z)$ and either $w(\tilde{z}) < w(z)$ or $h(\tilde{z}) > h(z)$.*

Proof. By Lemma 8 it suffices to consider paths $\mathbf{z}$ from z to any proto ν-robust state z'. Because z is ν-robust, all learners are content and play a ν-best response. Hence on any transition from z to some other proto ν-robust state $\hat{z}$ has $r(z, \hat{z}) \geq 1$, since by Lemma 2 at least one content learner that is playing a ν-best response must tremble for the system to leave z. We apply the following algorithm to construct least resistance paths between ν-robust states. In z, using $h(z)$ identify action $\tilde{a}^j$ for one player j that is played by the largest number of learners in $\text{supp}(\overline{\alpha}^j(z))$. Suppose that in $z \to z''$ one content player j agent in state $a^j \in A^j$ trembles and become discontent while all the other content agents are inactive and do not tremble. This implies that $r(z, z'') = 1$, and $w(z'') \leq w(z)$ by construction. If z'' turns out to be a proto ν-robust state, consider $z'' \to z'''$ where the unique discontent learner plays the action $\tilde{a}^j \neq a^j$ (notice that $\tilde{a}^j \in \text{supp}(\overline{\alpha}^j(z''))$), is inactive, and does not tremble, while the rest of the learners do not tremble and are inactive. Thus z''' is ν-robust and $h(z''') > h(z)$. Otherwise, z'' is not proto ν-robust, so there is a zero resistance path $\mathbf{z}$ from z'' to a state $\tilde{z}$ with $w(\tilde{z}) < w(z'')$ by Lemma 8. If $\tilde{z}$ is a proto ν-robust state we are done. If $\tilde{z}$ is not a proto ν-robust state we proceed as in the last step, applying repeatedly Lemma 8, we construct a zero resistance path $\mathbf{z}'$ from $z'_0 = \tilde{z}$ to other state $z'_t = \overline{z}$ with $w(z_{\tau+1}) < w(z_\tau)$ for $t \geq \tau \geq 0$ until we reach a proto ν-robust state $\overline{z}$ (which could be either pure or totally discontent). Recall that if $w(z') = 1$, z' cannot be a proto ν-robust state because it would be semi-content which we ruled out (Assumption 4). By Lemma 8, from a totally discontent state we can reach any ν-robust state. □

Equipped with these lemmas, we can now state the main result of this section:

Theorem 3. *There is a single circuit which contains all ν-robust states. For ν-robust states z and z' we have $\frac{\mu_z^\epsilon}{\mu_{z'}^\epsilon} \sim \epsilon^{r_{z'}-r_z}$, so in particular the stochastically stable states are those with the largest radius.*

Proof. The fact that all ν-robust states are connected by least resistance paths follows from Lemmas 12 and 13. The second conclusion follows from Theorem 9 of Levine and Modica (2016), and the third follows immediately from the second. □

A key implication of our characterization is the analysis of the relative likelihood of pure and mixed approximate equilibria. Suppose that the ν-robust state z corresponds to a mixed approximate equilibrium, so $r_z = 1$ by Lemma 13, and that the ν-robust state z' corresponds to a pure approximate equilibrium with a radius equals to the least $\lceil (N-M)\rho_a^j(\nu) \rceil$ by Lemma 12. Since we work with generic games, pure equilibria are strict, hence for large enough N we have $\lceil (N-M)\rho_a^j(\nu) \rceil > 1$ by Lemma 11 so in the limit that we consider, with N fixed and $\epsilon \to 0$, pure equilibria are far more likely than mixed equilibria.

7 Pure Versus Mixed Equilibria

In this section, we explore our results concerning the stability of pure and mixed equilibria in the context of an example. We argue that our model has sharp predictions in environment where agents have very limited information. We then propose a learning model that incorporates more memory and information, and show that we may obtain different predictions; in particular, we find that mixed equilibria may be stochastically stable even when some pure equilibria are not.

We begin by applying our results to the following *pure-vs-mixed* game G_2:

	H	T	P
H	5,3	3,5	1,1
T	2,5	5,2	1,1
P	1,1	1,1	2,2

This game has three Nash equilibria: one mixed equilibrium $((\frac{3}{5}H, \frac{2}{5}T), (\frac{2}{5}H, \frac{3}{5}T))$, the strict equilibrium (P, P), and the completely mixed equilibrium $((\frac{3}{19}H, \frac{2}{19}T, \frac{14}{19}P), (\frac{2}{19}H, \frac{3}{19}T, \frac{14}{19}P))$. Our theory says that the two mixed equilibria are less stable than the strict equilibrium.

Suppose that there is a population of $N > 13$ agents in each player role, 3 committed agents in each population, and that in every period all agents of player 1 are matched round robin with all agents of player 2 to play the pure-vs-mixed game G_2.[19] All agents are assumed to follow the learning procedure described in Section 2.2. The stage game induces a grid $\Delta^N(\{H, T, P\})$ for each player so $\alpha^j(a^j)$ is an integer multiple of $\frac{1}{N}$ for every $a^j \in \{H, T, P\}$. Each committed agent plays a specific action with probability 1. Let $\nu < 1$. We first compute the set of ν-robust states

[19] We can generalize this analysis to N large compared to M.

which consists of the set of states in which all learners are content and the current content actions correspond to the pure action profile (P, P), along with the following sets:

$$\mathcal{B} = \left\{\alpha \in \Delta(A) : \left|\frac{N-3}{N}(3\tilde{\alpha}^1(T) - 2\tilde{\alpha}^1(H)) + \frac{1}{N}\right| < \nu, \tilde{\alpha}^j(P) = 0 \text{ for } j = 1, 2, \right.$$
$$\left. \text{and } \left|\frac{N-3}{N}(3\tilde{\alpha}^2(H) - 4\tilde{\alpha}^2(T)) + \frac{1}{N}\right| < \nu\right\}$$

and

$$\mathcal{C} = \left\{\alpha \in \Delta(A) : \left|\frac{N-3}{N}(3\tilde{\alpha}^1(T) - 2\tilde{\alpha}^1(H)) + \frac{1}{N}\right| < \nu, \left|\frac{N-3}{N}(2\tilde{\alpha}^2(H) + 4\tilde{\alpha}^2(T) - \tilde{\alpha}^2(P)) + \frac{5}{N}\right| < \nu, \right.$$
$$\left. \left|\frac{N-3}{N}(3\tilde{\alpha}^2(H) - 4\tilde{\alpha}^2(T)) + \frac{1}{N}\right| < \nu, \text{ and } \left|\frac{N-3}{N}(4\tilde{\alpha}^2(H) + 2\tilde{\alpha}^2(T) - \tilde{\alpha}^2(P)) + \frac{5}{N}\right| < \nu\right\},$$

where $\tilde{\alpha}^j$ corresponds to the population play of content agents in j. Note that the sets $\mathcal{B}$ and $\mathcal{C}$ are approximate mixed equilibria which are close to the mixed equilibria of the pure-vs-mixed game G_2. We have established in Theorem 13 that ν-robust states for which the current distributions of content actions correspond to either the set $\mathcal{B}$ or $\mathcal{C}$ move along a path of resistance 1 to any other ν-robust state. We also know from Lemma 12 that ν-robust states in which the current action distribution induces the action profile (P, P) may transition to any ν-robust state along a path of resistance $\lceil \frac{N(1+\nu)-8}{5} \rceil$. Assume z is a ν-robust state with current distribution of content actions belonging to the set $\mathcal{B}$ (or the set $\mathcal{C}$) and z' is a ν-robust state in which currently the action distribution corresponds to (P, P). Our characterization of the relative likelihood of different equilibria (Theorem 3) enables us to conclude that relatively $\epsilon^{1-\lceil \frac{N(1+\nu)-8}{5} \rceil}$ times as long is spent at the pure ν-equilibrium as at the mixed ν-equilibrium, and since $N > 13$ and all mixed equilibria have a radius of 1, the pure equilibrium is far more likely than the mixed equilibria when ϵ is small.

Now let us turn away from the formal application of our model to examine our intuition about this game. Let us focus on the mixed equilibrium in which agents choose randomly between H and T versus the pure equilibrium where agents only play P. A key feature of our learning model is that agents learn how well they are doing relative to other strategies, but can never identify how they would be able to do better. In the case that participants can only use this limited information we expect that if all participants are playing an approximate mixed equilibrium in $\mathcal{B}$, then when a single individual experiments with P there is an appreciable chance that this will lead to a shift from the mixed equilibrium to the pure equilibrium and once they get there it is very unlikely that they leave.

It is useful to contrast our predictions in this example to best response with inertia that we discussed in Section 5. With some abuse of notation let denote the block $\{H, T\} \times \{H, T\}$ by HT, and as before P denotes the pure strategy equilibrium. Here we can easily show from the radius co-radius argument that the HT block contains the stochastically stable set rather than P.[20]

[20]To see this, the radius of the HT block is at least $2N/3$ because if $2/3$ of one population is playing in HT block any of these strategies must earn at least $(2/3)(3 + 4/5)$ while playing P yields at most $4/3$. But, the co-radius of HT block is about $N/5$ since at least $1/5$ of one population has to mutate to escape from P and is the only pure

Unfortunately, the dynamics of best response with inertia inside the HT block are not terribly plausible. If the inertia parameter $\lambda = 0$ players will follow a deterministic best response cycle, meaning that each outcome of the block game associated to the HT block will have equal weight in the unique limit distribution with $\epsilon = 0$. Notice that every agent switches to a best response with probability one. Since the unique limit distribution is continuous in λ this means that the time average payoff received by the players is approximately 15/4. However, this implies that the agents' time average payoff is less than their minmax, which is not a desirable property of a learning procedure (Fudenberg and Kreps (1993), Fudenberg and Levine (1995)). In the next section, we consider a variation on best response with inertia dynamic that represents the essential features of our model. We show that it converges with probability 1 to a ν-Nash equilibrium when $\epsilon = 0$, and captures our previous discussion that with more information the mixed equilibrium with support equal to the HT block is more likely than the equilibrium P.

8 High Information Social Learning

Our learning procedure thus far has focused on agents that have very limited social information and short recall about the past. We now consider "high information" models where agents both observe and remember more.

8.1 The Learning Procedure

We make three changes to the learning procedure concerning observability and memory. Previously we assumed that an agent observed $\phi^j(\alpha)[u^j] = \sum_{a^i \in \mathbf{A}^j(u^j, \alpha^{-j})} \alpha^j(a^i)$, the distribution of utilities corresponding to actions actually played. Now we assume that an agent observes the joint distribution of utilities and actions played in the stage game

$$\Phi^j(\alpha)[u^j, a^j] = \begin{cases} \alpha^j(a^i) & \text{for } a^i \in \mathbf{A}^j(u^j, \alpha^{-j}), \\ 0 & \text{for } a^i \notin \mathbf{A}^j(u^j, \alpha^{-j}). \end{cases}$$

Concerning recall, we now assume that at the beginning of a period agents can recall which actions were ν-best responses during the last $T \geq 1$ periods, and that all agents recall the last action they played, not only the content agents as in the low memory social learning model.[21] In Young (1993) adaptive dynamics agents also have social information but of a different kind, as their information is time series observations rather than cross-section. Specifically, every period only one agent per player role moves at that period, takes a size K random sample of play from the last T periods without replacement. Given this sample, certain actions are best responses, and only those have positive probability of being played. In contrast, our model allows agents to choose actions from

Nash equilibrium outside of HT block. Since the radius of HT block is larger than its coradius the stochastically stable states are contained in HT block.

[21] Note that actions that are best responses are those with the highest utilities since agents observe the payoff to each action given the presence of committed types.

the last T periods that were ν-best responses in the period they were used based on that period cross-section information.[22] Our model also differs from those papers in that agents do not take random samples. Notice that in the present model the cross-section information is not trivial since all agents in each population takes an action at once.

Formally an agent's type is $\theta_t^i \in (A^j \times \{0,1\} \times T^{A^j}) \cup \Xi^j \equiv \Theta_T^j$. There is a given initial type distribution. Subtype 0 indicates the agent is discontent, and subtype 1 indicates the agent is content. Thus, the first part of a type for learners $A^j \times \{0,1\}$ gives the previous action taken a_{t-1}^i for both content and discontent agents. The rules concerning the dynamics of contentment do not change. The final part characterizing the learner's type $T_t^i[a^j]$ is the amount of time since each action a^j was observed to be a ν-best response to α_{t-1}^{-j}: $T_t^i[a^j] = 0$ if a^j was a ν-best response to α_{t-1}^{-j}, otherwise $T_t^i[a^j] = \min\{T, T_{t-1}^i[a^j] + 1\}$. Since this is the same for all learners of player j we refer to this as the *common memory* of player j and the actions a^j for which $T_t^i[a^j] < T$ as the players's *common memory set*, which we denote by $\mathbf{A}_T^j$. This simplifies the description of the state since we can use a single memory that is relevant for all agents of a player.[23] The *individual memory set* of agent i of player j is the union of the common memory set and the last action that agent played, that is, $\mathbf{A}^i = \mathbf{A}_T^j \cup \{a_{t-1}^i\}$.

The impact of the memory set is only on the behavior of discontent agents: rather than engaging in uniform play over all actions A^j they engage in uniform play only over their individual memory set $\mathbf{A}^i$.

8.2 $T = 1$ Memory and Best Response Dynamics

Observe that discontent agents randomize over a subset of the uniform play in the low information social learning model. This is similar to best response with inertia since discontent agents may play the last period action as well as the current ν-best response. Furthermore, when $T = 1$, $\nu = 0$ and the unique best response property is satisfied, we will show that the high information social learning has similar features as best response with inertia.

In order to make this comparison formally, we must extend the state space to incorporate the current population play. Let $\Phi_t^j \in \Delta^N(\Theta_T^j \cup A^{-j})$ be a vector of population shares of the player j types in period t, which includes the description of play of the opposing population α_{t-1}^{-j} in period $t-1$. We assume that the memory length $T = 1$ in the spirit of best response with inertia. As in Section 4 we restrict attention to exact best responses, that is, $\nu = 0$. We remark that Assumption 6 implies that for each population j there is a single action a^j with $T_t^i[a^j] = 0$ and all the other actions $\tilde{a}^j \neq a^j$ have $T_t^i[\tilde{a}^j] = 1$. All actions that are not best responses to the previous population

[22] In a single-population stochastic evolutionary model, Oyama et al. (2015) consider sampling best response procedure under which agents take a random sample from the current actions played by their opponents and choose a best response against this empirical distribution.

[23] We think of this common memory set as the amount of public information available to each population. As we discussed the bounded memory assumption is motivated by the limitation of record-keeping devices: borrower's credit history is limited, insurance companies only have access to the most recent driving records that are cleared after a certain number of years; and in informal markets information is usually transmitted through word of mouth that naturally fades away.

play have been forgotten. Finally, for compatibility we assume that $\#\Xi^j = 0$ for each population, that is there are no committed agents, but that it is directly observed which actions are best responses.

Theorem 4. *High information social learning with $T = 1$ is equivalent to best response with inertia in the sense that they have the same recurrent classes and the same least resistance between any pair of such classes.*

Proof. Define z to be equivalent to z' if they have the same action distribution, and consider the equivalence classes $\{z\}$. In the best response with inertia dynamic the non-action part of the state (subtypes and common memory sets) never changes so, given the initial condition, there is a unique point in each $\{z\}$ that will ever occur. This in turn implies that, along the least resistance path from that unique point in $\{z_t\}$ to the unique point in $\{z_{t+1}\}$, the least resistance is given by taking all the actions that are not best responses to α_{t-1}^{-j} and the increase in the number of agents playing those actions by j summed for $j = 1, 2$. In high information social learning with $T = 1$ dynamic regardless of the starting point in $\{z_t\}$ the least resistance over all targets in $\{z_{t+1}\}$ is exactly the same since agents that are not playing a best response to α_{t-1}^{-j} must have trembled: content and discontent agents play the unique best response to α_{t-1}^{-j}. Hence if we have a recurrent class with respect to best response with inertia dynamics, a subset of the equivalence classes of states in that recurrent class are a recurrent class with respect to high information social learning with $T = 1$ dynamics, and the least resistance between recurrent classes is the same for both dynamics. □

We have shown the resistances between recurrent classes are the same under both dynamics, which in turn implies that the stochastically stable set and relative ergodic probability ratios of the recurrent classes are the same. To best of our knowledge there is no general characterization of convergence to mixed approximate equilibria under the best response with inertia dynamic. Still, as our pure-vs-mixed example G_2 illustrates, we believe it is more typical for the system to converge to a best response cycle than to a mixed approximate equilibrium. Intuitively, population play must correspond to the ratios induced by the mixed equilibrium, and the best response with inertia dynamic requires there be enough agents in one population to switch that the second population is no longer playing a best response, and in that case it may well be that too many agents in the first population have switched to be consistent with any mixed equilibrium. Moreover, it is impossible for the first population to go back to the original best response. The reason is that once there is a sufficiently better response that agents are willing to move they can stick with the original best response or move to the new best response, but they cannot move back from the new best response to the original best response.

8.3 Learning Dynamics with T Limited Memory

We next show global convergence to Nash equilibria when agents have possibly short memory. Young (1993) demonstrates that play converges to Nash equilibrium if the game is acyclic and

the sampling is taken over long enough time series.[24] We show convergence to Nash equilibria for generic games using a weaker version of Young's acyclicity if we allow more memory.

Recall that a block is any set $W = W^1 \times W^2$ with non-empty subsets of actions $W^j \subseteq A^j$ for $j = 1, 2$. We are interested in strict pure equilibria, and to some extent "strict" mixed equilibria given our assumptions on the grid (i.e. presence of committed agents). In this context a natural set-valued solution concept is curb blocks. A block W is *curb* ("closed under rational behavior") if $\arg\max_{a^j \in A^j} u^j(a^j, \alpha^{-j}) \subseteq W^j$ for every action profile $\alpha \in \Delta(A)$, where $\alpha^j(a^j) = 0$ for $a^j \notin W^j$, and every player j (see Basu and Weibull (1991)). That is, a set of action profiles is curb if it contains all best responses to itself. We define a $k \times l$ block W to be a block with $\#W^1 = k$ and $\#W^2 = l$. We say a game is $k \times l$ *acyclic* if for every action profile a there exists a best response path starting at a and leading to a $k \times l$ curb block W. Notice that every game is $\#A^1 \times \#A^2$ acyclic since the entire game is a curb block and that any 1×1 acyclic game is acyclic (Young (1993)). The following game is 2×2 acyclic but is not acyclic:

	H	T	U	D
H	2,0	0,2	0,0	0,0
T	0,2	2,0	0,0	0,0
U	0,0	0,0	5,5	8,2
D	0,0	0,0	9,1	2,8

A more general class is any $\#A^1 \times \#A^2$ game, where $\#A^1 = n^1 \cdot k$ and $\#A^2 = n^2 \cdot l$, with $k \times l$ blocks along the diagonal in which payoffs are strictly positive and in each block there is a unique mixed strategy equilibrium, and all other payoffs are zero. This class is similar to coordination games but with mixed equilibria on the blocks along the diagonal instead of pure strategy equilibria.

From Theorem 4 we know that play converges to Nash equilibrium for acyclic games and $T = 1$. We next show that our learning procedure converges to equilibrium if more memory is combined with our weaker notion of $k \times l$ acyclicity where best response paths need to end up in a curb block. In particular, as memory grows the requirement of $k \times l$ acyclicity is weakened, thereby encompassing a broader class of games. If we consider memory length equal to the largest curb block, we obtain convergence to equilibrium regardless of the payoff structure of the game.

Theorem 5. *If the game is $k \times l$ acyclic then, with memory $T \geq k \times l$ and $\epsilon = 0$, ν-robust states are absorbing and other states are transient.*

Proof. Starting at a ν-robust state z since all learners are playing a ν-best response, all content agents remain content with their action, so such states are absorbing. We next prove that from any non ν-robust state there is a zero resistance path to a ν-robust state.

[24] Unlike our learning procedure, for generic games play converges to a minimal curb block if agents are assumed to comply with adaptive dynamic (Young (1993)); similar results are obtained by Hurkens (1995). By virtue of assuming that agents experiment occasionally the stochastically stable states correspond to a single minimal curb block irrespective of the initial conditions, and in some games minimal curb blocks are strictly larger than the support of a Nash equilibrium.

Pick any state z_t and suppose it is not ν-robust. Then, there is zero resistance to a state z_{t+1} in which all learners of one population, say j, play the same action and are inactive, while one committed agent in population $-j$ plays the ν-best response a^{-j} to α_t^j, and all learners of population $-j$ are active and those agents that are not playing a ν-best response become discontent. From z_{t+1} there is zero resistance to a state z_{t+2} where learners of population j are inactive and hold their actions fixed, while all learners of population $-j$ play the same ν-best response a^{-j} to α_{t+1}^j in the common memory set. We proceed similarly starting at z_{t+2} and moving to z_{t+3}, we assume agents in population $-j$ hold their play fixed and are inactive, whereas one committed agents in population j plays the ν-best response a^j to α_{t+2}^{-j}, and agents of player j are all active and those not playing a ν-best response become discontent. Consider the transition to state z_{t+4} in which agents in population $-j$ play the previous action and are inactive, while learners in population j all play the same best response a^j to a^{-j} in the memory set and are inactive. The resulting state z_{t+4} is pure.

Take any pure state z_t. Since the game is finite and $k \times l$ acyclic, the best response path from this state goes to a $k \times l$ curb block W in a finite number of steps. Notice that in the following transitions when moving along best response path we only require to use the ν-best response to the play of the opposing population in the last period so it suffices to have $T = 1$. First both populations play the previous actions, a committed agent in one population, say j, play a ν-best response a^j to the population play $-j$, and all learners from population j when active become discontent. The next transition all discontent agents of population j played the ν-best response a^j which belongs to the common memory set $\mathbf{A}_T^j$, are inactive while all agents in population $-j$ play the same actions, are active and a committed agent plays a ν-best response a^{-j} to the population $-j$ play, so they become discontent. We continue until the state is such population play of learners corresponds to the $k \times l$ curb block.

Start at z_t where population play of learners lies in a $k \times l$ curb block W, and pick any $\mathbf{A}_T^j \subseteq W^j$ for each j with $T = k \times l$. If in each population j all content agents are playing a ν-best response, and for each j the common memory set $\mathbf{A}_T^j$ only contains actions that are ν-best responses to any $\alpha_t^{-j} \in \mathcal{A}^{-j}(z_t)$, then there is zero resistance to discontents choosing $a_t^i \in \mathbf{A}_T^j$, all agents being active and becoming or staying content, hence reaching a ν-robust state. Otherwise, there exists at least one agent in one of the populations that is not playing a ν-best response to $\alpha_t^{-j} \in \mathcal{A}^{-j}(z_t)$. Consider the transition where all agents play the same previous action and in one population j those agents that are not playing a ν-best response are active and become or stay discontent because they observe a ν-better response played by some committed agent which implies that $\#\mathbf{A}_T^j$ increases by 1 and that $\mathbf{A}_T^j \subseteq W^j$. If there are agents in population $-j$ that are not playing a ν-best response, we proceed to repeat the argument which results in a larger memory set $\mathbf{A}_T^{-j} \subseteq W^{-j}$. Eventually, after $k \times l$ steps we have not lost any relevant memory since $T = k \times l$ so all learners are discontent and we have expanded each memory set $\mathbf{A}_T^j$ to include all actions in the $k \times l$ curb block W, which contains a ν-Nash equilibrium by definition. There is zero resistance to having all discontents playing the action profile corresponding to such equilibrium, all agents being active and becoming

content; reaching the corresponding ν-robust state. $\square$

We showed that, unlike best response with inertia, high information social learning does converge globally to Nash equilibrium for generic two player games, not only acyclic, if memory is long enough $T = \#A^1 \times \#A^2$. We will show in Section 9.2 that no agent could improve his expected time average payoff by more than ν by using a different learning procedure than high information social learning.

As we have seen, only pure ν-robust states have radii that increase linearly with population size N. In the following result, we show that it is possible that radii of mixed ν-robust states increase with N under high information social dynamic and the support of those ν-robust states belongs to a curb block that does not include all equilibria.

Lemma 14. *If a W curb block does not contain all Nash equilibria and $T \geq 1$ then there exists a constant $\kappa > 0$ such that the radius of the ν-robust states, where the support of $\alpha(z) \in \mathcal{A}(z)$ is entirely contained within the W curb block, is at least κN.*

8.4 Revisiting the Examples

We now examine the high information social learning dynamics in our two examples. We first observe that for $\#\Xi^j = 0$ for $j = 1, 2$ and direct observation of best responses the computation of the radius and co-radius with high information social learning models is exactly the same as for best response with inertia.

With our high information social learning dynamics, as well as with best response with inertia, in the mixed-vs-pure game G_2 the HT block is stochastically stable by radius co-radius argument. Since the HT block contains the unique Nash equilibrium $((\frac{3}{5}H, \frac{2}{5}T), (\frac{2}{5}H, \frac{3}{5}T))$, the stochastically stable set is a subset of ν-robust states in a neighborhood of $((\frac{3}{5}H, \frac{2}{5}T), (\frac{2}{5}H, \frac{3}{5}T))$. This is in contrast to our low information social learning model, where P is stochastically stable. Importantly, the behavior in the HT block is starkly different since it does not exhibit deterministic best response cycles, as in best response with inertia dynamics, and; indeed, it converges to equilibrium. This serves to illustrate that the relative stability of pure and mixed equilibria may depend on whether agents are able to observe which actions corresponds to best responses, and offers new testable implications for future work.

In the coordination game G_1, let the BCD block be $\{B, C, D\} \times \{B, C, D\}$. Under the best response with inertia and high information social learning dynamics, the stochastically stable set is contained in the BCD block, again in contrast to our low memory social learning model. We observe that the block game associated with the BCD block has seven Nash equilibria: three pure strategy equilibria which we label B, C and D, three mixed equilibria $((\frac{10}{11}C, \frac{1}{11}D), (\frac{10}{11}C, \frac{1}{11}D))$, $((\frac{10}{11}B, \frac{1}{11}D), (\frac{10}{11}B, \frac{1}{11}D))$ and $((\frac{10}{11}B, \frac{1}{11}C), (\frac{10}{11}B, \frac{1}{11}C))$; and one mixed equilibrium in which players randomize uniformly across B, C and D. Notice that, more precisely, when analyzing ν-robust states there is a subset of ν-robust states in a neighborhood of each mixed equilibrium.

To determine what is stochastically stable inside the BCD block requires a bit more work. Note that in this example all equilibria of the best response with inertia dynamic do not belong

to the same circuit. This forces us to analyze circuits of circuits. To do so, we need to define the incremental resistance of going from one circuit to the next circuit (the *modified resistance* in Levine and Modica (2016)), which is the least resistance from one circuit to the next circuit minus the least resistance path out of the circuit.[25] For the least resistance path out of the 6-point circuit, we have computed that the radii of B, C and D are about $N/11$, and the radii of any mixed equilibria across two actions are all 1, hence the least resistance path out of the circuit is about $N/11$. The least resistance from the 6-point circuit to any neighborhood of the completely mixed equilibrium on the BCD block is the same regardless of the starting mixed equilibria in the 6-point circuit. In all cases this is about $N/2$ since $1/2$ of one population may play the remaining action to make it a ν-best response and appear in the memory set. Hence the incremental resistance from the 6-point circuit to the other circuits is about $N/2 - N/11$. Analogously, the least resistance path out of the other circuit is $N/3$, and the least resistance from those circuits to the 6-point circuit is about $N/3$ since $1/3$ of one population may play shift from one of the actions to some other.

We define the *modified radius* to be the resistance of moving away from an absorbing state, which is equal to the sum of the incremental resistance and the modified resistance within the circuit. In this case, the modified resistance within the circuit of the 6-point circuit is $N/11$ as all pure strategy equilibria have the same resistance to travel far from them and in the case of the other circuits is $N/3$. Then the modified radius of the 6-point circuit is about $N/2$ and about $N/3$ for the completely mixed equilibrium circuit. By Theorem 10 in Levine and Modica (2016) the 6-point circuit contains the stochastically stable states, which are the three pure strategy equilibria B, C and D, and each equilibrium is equally likely in the long run. These are also the stochastically stable states under best response with inertia dynamics.[26] Moreover, if we consider the coordination game G_1 with perturbed payoff where a player obtains $\kappa > 0$ instead of 0 when choosing B against C, the unique stable state is the pure equilibrium B as it has the largest radius in the BCD block.

9 Discussion and Extensions

9.1 Noisy Information

In the analysis so far, there is a fixed and small (relatively to N) number of committed agents, and agents play all their opponents in round-robin tournaments so there is no sampling error in agents' observations about whether they are playing a ν-best response. In practice, however, there could be noise about what agents observe either because of sampling or because utility is a random function of the actions that are played in matches. Our results are not robust if this noise has probability independent of ϵ because the noise would be the only driving force. On the other hand, in a noisy environment it seems natural to allow agents average over matches within a period to push the noise down.

[25] The least outgoing resistance is equivalent to the radius in this case.

[26] Basically this is the same calculation as the radius modified co-radius, but the circuits require less computations since they specify what sequence of absorbing states must take place to transition from one place to another; rather than computing all possible ways.

To allow for the possibility that within a period there is some noise we assume that agents observe a noisy signal of their payoff in each match. In every period t, instead of playing round robin, agents play K random opponents, sampling with replacement for simplicity.[27] We now assume that there is no trembling, but in any period an agent may draw a sample that is not representative with some probability, that we denote by ϵ. We then interpret ϵ as arising from sampling error rather than trembling, and note that ϵ goes to zero as K grows large, resulting in the dynamics we are studying. There are two differences between this model and the one we analyzed above. First, a discontent agent that is not playing a ν-best response can become content with positive probability. However, as this is probability will be bounded away from 1 the no cost to staying discontent principle still applies, because one cannot lower the resistance of the path constructed in the proof by having one agent accidentally become content, as the path may require that the agent tremble to play a specific action later on. Consequently all the resistance computations are the same as in the original model. Second, in the existing model it is possible for an agent to tremble onto a dominated strategy, and this is not possible with trembles arising from beliefs. However as we have made the generic assumption that there are no weakly dominated strategies (Assumption 4), only strictly dominated strategies, this does not matter either.

9.2 Performance of the Learning Rules

While the learning rules we describe seem intuitive given the information we assume is available, we might expect learning rules not to be used if they perform too poorly. One property that we would like a learning rule to satisfy is that in a reasonably broad class of environments it "learns" in the sense of getting it right asymptotically. We conclude by showing that this is the case for the learning rules we study in environments in which there is global convergence to approximate Nash equilibrium.

Specifically no agent could improve his expected time average payoff by more than ν by using a different learning procedure than ours, where we allow the alternative learning procedures to use any amount of information, including knowing in advance what the agents of the other player were going to do. Note that this is not a "universal consistency property,"[28] since it depends on the fact that the other agents are also using our learning procedure. Formally, in a state z agent i's learning rule gives expected utility $U^i(z)$ that depends only on z. Given the state z there is a unique probability distribution $\pi^{-j}(z)[\alpha^{-j}]$ over $\alpha^{-j} \in \Delta^N(A^{-j})$. Suppose that action distributions α^{-j} of the opposing population are drawn from $\pi^{-j}(z)$, that the agent i observes the outcome α^{-j} and chooses a best response to it. Let $V^i(z)$ be the corresponding expected utility with respect to $\pi^{-j}(z)$.[29] Taking expectations with respect to P_ϵ, we compute

[27] What is important is that K be large in an absolute sense, independent of the size of N, since it is the sample size K that determines the standard error.

[28] See Fudenberg and Levine (1995).

[29] No learning rule using any information can do better than this.

$$\limsup_{\epsilon \to 0} \limsup_{\tau \to \infty} \frac{1}{\tau} E \sum_{t=1}^{\tau} (V^i(z_t) - U^i(z_t)) \leq \nu.$$

The reason for this is simply that z_t most of the time is at a ν-Nash equilibrium, and so $U^i(z_t)$ cannot do more than ν-worse than any strategy regardless of how it is learned.

Putting differently, if the agent knew that agents of the other population were going to follow a stationary strategy for very long periods of time τ (where τ depends on ϵ) and that committed agents in his own population were going to reveal what the agents of the other population are doing, the agent could not do much better than our learning procedure despite its limited memory and information.

10 Conclusion

In many settings people have aggregate information about the payoffs and/or behaviors of others, and may use this information to help select their strategies. Most people also have bounded memory. We have considered two learning models that incorporate these ideas, and showed that behavior comes close to approximate Nash equilibria, and related the amount of social information and memory used to which equilibria we should expect to see in the long run.

We considered a low information social learning model in which agents observe aggregate information about how well others are doing, but not how they obtain those payoffs, so agents are not able to directly imitate successful actions. Here we assume that agents use their limited memory to keep track of their own actions that recently did well and a "search state" that indicates that there might be better actions to experiment with. In principle agents might do better by using more memory, for instance, building a picture of the payoff matrix by remembering past play. Nonetheless this is likely to be cognitively and computationally costly, and it will work well only if the environment is stationary. We demonstrated that pure strategy equilibria should be expected to be seen a larger fraction of the time than mixed strategy equilibria when people cannot easily see what actions did well. By way of examples, we compared the predictions of our learning model to those of the best response with inertia dynamic.

Our high information social learning model supposes that people observe aggregate information about how well and what others did, which might describe some sorts of consumption and financial decisions, and that when people experiment they use actions that performed well recently. When people recall only the last action and approximate best responses, we found that our learning dynamic predicts the same stochastically stable states as best response with inertia, and so can be trapped in cycles in the long run. When agents have more memory, cycles become improbable, and mixed strategy equilibria can be relatively more stable than pure strategy equilibria.

Which of these models is a better description for how people learn to play Nash equilibria will of course depend on the information available to the agents and to the cognitive effort they put into processing it. Neither one should be expected to apply literally to a wide spectrum of situations,

but we hope they will provide a useful complement to the widely-used best response dynamic in making predictions about long run social outcomes. We believe that it would be interesting to explore our learning models in controlled laboratory experiments because our results establish sharp predictions depending on observability and memory.

References

Agarwal, S., Driscoll, J. C., Gabaix, X., and Laibson, D. (2008). Learning in the credit card market. Technical report, National Bureau of Economic Research.

Basu, K. and Weibull, J. W. (1991). Strategy subsets closed under rational behavior. *Economics Letters*, 36(2):141–146.

Benaïm, M. and Hirsch, M. W. (1999). Mixed equilibria and dynamical systems arising from fictitious play in perturbed games. *Games and Economic Behavior*, 29(1):36–72.

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

Björnerstedt, J. and Weibull, J. (1996). Nash equilibrium and evolution by imitation. In K. Arrow, E. Colombatto, M. P. and Schmidt, C., editors, *The Rational Foundation of Economic Behavior*, pages 155–71. London: MacMillan.

Bott, R. and Mayberry, J. (1954). *Matrices and trees*. John Wiley and Sons, Inc., New York.

Dal Bó, P. and Fréchette, G. R. (2016). On the determinants of cooperation in infinitely repeated games: A survey. *Journal of Economic Literature*.

Ellison, G. (2000). Basins of attraction, long-run stochastic stability, and the speed of step-by-step evolution. *Review of Economic Studies*, 67(1):17–45.

Ellison, G., Fudenberg, D., and Imhof, L. A. (2009). Random matching in adaptive dynamics. *Games and Economic Behavior*, 66(1):98–114.

Erev, I. and Haruvy, E. (2013). Learning and the economics of small decisions. The Handbook of Experimental Economics, Vol. 2. Princeton University Press. Forthcoming.

Feenberg, D. R., Ganguli, I., Gaule, P., and Gruber, J. (2015). It's good to be first: Order bias in reading and citing NBER working papers. Technical report, National Bureau of Economic Research.

Foster, D. P. and Hart, S. (2015). Smooth calibration, leaky forecasts, finite recall, and nash dynamics. Working Paper.

Foster, D. P. and Young, H. P. (2003). Learning, hypothesis testing, and nash equilibrium. *Games and Economic Behavior*, 45(1):73–96.

Foster, D. P. and Young, H. P. (2006). Regret testing: learning to play nash equilibrium without knowing you have an opponent. *Theoretical Economics*, 1(3):341–367.

Freidlin, M. and Wentzell, A. (1984). *Random Perturbations of Dynamical Systems*. Springer.

Fudenberg, D. and Imhof, L. A. (2006). Imitation processes with small mutations. *Journal of Economic Theory*, 131(1):251–262.

Fudenberg, D. and Kreps, D. M. (1993). Learning mixed equilibria. *Games and Economic Behavior*, 5(3):320–367.

Fudenberg, D. and Levine, D. K. (1993). Self-confirming equilibrium. *Econometrica*, 61(3):523–545.

Fudenberg, D. and Levine, D. K. (1995). Consistency and cautious fictitious play. *Journal of Economic Dynamics and Control*, 19(5):1065–1089.

Fudenberg, D. and Levine, D. K. (1998). *The theory of learning in games*. MIT Press.

Fudenberg, D. and Levine, D. K. (2014). Recency, consistent learning, and nash equilibrium. *Proceedings of the National Academy of Sciences*, 111(3):10826–10829.

Fudenberg, D. and Peysakhovich, A. (2014). Recency, records and recaps: Learning and non-equilibrium behavior in a simple decision problem. *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, (16):971–986.

Hart, S. and Mas-Colell, A. (2006). Stochastic uncoupled dynamics and nash equilibrium. *Games and Economic Behavior*, 57(2):286–303.

Hofbauer, J. and Sandholm, W. H. (2002). On the global convergence of stochastic fictitious play. *Econometrica*, 70(6):2265–2294.

Hurkens, S. (1995). Learning by forgetful players. *Games and Economic Behavior*, 11(2):304–329.

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

Levine, D. K. and Modica, S. (2013). Conflict, evolution, hegemony, and the power of the state. Working paper.

Levine, D. K. and Modica, S. (2016). Dynamics in stochastic evolutionary models. *Theoretical Economics*, 11(1):89–131.

Malmendier, U. and Nagel, S. (2011). Depression babies: Do macroeconomic experiences affect risk taking? *Quarterly Journal of Economics*, 126(1):373–416.

Myerson, R. and Weibull, J. (2015). Tenable strategy blocks and settled equilibria. *Econometrica*, 83(3):943–976.

Nöldeke, G. and Samuelson, L. (1993). An evolutionary analysis of backward and forward induction. *Games and Economic Behavior*, 5(3):425–454.

Oyama, D., Sandholm, W. H., and Tercieux, O. (2015). Sampling best response dynamics and deterministic equilibrium selection. *Theoretical Economics*, 10(1):243–281.

Pradelski, B. S. R. and Young, H. P. (2012). Learning efficient nash equilibria in distributed systems. *Games and Economic Behavior*, 75(2):882–897.

Samuelson, L. (1994). Stochastic stability in games with alternative best replies. *Journal of Economic Theory*, 64(1):35–65.

Schelling, T. C. (1960). *The strategy of conflict*. Harvard University Press.

Van Huyck, J. B., Battalio, R. C., and Beil, R. O. (1990). Tacit coordination games, strategic uncertainty, and coordination failure. *American Economic Review*, 80(1):234–248.

Young, H. P. (1993). The evolution of conventions. *Econometrica*, 61(1):57–84.

Young, H. P. (2009). Learning by trial and error. *Games and Economic Behavior*, 65(2):626–643.

Appendix

A Proofs

Proof of Lemma 1. The determination of $P_\epsilon(x_{t+1}|x_t)$ has several steps involving interim variables. First, let $\mathcal{T}^j$ denote the learners of player j that tremble and let $\mathcal{N}^j$ be the non-trembling learners. The probability of exactly this set of tremblers and non-tremblers is $\epsilon^{\#\mathcal{T}^j}(1-\epsilon)^{\#\mathcal{N}^j}$. Second, choose any assignment of actions to all agents for each j denoted by $\sigma^j \in N^{A^j}$. Such an action profile has probability defined as $\Gamma^j(x_t, \mathcal{T}^j)[\sigma^j]$ that is calculated below. Given the action assignment σ^j and the corresponding frequencies α_t, we compute the aggregate statistic $\phi^j(\alpha_t)$ for each j. For the non-tremblers $i \in \mathcal{N}^j$ and each subset $\mathcal{R}^j \subseteq \mathcal{N}^j$ of these non-trembling learners who are active conditional on $\mathcal{T}^j$, there is probability $p^{\#\mathcal{R}^j}(1-p)^{\#\mathcal{N}^j - \#\mathcal{R}^j}$ that exactly this subset of agents is active and updates its type according to this period's highest payoff. In summary, we have the interim variables $\mathcal{T}^j, \sigma^j$ and $\mathcal{R}^j$.

According to the individual learning rule defined above, if $i \notin \mathcal{R}^j$ then $\theta^i_{t+1} = \theta^i_t$. If $i \in \mathcal{R}^j$ and $u^j(a^i_t, \alpha^{-j}_t) > \overline{u}^j(\phi^j(\alpha_t)) - \nu$ then $\theta^i_{t+1} = a^i_t$, otherwise $\theta^i_{t+1} = 0$. We also compute feasible action profiles conditional on $\mathcal{T}^j$. Define $D^j(x_t)$ to be the number of discontent types in x_t. Let $\mathcal{T}^j_C(x_t)$ be the subset of $\mathcal{T}^j$ and let $\mathcal{N}^j_C(x_t)$ be the subset of $\mathcal{N}^j$ corresponding to content agents in x_t. Let $\overline{\alpha}^j(x_t, \mathcal{T}^j) \in \Delta^{\#\Xi^j + \#\mathcal{N}^j_C(x_t)}(A^j)$ be the action profile corresponding to the aggregate play of the committed and non-trembling content types in x_t where the non-trembling content agents play the action corresponding to their type and the committed types

play their committed action. An action profile $\alpha^j \in \Delta^N(A^j)$ in agent state x_t is *feasible with respect to* $\mathcal{T}^j$ if $N\alpha^j = (\#\Xi^j + \#\mathcal{N}_C^j(x_t))\overline{\alpha}^j(x_t, \mathcal{T}^j) + (D^j(x_t) + \#\mathcal{T}_C^j(x_t))\tilde{\alpha}^j$ for some action profile $\tilde{\alpha}^j \in \Delta^{D^j(x_t)+\#\mathcal{T}_C^j(x_t)}(A^j)$, that is, if it is consistent with the play of the non-trembling content and committed types. In particular, let $\overline{\alpha}^j(z_t) \equiv \overline{\alpha}^j(x_t, \emptyset)$ be the action profile corresponding to the aggregate play of contents and committed agents in state z_t which is well-defined since $\overline{\alpha}^j(x_t, \emptyset)$ is independent of $x_t \in X(z_t)$, and define $\mathcal{A}^j(z_t)$ to be the set of all corresponding feasible α^j. Finally, let $\mathcal{T} = (\mathcal{T}^1, \mathcal{T}^2)$, $\mathcal{R} = (\mathcal{R}^1, \mathcal{R}^2)$ and $\sigma = (\sigma^1, \sigma^2)$.

We compute the joint conditional probability $P(\mathcal{T}, \sigma, \mathcal{R}, x_{t+1}|x_t)$ of the terminal agent state x_{t+1} and the interim variables $\mathcal{T}, \sigma, \mathcal{R}$ considering two sets of events. In the first case, if σ^j is not feasible given $\mathcal{T}^j$ and x_t, or if $x_{t+1} \notin X(z_{t+1})$ this probability is zero. Observe that the non-trembling content agents are playing the action with which they are content and all other learners are playing uniformly; this implies that $\Gamma^j(x_t, \mathcal{T}^j)[\sigma^j] = (1/\#A^j)^{D^j(x_t)+\#\mathcal{T}_C^j(x_t)}$. Then for the other case, the probability is given by

$$P(\mathcal{T}, \sigma, \mathcal{R}, x_{t+1}|x_t) = \prod_{j=1,2} \epsilon^{\#\mathcal{T}^j}(1-\epsilon)^{\#\mathcal{N}^j} \left(\frac{1}{\#A^j}\right)^{D^j(x_t)+\#\mathcal{T}_C^j(x_t)} p^{\#\mathcal{R}^j}(1-p)^{\#\mathcal{N}^j - \#\mathcal{R}^j}.$$

Now we can compute $P_\epsilon(x_{t+1}|x_t) = \sum_{\mathcal{T},\sigma,\mathcal{R}} P(\mathcal{T}, \sigma, \mathcal{R}, x_{t+1}|x_t)$. □

Proof of Lemma 3. The hypothesis $\nu < g$ implies that ν-best responses are strict 0-best responses,[30] and for each pure opponent's action a^{-j} for which some a^j is the (unique) strict best response, there is a $\gamma \geq 0$ such that a^j is also a best response to any mixed strategy $\alpha^{-j} \in \Delta(A^{-j})$ such that $\alpha^{-j}(a^{-j}) \geq 1 - \gamma$. By finiteness of A^{-j} there is a $\overline{\gamma}$ such that for all $\gamma \in (0, \overline{\gamma})$ the previous conclusion holds for all such best responses a^j, which proves part (1).

For part (2), since the game is finite it has a mixed strategy Nash equilibrium, and for any $\nu > 0$ and any such Nash equilibrium $\hat{\alpha} \in \Delta(A)$, there is an open neighborhood $\mathscr{U}$ of $\hat{\alpha}$ in which every element is a $\nu/2$ equilibrium. For N sufficiently large there is a grid point $\alpha \in \Delta^N(A)$ in $\mathscr{U}$, and consequently for large enough N/M if the learners are content with this grid point it is ν-robust. Because $M \geq 1$, the hypothesis that N/M is large implies that N is large as well, so we may choose N/M large enough that both of these hypotheses are true. Finally, if the game has a pure strategy equilibrium it is strict from Assumption 1 so if the learners play these actions the corresponding state is 0-robust for N/M sufficiently large. □

Proof of Lemma 4. In light of Lemma 3 part (1) we see that $u^j(\hat{a}^j, a^{-j}) > u^j(a^j, a^{-j}) - \nu$ for $a^j \neq \hat{a}^j$ and all $a^{-j} \in A^{-j}$, and so $u^j(\hat{a}^j, \alpha^{-j}) > u^j(a^j, \alpha^{-j}) - \nu$ for $a^j \neq \hat{a}^j$ and all α^{-j}. It follows that in any ν-robust state all of the population j learners must be content with $\hat{a}^j$. This means that at the ν-robust state $\alpha^j(\hat{a}^j) > 1 - M/N$, so by Lemma 3 part (1) when $N/M \geq \nu$ there is only one possible ν-best response to this for player $-j$ and so player $-j$ learners must be in that state which is therefore ν-robust. □

[30] Note that this is true even for $\nu = 0$.

Lemma 15. *When $\epsilon > 0$ the Markov process is irreducible and aperiodic.*

Proof. Pick any state $\hat{z}$ where $D^j(\hat{z}) = N - \#\Xi^j$ for each population j. Start with any state $z_t \in Z$ and take any agent state $x_t \in X(z_t)$. There is probability $\epsilon^{\#\mathcal{T}^j}$ that all learners tremble, and $\#\mathcal{T}_t^j = N - \#\Xi^j$, so $D^j(z_{t+1}) = N - \#\Xi^j$ for $j = 1, 2$. Take $\alpha_{t+1}^j \in \mathcal{A}^j(\hat{z})$ and choose $\hat{x}_{t+1} \in X(\hat{z})$ with an action assignment $\hat{\sigma}^j$ consistent with α_{t+1}^j. Starting at $\hat{x}_t$there is probability $(1/\#A^j)^{2N-\#\Xi^1-\#\Xi^2}$ that all agents play $\hat{\sigma}^j$. There is probability $(1-p)^{2N-\#\Xi^1-\#\Xi^2}$ that all agents are inactive so they all stay discontent, hence entering $\hat{z}$.

Next we observe that once at $\hat{z}$ there is positive probability of staying there for any finite length of time. That is, starting at an agent state $\hat{x} \in X(\hat{z})$ consistent with $\hat{z}$ there is positive probability that no agent trembles and is active so that learners will all remain with their contentment and action. Since starting at any state there is a positive probability of reaching a single state $\hat{z}$ where the system may rest for any length of time with positive probability implies that the system is irreducible and aperiodic. □

Proof of Lemma 6. Let $x_t \in X(z)$ and $z_t = z$. Since $z \succeq \hat{z}$ and $\hat{z}$ is ν-robust we have for each j that $N\overline{\alpha}^j(\hat{z}) = (N - D^j(z))\overline{\alpha}^j(z) + D^j(z)\tilde{\alpha}^j$ for some $\tilde{\alpha}^j \in \Delta^{D^j(z)}(A^j)$. This implies that $\mathcal{A}^j(z) \supseteq \mathcal{A}^j(\hat{z})$. Hence if $\alpha_t^j \in \mathcal{A}^j(\hat{z})$ then $\alpha_t^j \in \mathcal{A}^j(z)$ and the former implies that in α_t^j all learners are playing ν-best responses. There is zero resistance to no trembles and $\mathcal{R}$ including all learners so all become content with a_t^i. The resulting agent state x_{t+1} therefore satisfies $x_{t+1} \in X(\hat{z})$ and by construction the resistance of this transition is 0. □

Proof of Lemma 8. Suppose $z_t = z$ is totally discontent and $\hat{z}$ is ν-robust. Take $x_t \in X(z)$ and action profile σ_t in which agents play some actions that are feasible in $\hat{z}$ which has no resistance as $\mathcal{A}^j(\hat{z}) \subseteq \mathcal{A}^j(z)$ for $j = 1, 2$. There is then zero resistance to all the learners becoming content, as they are active and all are playing a ν-best response since $\hat{z}$ is ν-robust, and none trembling. The resulting state $x_{t+1} \in X(\hat{z})$; hence reaching $\hat{z}$ with zero resistance and showing part (1).

Now consider a proto ν-robust state $z_t = z$ that is not totally discontent with $w(z) > 0$. Let population j have at least one content agent in state a^j so $w^j(z) \geq 1$. If there is a content agent in population $-j$, that is $w^{-j}(z) = 1$, suppose she is playing a ν-best response a^{-j}. If there is not, $w^{-j}(z) = 0$ and there is at least one content in population j playing an action $\hat{a}^j$ that is a ν-best response to any $\alpha^{-j} \in \Delta^N(A^{-j})$ such that $N\alpha^{-j} = \#\Xi^{-j}\overline{\alpha}^{-j} + (N - \#\Xi^{-j})\tilde{\alpha}^{-j}$ for some $\tilde{\alpha}^{-j} \in \Delta^{N-\#\Xi^{-j}}(A^{-j})$ as in population $-j$ all learners are discontent. In particular $\hat{a}^j$ is a ν-best response to any pure action a^{-j} which implies $\hat{a}^j$ is weakly ν-dominant. By Lemma 3 part (1) all content agents in j must be playing $a^j = \hat{a}^j$, the unique dominant action hence $w^j(z) = 1$. In this case, by Lemma 4, let a^{-j} be the unique ν-best response to that action. Take any $x_t \in X(z)$, and a feasible action profile σ_t so that all learners play the actions a^j, a^{-j}. Next suppose that all discontent agents are active, become content and nobody trembles; thereby the target state $x_{t+1} \in X(\hat{z})$. This transition has no resistance. By the proto ν-robust state assumption the state $\hat{z}$ is in fact a ν-robust state and $w(\hat{z}) = 2$. By construction unless z was semi-discontent we did not increase the width which is claimed in part (2).

Finally, to show part (3) suppose that $z_t = z$ is not proto ν-robust with $w(z) > 0$. Then in at least one population j there is at least one content agent in a state a^j that is not a ν-best response to some feasible $\alpha^{-j} \in \mathcal{A}^{-j}(z)$. Pick any $x_t \in X(z)$. There is zero resistance to having population $-j$ play α^{-j}, one committed agent of player j plays a ν-better response than a^j and some feasible $\alpha^j \in \mathcal{A}^j(z)$. Moreover, there is zero resistance when all agents of player j in state a^j are active hence become discontent while not trembling; the rest of the agents do not tremble as well. Then $x_{t+1} \in X(z_{t+1})$ with $w(z_{t+1}) < w(z)$. □

Lemma 16. *Every state that does not correspond to a pure strategy Nash equilibrium is transient under best response with inertia dynamic.*

Proof. Fix a time t and suppose that the state does not correspond to a pure strategy equilibrium. There is positive probability that this period all agents of one player, say j, do not adjust their play while all agents of the other player $-j$ play the best response to the date-t state, and that at date $t+1$ all agents of j play the best response to the date $t+1$ state while all agents of player $-j$ hold their actions fixed. Thus there is positive probability that play in each population corresponds to a pure strategy from period $t+2$ on. Because the game is finite and acyclic, the best response path from this state converges to a pure strategy Nash equilibrium in a number of steps no greater than $J =: \#A^1 \times \#A^2$. There is positive probability that the populations will take turns adjusting, all of the $-j$ agents adjusting in periods $t, t+2, t+4, \ldots$, and all of the j agents adjusting at $t+1, t+3, t+5, \ldots$, so this equilibrium has probability bounded away from 0 of being reached in $2+J$ steps, showing the initial time t state is transient. □

Proof of Lemma 11. Let $a \in A$ be any pure strategy Nash equilibrium, and notice that $\underline{\rho}_a^j(\nu)$ and $\rho_a^j(\nu)$ are continuous at $\nu = 0$ by strictness of the equilibrium. By Assumption 7 and by continuity of $\underline{\rho}_a^j$ and ρ_a^j it follows that $\rho_a^j(\nu) < \underline{\rho}_a^1(\nu) + \underline{\rho}_a^2(\nu)$ for one player j. By Assumption 1, for each pure strategy Nash equilibrium and each player j, $\underline{\rho}_a^j(0) > 0$, so that for small enough ν, $\underline{\rho}_a^j(\nu) > 0$. Since there are finitely many pure Nash equilibria, we may choose $\overline{\nu}$ these conditions are satisfied at all pure Nash equilibria a for all $\nu \leq \overline{\nu}$. Take any $\nu \leq \overline{\nu}$. Define $\lambda_a^j \equiv \underline{\rho}_a^1(\nu) + \underline{\rho}_a^2(\nu) - \rho_a^j(\nu) > 0$ for any such equilibrium a. Pick M/N such that $(N-M)\lambda_a^j \geq 4$, then

$$\begin{aligned} \lceil (N-M)\rho_a^j(\nu) \rceil \leq 1 + (N-M)\rho_a^j(\nu) &< (N-M)\underline{\rho}_a^1(\nu) + (N-M)\underline{\rho}_a^2(\nu) - 3 \\ &\leq \lfloor (N-M)\underline{\rho}_a^1(\nu) \rfloor + \lfloor (N-M)\underline{\rho}_a^2(\nu) \rfloor - 1. \end{aligned}$$

Next, for each player j choose N/M so that $(N-M)\underline{\rho}_a^j(\nu) \geq 2$ and notice that $\lfloor (N-M)\underline{\rho}_a^j(\nu) \rfloor \geq (N-M)\underline{\rho}_a^j(\nu) - 1 \geq 1$. Then since there are finitely many pure equilibria we pick the largest $\overline{N/M}$. For any pure ν-robust state z notice that $\overline{r}_z^j = \lceil (N-M)\rho_a^j(\nu) \rceil$ and $\underline{r}_z^j = \lfloor (N-M)\underline{\rho}_a^j(\nu) \rfloor$. Then for all $\nu \leq \overline{\nu}$ and $N/M \geq \overline{N/M}$ for one player j we have that $\overline{r}_z^j \leq \underline{r}_z^1 + \underline{r}_z^2$ and for both players j that $\underline{r}_z^j \geq 1$ in each pure ν-robust state. □

Proof of Lemma 14. Let z be any ν-robust state with the support of $\alpha \in \mathcal{A}(z)$ being a W curb block. Let $\hat{z}$ be any ν-robust state such that the support of $\hat{\alpha} \in \mathcal{A}(\hat{z})$ intersects $A \setminus W$. Define κ_z^j to

be the least fraction of learners from population $-j$ that play $a^{-j} \in A^{-j} \setminus W^{-j}$ such that any ν-best response played by the agents from population j lies in $A^j \setminus W^j$. Let $\kappa_z = \min\{\kappa_z^1, \kappa_z^2\}$. Any z' such that for either population $D^j(z') < \kappa N$ belongs to the basin of z since the system returns to z with probability 1. This is because $\mathbf{A}_T^j \subseteq W^j$ for both j which in turn implies that discontent agents choose a ν-best response, and when active become content, and $\text{supp}(\alpha) = W$. If $D^j(z') \geq \kappa N$ for at least one population j, then committed agents in population $-j$ may reveal a ν-better response $\hat{a}^{-j}$ in the support of $\hat{\alpha}^{-j}$ so that $\mathbf{A}_T^{-j}$ is not contained in W^{-j} and all agents in population $-j$ that are active become discontent. Next all discontent agents in population $-j$ play $\hat{a}^{-j} \notin W^{-j}$ with positive probability and a committed agent in population j may play a ν-better response $\hat{a}^j$ in the support of $\hat{\alpha}^j$ so that all agents with positive probability are active. Then, discontent agents in population j play $\hat{a}^j$ in $\mathbf{A}_T^j$ with positive probability, reaching the state $\hat{z}$. $\square$