Pathological Outcomes of Observational Learning[*]

Lones Smith[†]
Department of Economics
University of Michigan

Peter Sørensen[‡]
Institute of Economics
University of Copenhagen

revised for *Econometrica*

October 20, 1999

(original version: July, 1994; first submitted: June, 1996)

Abstract

This paper explores how Bayes-rational individuals learn sequentially from the discrete actions of others. Unlike earlier informational herding papers, we admit heterogeneous preferences. Not only may type-specific 'herds' eventually arise, but a new robust possibility emerges: *confounded learning.* Beliefs may converge to a limit point where history offers no decisive lessons for anyone, and each type's actions forever nontrivially split between two actions.

To verify that our identified limit outcomes do arise, we exploit the Markov-martingale character of beliefs. Learning dynamics are stochastically stable near a fixed point in many Bayesian learning models like this one.

[*]This second revision is based on suggestions from three referees and the Editor. We have also benefited from seminars at MIT, Yale, IDEI Toulouse, CEPREMAP, IEA Barcelona, Pompeu Fabra, Copenhagen, IIES Stockholm, Stockholm School of Economics, Nuffield College, LSE, CentER, Free University of Brussels, Western Ontario, Brown, Pittsburgh, Berkeley, UC San Diego, UCLA, UCL, Edinburgh, Cambridge, Essex and IGIER. We wish also to specifically acknowledge Abhijit Banerjee, Kong-Pin Chen, Glenn Ellison, John Geanakoplos, David Hirshleifer, David Levine, Preston McAfee, Gerhard Orosel, Thomas Piketty, Ben Polak, and Jörgen Weibull for useful comments, and Chris Avery for discussions at a very early stage in this project. All errors remain our responsibility. Smith and Sørensen respectively acknowledge financial support for this work from the National Science Foundation and the Danish Social Sciences Research Council.

[†]e-mail address: `econ-theorist@earthling.net`

[‡]e-mail address: `peter.sorensen@econ.ku.dk`

1. INTRODUCTION

Suppose that a countable number of individuals each must make a once-in-a-lifetime binary decision — encumbered solely by uncertainty about the state of the world. If preferences are identical, there are no congestion effects or network externalities, and information is complete and symmetric, then all ideally wish to make the same decision.

But life is more complicated than that. Assume instead that the individuals must decide sequentially, all in some preordained order. Suppose that each may condition his decision both on his endowed private signal about the state of the world and on all his predecessors' decisions, but not their hidden private signals. The above simple framework was independently introduced in Banerjee (1992) and Bikhchandani, Hirshleifer, and Welch (1992) (hereafter, simply BHW). Their perhaps unexpected common conclusion was that with positive probability an 'incorrect herd' would arise: Despite the surfeit of available information, after some point, everyone might just settle on the identical less profitable decision. This result has really sparked a welcome renaissance in the field of informational economics, as various twists on the herding phenomenon have been woven into a host of applications from finance to organizational theory, and even lately into experimental work.

In this paper, we study more generally how Bayes-rational individuals sequentially learn from the actions of others. This leads us to a greater understanding of herding, and why and when it occurs. Crucially, it leads also to the discovery of a co-equal robust rival phenomenon to herding that has so far been missed, and that is economically important.

To motivate our point of departure from Banerjee and BHW, consider the following counterfactual. Assume that we are in a potential herd in which one million consecutive individuals have acted alike, but suppose that the next individual deviates. What then could Mr. one million and two conclude? First, he could decide that his predecessor had a more powerful signal than everyone else. To capture this, we generalize the private information beyond discrete signals, and admit the possibility that there is no uniformly most powerful yet nonrevealing signal. Second, he might opine that the action was irrational or an accident. We thus add noise to the herding model. Third, he possibly might decide that different preferences provoked the contrary choice. On this score, we consider the model with multiple types. Here, we find that herding is not the only possible 'pathological' outcome: We may well converge to an informational pooling equilibrium where history offers no decisive lessons for anyone, and everyone must forever rely on his private signal!

The paper is unified by two natural questions: (1) What are the robust long-run outcomes of observational learning? (2) Do we in fact settle on any one? Our inquiry is focused through the two analytic lenses of convergence of beliefs (learning) and convergence of actions (in frequency or, more strongly, with herds). BHW introduced the colorful terminology of a *cascade* for an infinite train of individuals acting irrespective of the content of their signals. With a single rational type and no noise (henceforth, the *herding model*), individuals always eventually settle on an action (a herd). Yet the label 'cascades literature' is inappropriate outside the discrete signal world. Among our simplest findings is that outside BHW's discrete signal world, cascades need not arise: No decision need ever be a foregone conclusion even during a herd. With these two notions decoupled, the analysis is much richer, and it suggests why we must admit a general signal space, and adopt our general stochastic process approach. For instance, we show that learning is incomplete exactly when private signals are uniformly bounded in strength (Theorem 1). Then and only then can bad herds possibly arise in the herding model (Theorem 3).

The explanation we provide for herding is that (the standard) convergence of beliefs implies action convergence: The action frequency settles down, and is consistent with the limit belief. Perfect conformity arises in the pure herding model because contrary actions radically swing beliefs; for any rational desire to deviate must then be shared by all successors. But uniformly identical preferences is neither a realistic nor general assumption. Adding 'noise' (some individuals randomly committed to different actions) to this model is a useful interim step. Yet it is far short of our main contribution, being equivalent to rational agents with different dominant preference types and strategies. Not surprisingly, this statistical noise washes out in the long run, and does not affect convergence.

In this paper, we more generally assume that individuals entertain possibly different preferences over actions; further, types are unobserved, so that only statistical inferences may be drawn about any given individual. Taste diversity with hidden preferences aptly describes numerous cited or motivating examples of herding in the literature, such as restaurant choice, or financial decisions. This twist yields our most novel economic findings. The standard herding outcome is robust to individuals having identical ordinal but differing cardinal (vNM) preferences. With multiple rational preference types, not all ordinally alike, an interior rational expectations dynamic steady-state non-degenerately emerges: It may be impossible to draw any clear inference from history even while it continues to accumulate privately-informed decisions. Further, this incomplete learning pooling outcome exists even with unbounded beliefs, when an incorrect herd is impossible.

Let us fix ideas and illustrate this *confounded learning* possibility with a perhaps familiar example. Suppose that on a highway under construction, depending on how the detours are arranged, those going to Houston should take either the high or low off-ramps (in states H and L), with the opposite for those headed toward Dallas. If 70% are headed toward Houston, then absent any strong signal to the contrary, Dallas-bound drivers should take the lane 'less traveled by'. This yields two separating herding outcomes: 70% high or 70% low, as predicted by armchair application of the herding logic. But another rather subtle possibility may arise, revealed by a careful analysis. For as the chance q that observed history accords state H rises from 0 to 1, the probability that a Houston driver takes the high road gradually rises from 0 to 1, and conversely for Dallas drivers. Thus, the fractions $\psi_H(q)$ and $\psi_L(q)$ in the right lane in states H, L each rises (perhaps nonmonotonically) from 0.3 to 0.7. If for some q, a random car is equilikely in states H and L to go high, or $\psi_H(q) = \psi_L(q)$, then no inference can be drawn from additional decisions: Learning stops. While existence of such a fixed point exists is not obvious, Theorems 1 and 2 prove that for nondegenerate models, confounding outcomes co-exist with the cascade possibilities.

Our confounded learning outcome is generic when two types have opposed preferences, assuming uniformly bounded private signals. With unbounded signals, it emerges for sufficiently strongly opposed vNM preferences, and not too unequal population frequencies. In either case, $\psi_H(q) > \psi_L(q)$ for small enough q, and $\psi_H(q) < \psi_L(q)$ for large enough q.

Two stochastic processes constitute the building blocks for our theory: the public likelihood ratio is a conditional martingale, and the vector (action taken, likelihood ratio) a Markov chain. Martingale and Markovian methods are standard methods for *ruling out* potential limit outcomes of learning. But our major technical innovation concerns their stability: Given multiple limit beliefs, must we converge upon any given one? How can we *rule in* any limit? For instance, even if our earlier confounding outcome with driving robustly exists, must we converge upon it? We have found a simple easily checked condition for the *local stochastic stability* of a Markov-martingale process near a fixed point (Theorem 4). This yields a general and new property of Bayesian learning dynamics. In our context, assume that near any fixed point, posterior beliefs are not degenerately equally responsive to priors for every action taken, but are monotonely so (a higher prior yields a higher posterior belief). Then the belief process tends with positive chance and exponentially fast to that fixed point if starts nearby. Thus, (i) an action which can be herded upon, will then be herded upon for nearby beliefs, while (ii) convergence to our new confounding outcome occurs with positive chance, and necessarily rapidly (Theorem 5).

Section 2 gives a common framework for the paper. Section 3 illustrates our findings in three examples. We then proceed along two technical themes. Via Markov-martingale means, section 4 describes the action and belief limits; the confounding outcome is our key innovation here. Section 5 presents our new stability result, and shows when a long-run outcome arises. Extension to finitely many states is addressed in the conclusion; there we also describe more substantial extensions of the paper, as well as related literature. More detailed proofs and some essential new math results are appendicized.

2. THE COMMON FRAMEWORK

2.1 The Model

States. There are $S = 2$ payoff-relevant *states of the world,* the high state $s = H$ and the low state $s = L$. As is standard, there is a common prior belief — WLOG, a flat prior $\Pr(H) = \Pr(L) = 1/2$. Our results extend to any finite number S of states, but at significant algebraic cost, and so this extension is addressed in the conclusion (§6.1).

Private Beliefs. An infinite sequence of individuals $n = 1, 2, \ldots$ enters in an exogenous order. Individual n receives a random *private signal* about the state of the world, and then, computes via Bayes' rule his *private belief* $p_n \in (0, 1)$ that the state is H. Given the state $s \in \{H, L\}$, the private belief stochastic process $\langle p_n \rangle$ is i.i.d., with conditional c.d.f. F^s. These distributions are sufficient for the state signal distribution, and obey a joint restriction implicit below. The curious reader may jump immediately to Appendix A, which summarizes this development, and explores the results we need.

We assume that no private signal, and thus no private belief, perfectly reveals the state of the world: This ensures that F^H, F^L are mutually absolutely continuous, with common support, say $\mathrm{supp}(F)$. Thus, there exists a positive, finite Radon-Nikodym derivative $f = dF^L/dF^H : (0, 1) \to (0, \infty)$. And to avoid trivialities, we assume that some signals are informative: This rules out $f = 1$ almost surely, so that F^H and F^L do not coincide. When F^s is differentiable ($s = L, H$), we shall denote its derivative by f^s.

The convex hull $\mathrm{co}(\mathrm{supp}(F)) \equiv [\underline{b}, \bar{b}] \subseteq [0, 1]$ plays a major role in the paper. Note that $\underline{b} < 1/2 < \bar{b}$ as some signals are informative. We call the private beliefs *bounded* if $0 < \underline{b} < \bar{b} < 1$, and *unbounded* if $\mathrm{co}(\mathrm{supp}(F)) = [0, 1]$.

Individual Types and Actions. Every individual makes one choice from a finite action menu $\mathcal{M} = \{1, \ldots, M\}$, with $M \geq 2$ actions. We allow for heterogeneous preferences of successive individuals — the only other random element. A model with multiple but observable types is informationally equivalent to a single preference world. So assume instead that all types are private information. There are finitely many *rational types* $t = 1, \ldots, T$ with different preferences. Let λ^t be the known proportion of type t.

We also introduce M *crazy* types. Crazy type m arrives with chance $\kappa_m \geq 0$, and always chooses action m. One could view these as rational types with state independent preferences, and unlike everyone else, a single dominant action. We assume a positive fraction $\kappa = 1 - (\kappa_1 + \cdots + \kappa_M) > 0$ of payoff-motivated rational individuals. Rational and crazy types are spread i.i.d. in sequence, and independently of the belief sequence $\langle p_n \rangle$.

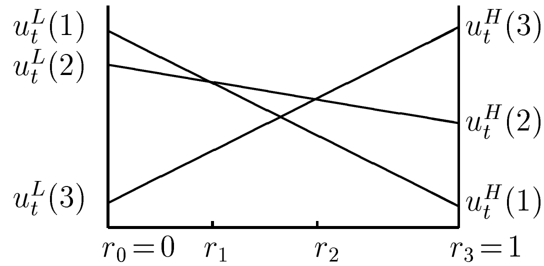

Figure 1: **Expected Payoff Frontier.** The diagram depicts the expected payoff of each of three actions as a function of the posterior belief r that the state is H. A rational individual simply chooses the action yielding the highest payoff. Here 2 is an insurance action, and 1 and 3 are extreme actions.

Payoffs. In state $s \in \{H, L\}$, each rational type t earns payoff $u_t^s(m)$ from action m (for precision, sometimes a_m^t), and seeks to maximize his expected payoff. For each rational type, $(M \geq) \, M_t \geq 2$ actions are not weakly dominated, and generically no one action is optimal at just one belief, and no two actions provide identical payoffs in all states. Each type t thus has $(S=) \, 2$ *extreme actions,* each strictly optimal in some state. The other $M_t - 2$ *insurance actions* are each taken at distinct intervals of unfocused beliefs.

Given a posterior belief $r \in [0, 1]$ that the state is H, the expected payoff to type t of choosing action m is $r u_t^H(m) + (1 - r) u_t^L(m)$. Figure 1 depicts the next summary result.

Lemma 1 *For each rational type t, $[0, 1]$ partitions into subintervals $I_1^t, \ldots, I_{M_t}^t$ touching at endpoints only, with undominated action $m \in \mathcal{M}_t \subseteq \mathcal{M}$ optimal exactly for beliefs $r \in I_m^t$.*

With multiple types, we must introduce T labels for every action. Permuting $\mathcal{M}_t$, we order rational type t's actions $a_1^t, \ldots, a_{M_t}^t$ by relative preference in state H, with a_M^t most preferred. So to be clear, if we order actions from least to most preferred by type t in state H, then action m has rank $\xi = \xi_m^t$ if $m = a_\xi^t$. By Lemma 1, type t's mth *action basin* is $I_m^t \equiv [r_{m-1}^t, r_m^t]$, with ordered boundaries $0 = r_0^t < r_1^t < \cdots < r_{M_t}^t = 1$; thus, extreme actions a_1^t and $a_{M_t}^t$ are optimal for type t in states L and H, and insurance actions $a_2^t, \ldots, a_{M_t - 1}^t$ are each best for some interior beliefs. The tie-breaking rule is WLOG that type t chooses a_m^t over a_{m+1}^t at belief r_m^t. Type t has a *stronger preference* for action a_m^t the larger is the basin I_m^t. Rational types t and t' have *opposed preferences* over actions m and m' if $(\xi_m^t - \xi_{m'}^t)(\xi_m^{t'} - \xi_{m'}^{t'}) < 0$ — i.e. their ordinal preferences for them in state H, and thus in state L, are reversed. With just a single rational type, we suppress t-superscripts, and likewise strictly order belief thresholds as $0 = r_0 < r_1 < \cdots < r_M = 1$.

2.2 The Individual Bayesian Decision Problem

Before acting, every rational individual observes his type t, his private belief p, and the entire ordered action *history* h. His decision rule then maps p and h into an action. We look for a Bayesian equilibrium, where everyone knows all decision rules, and can compute the chance $\pi^s(h)$ of any history h in each state s. This yields a *public belief* $q(h) = \pi^H(h)/(\pi^H(h)+\pi^L(h))$ that the state is H, i.e. the posterior given h and a neutral private belief $p = 1/2$. Applying Bayes rule again yields the *posterior belief* r in terms of q and p:

$$r = r(p,q) = \frac{p\ \pi^H(h)}{p\ \pi^H(h) + (1-p)\ \pi^L(h)} = \frac{pq}{pq + (1-p)(1-q)} \tag{1}$$

As belief q is informationally sufficient for the underlying history data h, we now suppress h.

Since the RHS of (1) is increasing in p, there are *private belief thresholds* $0 = p_0^t(q) \leq p_1^t(q) \leq \ldots \leq p_M^t(q) = 1$, such that type t optimally chooses action a_m^t iff his private belief satisfies $p \in (p_{m-1}^t(q), p_m^t(q)]$, given the earlier tie-break rule. Furthermore, each threshold $p_m^t(q)$ is decreasing in q. A *type-t cascade set* is the set of public beliefs $J_m^t = \{q \,|\, \mathrm{supp}(F) \subseteq [p_{m-1}^t(q), p_m^t(q)]\}$. So type t a.s. takes action a_m^t for any $q \in \mathrm{int}\,(J_m^t)$, since the posterior $r(p,q) \in I_m^t$ for all p. It follows that any cascade set lies inside the corresponding action basin, so that $J_m^t \subseteq \mathrm{int}\,(I_m^t)$. For if all private beliefs yield action a_m^t, then so must the neutral belief.

As is standard, call a property *generic* (resp. *nondegenerate* or *robust*) if the subset of parameters for which it holds is open and dense (resp. open and nonempty).

Lemma 2 *For each action a_m^t and type t, J_m^t is a possibly empty interval. Also,*
(a) With bounded private beliefs, $J_1^t = [0, \underline{q}^t]$ and $J_{M_t}^t = [\bar{q}^t, 1]$ for some $0 < \underline{q}^t < \bar{q}^t < 1$.
(b) With unbounded private beliefs, $J_1^t = \{0\}$, $J_{M_t}^t = \{1\}$, and all other J_m^t are empty.
(c) For generic payoffs u^s, no interior cascade set J_m^t is a single point.
(d) Each J_m^t is larger the smaller is the support $[\underline{b}, \bar{b}]$, and the larger is the action basin $[r_{m-1}^t, r_m^t]$. Only extreme cascade sets *J_1^t and $J_{M_t}^t$ are nonempty for large enough $[\underline{b}, \bar{b}]$.*

The appendicized proofs of Lemma 2-a, b are also intuitive: With bounded private beliefs, posteriors are not far from the public belief q; so for q near enough 0 or 1, or well inside an insurance action basin, all private beliefs lead to the same action. With unbounded beliefs, every public belief $q \in (0,1)$ is swamped by some private beliefs. Next, given the continuous map from payoffs to thresholds $\langle u_t^s(m)\rangle \mapsto \langle r_m^t\rangle$, Lemma 2-$c, d$ follows from (1):

$$q \in J_m^t \quad \text{iff} \quad r_{m-1}^t \leq \frac{\underline{b}q}{\underline{b}q + (1-\underline{b})(1-q)} \quad \text{and} \quad \frac{\bar{b}q}{\bar{b}q + (1-\bar{b})(1-q)} \leq r_m^t \tag{2}$$

The paper requires some more notation. Define $J^t = J_1^t \cup J_2^t \cup \cdots \cup J_{M_t}^t$. A type t is *active* when choosing at least two actions with positive probability. A *cascade* arises — each type's action choice is independent of private beliefs — for public beliefs strictly inside $J = J^1 \cap J^2 \cap \cdots \cap J^T$; however, with unbounded beliefs, there are two cascade beliefs: $q \in J = \{0\} \cup \{1\}$. We put $J^H \equiv \cap_t J_{M_t}^t$, namely where each type t cascades on the action $a_{M_t}^t$, which is optimal in state H. Similarly, we define J^L for state L.

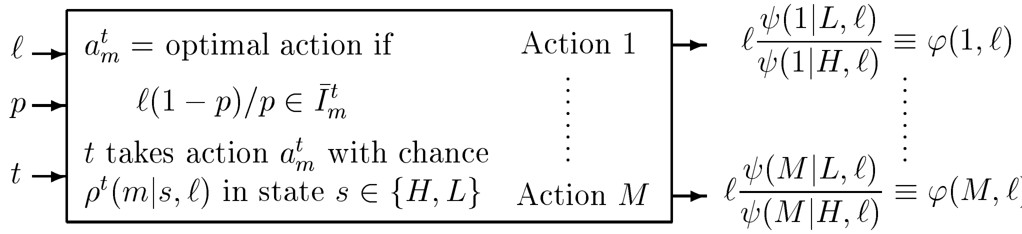

Figure 2: **Individual Black Box.** Everyone bases his decision on both the public likelihood ratio ℓ and his private belief p, resulting in his action choice a_m^t with chance $\psi(a_m^t|L,\ell)$, and a likelihood ratio $\varphi(a_m^t,\ell)$ to confront successors. Type t takes action a_m^t iff his *posterior likelihood* $\ell(1-p)/p$ lies in the interval $\bar{I}_m^t$, where $\bar{I}_1^t,\ldots,\bar{I}_{M_t}^t$ partition $[0,\infty]$.

2.3 Learning Dynamics

Let q_n be the public belief after Mr. n chooses action m_n. It is well-known that $\langle q_n\rangle_{n=1}^\infty$ obeys an unconditional martingale, $\mathcal{E}(q_{n+1}|q_n)=q_n$, and hence almost surely converges to a limit random variable. While we could in principle employ this posterior belief process, we care about the conditional stochastic properties in a given state H. Thus, the *public likelihood ratio* $\ell_n \equiv (1-q_n)/q_n$ that the state is L versus H offers distinct conceptual advantages, and saves time, as it conditions on the assumed true state. The likelihood process $\langle \ell_n\rangle$ similarly will be a convergent conditional martingale on state H.

We then have likelihood analogues of previous notation, now barred: private belief thresholds $\bar{p}_m^t((1-q)/q)\equiv p_m^t(q)$; action basins $\bar{I}_m^t$, where $r\in I_m^t$ iff $(1-r)/r\in\bar{I}_m^t$; and cascade sets $\bar{J}_m^t$, where $q\in J_m^t$ iff $(1-q)/q\in\bar{J}_m^t$; as well as $(\bar{J}^t,\bar{J},\bar{J}^H)$ for (J^t,J,J^H). With bounded private beliefs, $\bar{J}_1^t=[\bar{\ell}^t,\infty]$ and $\bar{J}_{M_t}^t=[0,\underline{\ell}^t]$ for some $0<\underline{\ell}^t<\bar{\ell}^t<\infty$. (Note: this natural notation implies a reverse correspondence: $\bar{\ell}^t=(1-\underline{q}^t)/\underline{q}^t$ and $\underline{\ell}^t=(1-\bar{q}^t)/\bar{q}^t$.) With unbounded private beliefs, $\bar{J}_1^t=\{\infty\}$, $\bar{J}_{M_t}^t=\{0\}$, and all other cascade sets are empty.

Likelihood ratios $\langle \ell_n\rangle_{n=1}^\infty$ are a stochastic process, described by $\ell_0=1$ (as $q_0=1/2$) and transitions

$$\rho^t(a_m^t|s,\ell)=F^s(\bar{p}_m^t(\ell))-F^s(\bar{p}_{m-1}^t(\ell)) \tag{3}$$

$$\psi(m|s,\ell)=\kappa_m+\kappa\textstyle\sum_{t=1}^T\lambda^t\rho^t(m|s,\ell) \tag{4}$$

$$\varphi(m,\ell)=\ell\psi(m|L,\ell)/\psi(m|H,\ell) \tag{5}$$

Here, $\rho^t(m|s,\ell)$ is the chance that a rational type t takes action m, given ℓ, and the true state $s\in\{H,L\}$. So the cascade set $\bar{J}_m^t$ is the interval of likelihoods ℓ yielding $\rho^t(a_m^t|H,\ell)=\rho^t(a_m^t|L,\ell)=1$. Faced with ℓ_n, Mr. n takes action m_n with chance $\psi(m_n|s,\ell_n)$ in state s, whereupon we move to $\ell_{n+1}=\varphi(m_n,\ell_n)$. Figure 2 summarizes (3)–(5).

Our insights are best expressed by considering the pair $\langle m_n,\ell_n\rangle$ as a discrete-time, time-homogeneous *Markov process* on the state space $\mathcal{M}\times[0,\infty)$. Given $\langle m_n,\ell_n\rangle$, the next state is $\langle m_{n+1},\varphi(m_{n+1},\ell_n)\rangle$ with probability $\psi(m_{n+1}|s,\ell_n)$ in state s. Since $\langle\ell_n\rangle$ is a martingale, convergence to any dead wrong belief almost surely cannot occur, since the odds against the truth cannot explode. In summary:

Lemma 3 *(a) The likelihood ratio process $\langle \ell_n \rangle$ is a martingale conditional on state H.*
(b) Assume state H. The process $\langle \ell_n \rangle$ converges almost surely to a r.v., $\ell_\infty = \lim_{n\to\infty} \ell_n$, with $\text{supp}(\ell_\infty) \subseteq [0,\infty)$. So fully incorrect learning ($\ell_n \to \infty$) almost surely cannot occur.

Proof: See Doob (1953) for the martingale character of $\langle \ell_n \rangle$. Convergence follows from the Martingale Convergence Theorem for non-negative, perhaps unbounded, random variables (Breiman (1968), Theorem 5.14); Bray and Kreps (1987) prove this with public beliefs. $\diamondsuit$

Learning is *complete* if 'beliefs' (likelihoods) converge to the truth: $\ell_n \to 0$ in state H. Otherwise, learning is *incomplete*: Beliefs are not eventually focused on the true state H.

Belief convergence then forces *action convergence*: $\langle m_n \rangle$ settles down in frequency, or $\lim_{n\to\infty} N_n(m)/n$ a.s. exists for all m, where $N_n(m) \equiv \#\{m_k = m, k \leq n\}$.

Corollary *Action convergence almost surely obtains, if F^s has no atoms.*

Proof: Let $\varepsilon > 0$. Given belief convergence (Lemma 3), continuity of $\psi(m|s,\ell)$, and the law (4), the chance $\alpha_m(n)$ that action m is chosen almost surely converges, say $\alpha_m(n) \to \alpha_m$. Thus, $\alpha_m(n) \leq \alpha_m + \varepsilon$ for large n, and so $\limsup_{n\to\infty} N_m(n)/n \leq \alpha_m + \varepsilon$. Similarly, $\liminf_{n\to\infty} N_m(n)/n \geq \alpha_m - \varepsilon$. Since $\varepsilon > 0$ is arbitrary, $\lim_{n\to\infty} N_m(n)/n = \alpha_m$. $\diamondsuit$

The literature has so far focused on two more powerful convergence notions. As noted, a cascade means $\ell_n \in \bar{J}$, or finite time belief convergence. Since every later rational type's action is dictated by history, this forces a *herd*, where rational individuals of the same type all act alike. By corollary, the weaker action convergence obtains. We also need the weaker notion of a *limit cascade,* or eventual belief convergence to the cascade set: $\ell_\infty \in J$.

3. EXAMPLES

3.1 Single Rational Type

Consider a simple example, with individuals deciding whether to 'invest' in or 'decline' an investment project of uncertain value. Investing (action $m = 2$) is risky, paying off $u > 0$ in state H and -1 in state L; declining (action $m = 1$) is a neutral action with zero payoff in both states. Indifference prevails when $0 = ru - (1-r)$, so that $r = 1/(1+u)$. Thus, equation (1) defines the private belief threshold $\bar{p}(\ell) = \ell/(u+\ell)$.

A. Unbounded Beliefs Example. Let the private signal $\sigma \in (0,1)$ have state-contingent densities $g^H(\sigma) = 2\sigma$ and $g^L(\sigma) = 2(1-\sigma)$ — as in the left panel of Figure 3. With a flat prior, the private belief $p = p(\sigma)$ then satisfies $(1-p)/p = g^L(\sigma)/g^H(\sigma) = (1-\sigma)/\sigma$ by Bayes' rule, and has the same conditional densities $f^H(p) \equiv 2p$ and $f^L(p) \equiv 2(1-p)$, and c.d.f.'s $F^H(p) = p^2$ and $F^L(p) = 2p - p^2$. So $\text{supp}(F) = [0,1]$, and private beliefs are unbounded; the cascade sets collapse to the extreme points, $\bar{J}_1 = \{\infty\}$, $\bar{J}_2 = \{0\}$.

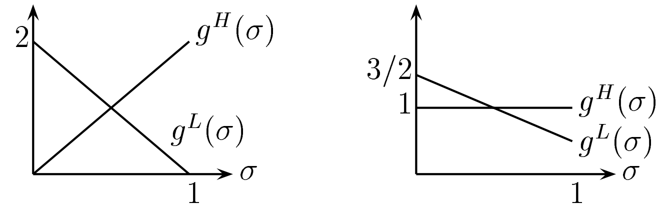

Figure 3: **Signal Densities.** Graphs for the unbounded (left) and bounded (right) beliefs examples. Observe how, in the left panel, signals near 0 are very strongly in favor of state L.

Given private belief c.d.f's F^H, F^L and threshold $\bar{p}(\ell) = \ell/(u+\ell)$, transition chances are $\rho(1|H,\ell) = \ell^2/(u+\ell)^2$ and $\rho(1|L,\ell) = \ell(\ell+2u)/(u+\ell)^2$ by (3). Continuations are $\varphi(1,\ell) = \ell + 2u$ and $\varphi(2,\ell) = u\ell/(u+2\ell)$ by (4), (5). As in Figure 4 (left panel), the only stationary finite likelihood ratio in state H is 0; the limit ℓ_∞ of Lemma 3 is focused on the truth. As the suboptimal action 1 lifts $\ell_n \geq 2u$, an infinite subsequence of such choices would preclude belief convergence: Hence, there must be action convergence (i.e. a herd).

B. Bounded Beliefs Example. Let private signals have density $g^L(\sigma) = 3/2 - \sigma$ on $(0,1)$ in state L, and uniform on $(0,1)$ in state H. Given a flat prior, Bayes' rule yields the private belief $p(\sigma) = 1/(g^L(\sigma)+1) = 2/(5-2\sigma)$, i.e. $p(\sigma) \leq p \Leftrightarrow 2/(5-2\sigma) \leq p \Leftrightarrow (5p-2)/2p \geq \sigma$. Thus, $F^H(p) = (5p-2)/2p$ in state H and $F^L(p) = \int_{p(\sigma)\leq p} g^L(\sigma)\,d\sigma = (5p-2)(p+2)/(8p^2)$ in state L. Each thus has bounded support $[\underline{b}, \bar{b}] = [2/5, 2/3]$.

Since $\bar{p}(\ell) = \ell/(u+\ell)$, active dynamics occur when $\ell \in (2u/3, 2u)$, where equations for F^H, F^L, and (3)–(5) yield $\rho(1|H,\ell) = (3\ell-2u)/2\ell$ and $\rho(1|L,\ell) = (3\ell-2u)(3\ell+2u)/8\ell^2$, as well as $\varphi(1,\ell) = u/2 + 3\ell/4$ and $\varphi(2,\ell) = u/2 + \ell/4$. The likelihood ratio converges by Lemma 3, so that $\bar{J}_1 \cup \bar{J}_2$ contains all possible stationary limits, as in Figure 4 (right panel). A herd on action 1 or 2 must then start eventually — again, lest beliefs not converge. If $\ell_0 \leq 2u/3$, then $\rho(1|H,\ell) = \rho(1|L,\ell) = 0$, i.e. action 2 is always taken, and thus $\langle \ell_n \rangle$ is absorbed in the set $\bar{J}_2 = [0, 2u/3]$. For $\ell_0 \geq 2u$, we similarly find $\bar{J}_1 = [2u, \infty]$.

Here, we may strongly conclude that each limit outcome $2u/3$ and $2u$ arises with positive chance for any $\ell_0 \in (2u/3, 2u)$. As in Figure 4 (right), dynamics are forever trapped in $(2u/3, 2u)$. Since $|\ell_n| \leq 2u$, the Dominated Convergence Theorem yields $E[\ell_\infty \mid H] = \ell_0$. Since $\ell_0 = \pi(2u/3) + (1-\pi)(2u)$ for some $0 < \pi < 1$ whenever $2u/3 < \ell_0 < 2u$, in state H, a herd on action 2 arises with chance π, and one on action 1 with chance $1-\pi$.

In contrast to BHW, if public beliefs are not initially in a cascade set, they never enter one. This holds even though a herd always eventually starts. Visually, it is clear that $\ell_n \in (2u/3, 2u)$ for all n if $\ell_0 \in (2u/3, 2u)$. This also follows analytically: If $\ell_n < 2u$, then $\ell_{n+1} = \varphi(1,\ell_n) = u/2 + 3\ell/4 < u/2 + 3u/2 = 2u$ too. So herds must arise even though a contrarian is never ruled out. This result obtains whenever continuation functions $\varphi(i,\cdot)$ are always increasing. For then $\langle \ell_n \rangle$ never jumps into the closed set $\bar{J}_1 \cup \bar{J}_2$, starting a cascade. Monotonicity is crucially violated in BHW's discrete signal world.

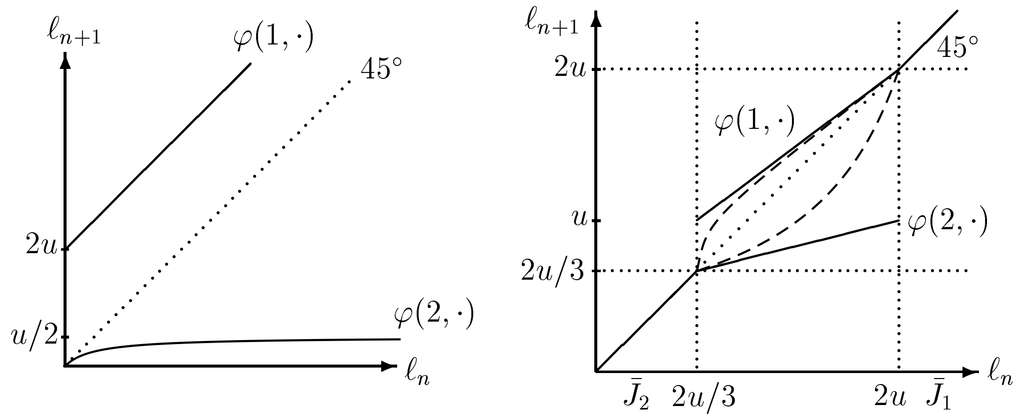

Figure 4: **Continuations and Cascade Sets.** Continuation functions for the examples: unbounded private beliefs (left), and bounded private beliefs (right) with ($\kappa_1 = \kappa_2 = 1/10$) and without noise (dashed and solid lines). The stationary points are where both arms hit the diagonal (as with noise), or where one arm is taken with zero chance ($\ell = 0$ or $\ell = \infty$ in the left panel; $\ell \leq 2u/3$ or $\ell \geq u$ in the right panel without noise). With crazy types the discontinuity vanishes, and isolated deviations no longer have drastic consequences. Graphs here and in figure 5 were generated analytically with PostScript.

C. Bounded Beliefs Example with Noise. Suppose that a fraction of individuals randomly chooses actions. This introduction of a small amount of noise radically affects dynamics, as seen in the right panel of Figure 4. For since all actions are expected to occur, none can have drastic effects. Namely, each $\varphi(i,\cdot)$ is now continuous near the cascade sets at $\underline{\ell} = 2u/3$ and $\bar{\ell} = 2u$. Yet, the limit beliefs are unaffected by the noise, contrary actions being deemed irrational (and ignored) inside the cascade sets.

3.2 Multiple Rational Types

With multiple types, one can still learn from history by comparing proportions choosing each action with the known type frequencies. This inference intuitively should be fruitful, barring nongenericities. A new twist arises: Dynamics may lead to each action being taken with the same positive chance in all states, choking off learning. This incomplete learning outcome is economically different than herding — informational pooling, where actions do not reveal types, rather than the perfect separation that occurs with a type specific informational herd. Mathematically it is a robust interior sink to the dynamics.

Let us consider the driving example from the introduction. Posit that Houston (type U, our 'usual' preferences) drivers should go high (action 2) in state H, low (action 1) in state L, with the reverse true for Dallas (type V) drivers. Going to the wrong city always yields zero, WLOG. The payoff vector of the Houston-bound is $(0, u)$ in state H and $(1, 0)$ in state L; for Dallas drivers, it is $(1, 0)$ and $(0, v)$, where WLOG $v \geq u > 0$. So the preferences are opposed, but not exact mirror images if $v > u$. Types U, V then respectively choose action 1 for private beliefs below $\bar{p}^U(\ell) = \ell/(u + \ell)$, and above $\bar{p}^V(\ell) = \ell/(v + \ell)$.

Assume the same bounded beliefs structure introduced earlier. Assume we start at $\ell_0 \in (2v/3, 2u)$. The transition probabilities for type U are then just as in section 3.1.B:

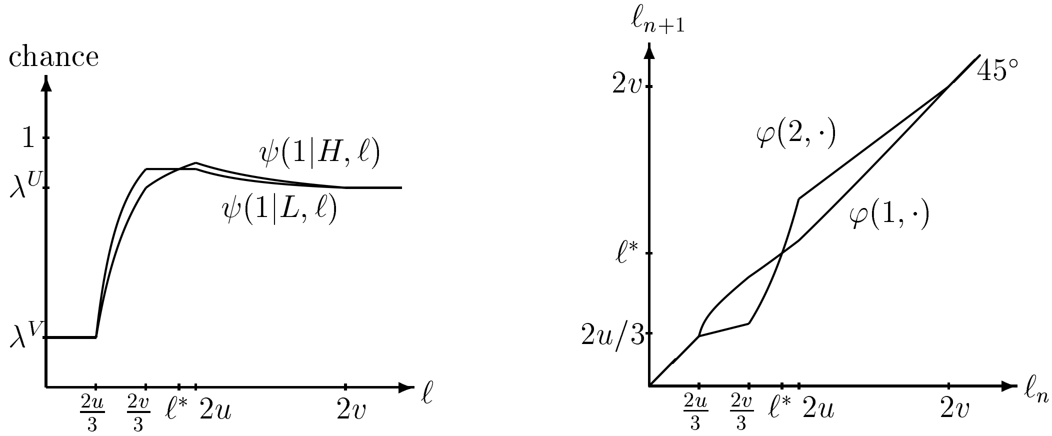

Figure 5: **Confounded Learning.** Based on our BOUNDED BELIEFS EXAMPLE, with $\lambda^U = 4/5$, $u = v/2$. In the left graph, the curves $\psi(1|H,\ell)$ and $\psi(1|L,\ell)$ cross at the confounding outcome $\ell^* = 8v/9$, where no additional decisions are informative. At ℓ^*, 7/8 choose action 1, and strangely 7/8 lies outside the convex hull of λ^V and λ^U — eg. in the introductory driving example, more than 70% of cars may take the high ramp in a confounding outcome. The right graph depicts continuation likelihood dynamics.

$\rho^U(1|H,\ell) = (3\ell - 2u)/2\ell$ and $\rho^U(1|L,\ell) = (3\ell - 2u)(3\ell + 2u)/8\ell^2$, where $\ell \in (2u/3, 2u)$; for type V, we likewise have $\rho^V(1|H,\ell) = (2v - \ell)/2\ell$ and $\rho^V(1|L,\ell) = (2v + \ell)(2v - \ell)/8\ell^2$ by applying (3). The two types take action 2 with certainty in the intervals $\bar{J}_2^U = [0, 2u/3]$ and $\bar{J}_2^V = [2v, \infty]$, respectively. If either these sets or $\bar{J}_1^U = [2u, \infty]$ and $\bar{J}_1^V = [0, 2v/3]$ overlap, as happens with $2v/3 < 2u$ or $2u/3 < 2v$, then only one type ever makes an informative choice for any ℓ, and the resulting analysis for the other type is similar to the single rational type model: just limit cascades, and therefore herds, arise.

Assume no cascade sets overlap. Consider dynamics for $\ell \in (2v/3, 2u)$ given by (4):

$$\psi(1|H,\ell) = \lambda^U \frac{3\ell - 2u}{2\ell} + \lambda^V \frac{2v - \ell}{2\ell} \quad \& \quad \psi(1|L,\ell) = \lambda^U \frac{(3\ell - 2u)(3\ell + 2u)}{8\ell^2} + \lambda^V \frac{(2v - \ell)(2v + \ell)}{8\ell^2}$$

We are interested in a different fixed point depicted in Figure 5, where neither rational type takes any action for sure, and decisions always critically depend on private signals. The two continuations (5) then coincide: $\psi(1|H,\ell^*) = \psi(1|L,\ell^*) \in (0,1)$, and actions convey no information:

$$\frac{\lambda^U}{\lambda^V} = \frac{(2v - \ell)(3\ell - 2v)}{(2u - \ell)(3\ell - 2u)} \equiv h(\ell)$$

If $v > u$ then h maps $(2v/3, 2u)$ onto $(0, \infty)$, and $h(\ell^*) = \lambda^U/\lambda^V$ is solvable for any λ^U, λ^V.

For this example, we can argue that with positive chance, the process $\langle \ell_n \rangle$ tends to ℓ^* if it does not start in a cascade, i.e. in $[0, 2u/3]$ or $[2v, \infty]$. Since each likelihood continuation $\varphi(i, \cdot)$ is increasing, if dynamics start in $[2u/3, \ell^*]$ or $[\ell^*, 2v]$, they are forever trapped there. Assume $\ell_0 \in (2u/3, \ell^*)$. Then $\langle \ell_n \rangle$ is a bounded martingale that tends to the end-points; therefore, the limit ℓ_∞ places positive probability weight on both a limit cascade on $\ell = 2u/3$ and convergence to $\ell = \ell^*$. This example verifies that the latter fixed point robustly exists and is stable, but does not yet explain why.

4. LONG RUN LEARNING

4.1 Belief Convergence: Limit Cascades and Confounded Learning

A. Characterization of Limit Beliefs. That $\langle \ell_n \rangle$ is a martingale in state H rules out non-stationary limit beliefs (such as cycles), and convergence to incorrect point beliefs. Markovian methods then prove that with a single rational type, as in section 3.1, limit cascades arise, or $\ell_n \to \bar{J}$. But with multiple rational types, the example in section 3.2 exhibits another possibility: $\langle \ell_n \rangle$ may converge to where each action is equilikely in all states. Then define the set $\bar{K}$ of *confounding outcomes* as those likelihood ratios $\ell^* \notin \bar{J}$ satisfying

$$\psi(m|s, \ell^*) = \psi(m|H, \ell^*) \quad \text{for all actions } m \text{ and states } s \tag{6}$$

Observe that since $\ell^* \notin \bar{J}$, decisions generically depend on private beliefs. Yet decisions are not informative of beliefs, given the pooling across types. Also, history is still of consequence, for otherwise decisions would then generically be informative. Rather history is precisely so informative as to choke off any new inferences. The distinction with a cascade is compelling: Decisions reflect own private beliefs and history at a confounding outcome, whereas in a limit cascade, history becomes totally decisive, and private beliefs irrelevant.

Markovian and martingale methods together imply that with bounded private beliefs, fully correct learning is impossible. Only a confounding outcome or limit cascade on the wrong action are possible incomplete learning outcomes. Summarizing:

Theorem 1 *Suppose WLOG that the state is H.*

(a) *With a single rational type, a not fully wrong limit cascade occurs:* $\text{supp}(\ell_\infty) \subseteq \bar{J} \backslash \{\infty\}$.

(b) *With a single rational type, and unbounded private beliefs, $\ell_n \to 0$ almost surely.*

(c) *With $T \geq 2$ rational types with different preferences, only a limit cascade that is not fully wrong or a confounding outcome may arise:* $\text{supp}(\ell_\infty) \subseteq \bar{J} \cup \bar{K} \setminus \{\infty\}$.

(d) *With bounded private beliefs, $\ell_\infty \in \bar{J} \setminus \bar{J}^H$ with positive chance provided $\ell_0 \notin \bar{J}^H$. Likewise, if $\ell_0 \neq \ell^*$, no single confounding outcome ℓ^* arises almost surely.*

(e) *Fix payoff functions for all types, and a sequence of private belief distributions with supports $[\underline{b}_k, \bar{b}_k]$ $(k = 1, 2, \ldots)$. If $\bar{b}_k \to 1$ (or $\underline{b}_k \to 0$ in state L), then the chance of an incorrect limit cascade $\ell_\infty \in \bar{J} \setminus \bar{J}^H$ vanishes as $k \to \infty$.*

Proof: First, Lemma 3 asserts $\text{supp}(\ell_\infty) \subseteq [0, \infty)$ in state H, i.e. $\ell_\infty < \infty$ a.s.

Theorems B.1,2 tightly prescribe ℓ_∞: Any $\hat{\ell} \in \text{supp}(\ell_\infty)$ is *stationary* for the Markov process, i.e. either an action doesn't occur, or it teaches us nothing: Absent signal atoms, with continuous transitions, this means that $\psi(m|s, \hat{\ell}) = 0$ or $\varphi(m, \hat{\ell}) = \hat{\ell}$ for all $m \in \mathcal{M}$.

Proof of (a): Assume φ continuous in ℓ. A single rational type must take some action m

with positive chance in the limit $\hat{\ell}$, and thus $\varphi(m,\hat{\ell})=\hat{\ell}$. Since $\psi(m|s,\ell)=\kappa_m+\kappa\rho(m|s,\ell)$,

$$\hat{\ell}=\varphi(m,\hat{\ell})=\hat{\ell}\frac{\kappa_m+\kappa\rho(m|L,\hat{\ell})}{\kappa_m+\kappa\rho(m|H,\hat{\ell})}$$

and so $\rho(m|H,\hat{\ell})=\rho(m|L,\hat{\ell})>0$. Intuitively, statistically constant noisy behavior does not affect long run learning by rational individuals, as it is filtered out. Next, pick the least action m taken in the limit by the rational type with positive chance (i.e. for low enough private beliefs). Since private beliefs are informative (Lemma A.1(c)), m is strictly more likely in state L than H, and is thus informative — unless $\rho(m|H,\hat{\ell})=\rho(m|L,\hat{\ell})=1$. Hence, $\hat{\ell}\in\bar{J}_m$. The appendix analyzes the case of a discontinuous function φ. □

Proof of (b): $\bar{J}=\{0,\infty\}$ by Lemma 2 with unbounded beliefs. So $\text{supp}(\ell_\infty)\subseteq\bar{J}$ and $\ell_\infty<\infty$ a.s. together jointly imply $\ell_\infty=0$ a.s. □

Proof of (c): With $T\geq 2$ types, we now have $\psi(m|s,\hat{\ell})=\kappa_m+\kappa\rho(m|s,\hat{\ell})$, and the previous complete learning deduction fails: $\psi(m|H,\hat{\ell})=\psi(m|L,\hat{\ell})$ for all m need not imply $\rho^t(m|H,\hat{\ell})=\rho^t(m|L,\hat{\ell})=1$ for all t. Instead, we can only assert $\hat{\ell}\in\bar{J}$ or $\hat{\ell}\in\bar{K}$. □

Proof of (d): Recall that $\bar{J}^H=[0,\underline{\ell}]$ where $\underline{\ell}=\min_t\underline{\ell}^t$, and $[\bar{\ell},\infty]\subseteq\bar{J}$ where $\bar{\ell}=\max_t\bar{\ell}^t$, across $t=1,\dots,T$. If $\ell_\infty\in[\bar{\ell},\infty]$ with positive chance, then we are done. Otherwise $\ell_n\leq\bar{\ell}<\infty$ a.s. for all n. By the Dominated Convergence Theorem, the mean of the bounded martingale $\langle\ell_n\rangle$ is preserved in the limit: $E[\ell_\infty]=\ell_0$. So $\text{supp}(\ell_\infty)\subseteq\bar{J}^H\equiv[0,\underline{\ell}]$ fails if $\ell_0>\underline{\ell}$. Similarly, $\ell_\infty=\ell^*$ with probability 1 cannot obtain if $\ell_0\neq\ell^*$. □

Proof of (e): By Lemma 2-a,d, only the extreme cascade sets $\bar{J}_k=[0,\underline{\ell}_k]\cup[\bar{\ell}_k,\infty]$ exist for $\text{co}(\text{supp}(F_k))$ close enough to $[0,1]$ (i.e. large k). If π_k is the chance of an incorrect limit cascade — namely, on action 1 — then $E\ell_\infty\geq\pi_k\bar{\ell}_k$. But $E\ell_\infty\leq\ell_0$ by Fatou's Lemma, so that $\pi_k\leq\ell_0/\bar{\ell}_k$. By (2), the knife-edge cascade public likelihood ratio $\bar{\ell}_k^t$ and the highest private belief $\bar{b}_k^t$ yield the posterior r_1^t, by Bayes rule: $(1-r_1^t)/r_1^t=\bar{\ell}_k^t(1-\bar{b}_k^t)/\bar{b}_k^t$. So $\bar{\ell}_k\to\infty$ as $\bar{b}_k\to 1$, and thus $\pi_k\to 0$. (In state L, $\pi_k\to 0$ as $\underline{b}_k\to 0$, and $\underline{\ell}_k\to 0$.) □

Observe how part (e) makes sense of the bounded versus unbounded beliefs knife-edge, since there is a continuous transition from incomplete to complete learning.[1]

B. Robustness of Confounding Outcomes. Lemma 2 establishes the robustness of cascade sets. However, unlike cascade sets, the existence of confounding outcomes is not foretold by the Bayesian decision problem. They are only inferred ex post, and are nondegenerate phenomena. Fleshing this out, the model *parameters* are the preferences and type/noise proportions $\langle u_t^s(m),\kappa_m,\lambda^t\rangle$, elements of the Euclidean normed space $\mathbb{R}^{SMT+M+T}$. Genericity and robustness are defined with respect to this parameter set.

[1]We think it noteworthy that Milgrom's (1979) auction convergence theorem, which also concerns information aggregation but in an unrelated context, turns on a bounded-unbounded signal knife-edge too.

Theorem 2 *Assume there are $T \geq 2$ rational types.*
(a) Confounding outcomes robustly exist, and are invariant to noise.
(b) At any confounding outcome, at least two rational types are not in a cascade set.
(c) For generic parameters, at a confounding outcome, at most two actions are taken by active rational types *(i.e. those who are not in a cascade).*
(d) If belief distributions are discrete, confounded learning is non-generic.
(e) With $M > 2$ actions and unbounded beliefs, generically no confounding outcome exists.
(f) At any confounding outcome, some pair of types has opposed preferences.
(g) Assume $M = 2$ and some types with opposed preferences. With atomless bounded beliefs and $T = 2$, a confounded learning point exists generically, provided the two types are both active over some public belief range. With atomless unbounded beliefs and $f^H(1), f^L(0) > 0$, a confounding point exists if the opposed types have sufficiently different preferences.

Before the proofs, observe that while generically only two actions are active at any given confounding outcome (part (c)), nondegenerate models with $M > 2$ actions still have confounding outcomes. For with bounded beliefs, only two actions may well be taken over a range of possible likelihoods ℓ. Second, note that a confounding outcome is in one sense a more robust failure of complete learning than is an incorrect limit cascade, since it arises even with unbounded private beliefs (part (g)).

Proof of (a): This has almost been completely addressed by the third example in section 3, which is nondegenerate in the specified sense. Invariance to noise follows because shifting κ_m identically affinely transforms both sides of (6), given (4). □

Proof of (b): By the proof of Theorem 1-a, if all but one rational type is in a cascade in the limit, then so is that type, given (6). So at least two rational types are active. □

Proof of (c): Consider the equations that a confounding outcome ℓ^* must solve. First, with bounded beliefs, some actions may never occur at ℓ^*. Assume that $M_0 \leq M$ actions are taken with positive probability at ℓ^*. Next, given the adding-up identity $\sum_{i=1}^{M_0} \psi(m_i|H,\ell) = \sum_{i=1}^{M_0} \psi(m_i|L,\ell) = 1$, (6) reduces to $M_0 - 1$ equations of the form $\psi(m|H,\ell) = \psi(m|L,\ell)$, in a single unknown ℓ. As the equations generically differ for active rational types, they can only be solved when $M_0 = 2$. □

Proof of (d): Rewrite (6) as $\sum_t \lambda^t \rho^t(m|H,\ell) = \sum_t \lambda^t \rho^t(m|L,\ell)$. For F^H, F^L discrete, each side assumes only countably many values. As $\rho^t(m|H,\ell) - \rho^t(m|L,\ell)$ generically varies in t, the solution of (6) generically vanishes as the λ^t weights are perturbed. □

Proof of (e): With unbounded private beliefs, for any $\ell \in (0, \infty)$, all M actions are taken with positive chance. By part (c), confounding outcomes generically can't exist. □

Proof of (f): If actions 1 and 2 are taken at ℓ^*, and all types prefer $m = 2$ in state H, then $m = 2$ is good news for state H, whence ℓ^* could not be a confounding outcome. □

Proof of (g): Consider an interval $[\underline{\ell}, \bar{\ell}]$ between any two consecutive portions of the cascade set $\bar{J}$. The appendix proves that under our assumptions, $\psi(1|H,\ell)$ exceeds $\psi(1|L,\ell)$ near $\underline{\ell}$ iff $\psi(1|L,\ell)$ exceeds $\psi(1|H,\ell)$ near $\bar{\ell}$. Without signal atoms, both are continuous functions, and therefore must cross at some interior point $\ell^* \in (\underline{\ell}, \bar{\ell})$. □

For an intuition of why confounding points exist (Theorem 2-g), consider our binary action Texas driving example depicted in Figure 5. By continuity, it suffices to explain when one should expect the antisymmetric ordering $\psi(1|L,\ell) \gtrless \psi(1|H,\ell)$, respectively, for ℓ small (near $\underline{\ell}$) and large (near $\bar{\ell}$). The critical idea here is that barring a cascade, a partially-informed individual is more likely to choose a given course of action when he is right than when he is wrong (Lemma A.1).

Since Houston drivers wish to go high in state H, uniformly across public beliefs, more Houston drivers will go high in state H than in state L. The reverse is true for Dallas drivers. The required antisymmetric ordering clearly occurs if and only if most contrarians are of a different type near $\underline{\ell}$ than near $\bar{\ell}$. With bounded beliefs, as in the example of §3.1B, this is true simply because a different type is active at each extreme, for generic preferences.

With unbounded beliefs, the above shortcut logic fails, as both types are active for all unfocused beliefs. The key economic ingredients for the existence of a confounding point are then (i) not too unequal population type frequencies, and (ii) sufficiently strongly opposed preferences by the rational types. To see (i), assume the extreme case with rather disparate type frequencies, and nearly everyone Houston-bound. No antisymmetric ordering can then occur, as $\psi(1|L,\ell) < \psi(1|H,\ell)$ for all $\ell \in (\underline{\ell}, \bar{\ell})$. Next, condition (ii) ensures that the two types' action basins for opposing actions grow, and contrarians of each type — those whose private beliefs oppose the public belief — increase as we approach one extreme, and decrease as we approach the other, in opposition.

To illustrate the intuition in an unbounded beliefs variant of the driving example (from the appendix proof of Theorem 2 (g)), a confounding point exists for $u/v \gtrless \lambda^U/\lambda^V \gtrless v/u$. In words, type frequencies are not too far apart (λ^U/λ^V not too big or small) — and given their disparity, preferences are sufficiently strongly opposed (u/v big or small enough).

4.2 The Traditional Case: Herds with a Single Rational Type and No Noise

We have already argued that actions converge in frequency. Without noise, this can be easily strengthened. After any deviation from a potential herd, or finite string of identical actions, an uninformed successors will necessarily follow suit. In other words, the public belief has radically shifted. As in section 3, this logic proves that herds arise.

Theorem 3 *Assume a single rational type and no noise.*
(*a*) *A herd on some action will almost surely arise in finite time.*
(*b*) *With unbounded private beliefs, individuals almost surely settle on the optimal action.*
(*c*) *With bounded private beliefs, absent a cascade on the most profitable action from the outset, a herd arises on another action with positive probability.*

Proof: Part (*a*) follows from the convergence result of Theorem 1(*a*) and the following *Overturning Principle*: If Mr. n chooses any action m, then before $n+1$ observes his own private signal, he should find it optimal follow suit because he knows no more than n, who rationally chose m. To wit, $\ell_{n+1} \in \bar{I}_m$ after n's action, and a single individual can overturn any given herd. The appendix analytically verifies this fact. From Lemma 2, $\bar{J}_m \subset \operatorname{int}(\bar{I}_m)$, so that when $\operatorname{supp}(\ell_\infty) \subseteq \bar{J}_m$, eventually $\ell_n \in \operatorname{int}(\bar{I}_m)$, precluding further overturns.

Finally, parts (*b*) and (*c*) are corollaries of part (*a*) and Theorem 1 (*b*) and (*d*). $\diamondsuit$

This characterization result extends the major herding finding in BHW to general signals and noise. (BHW also handled several states, addressed here in §6.1). The analysis of BHW — which did not appeal to martingale methods — only succeeded because their stochastic process necessarily settled down in some stochastic finite time. Strictly bounded beliefs so happens to be the mainstay for their bad herds finding. Full learning doesn't require the perfectly revealing signals in BHW, ruled out here.

5. STABLE OUTCOMES AND CONVERGENCE

Above, we have identified the candidate limit outcomes. But one question remains: Are these limits merely possibilities, or probabilities? We address this with local stability results. Convergence is rapid, and this affords insights into why herding occurs.

5.1 Local Stability of Markov-Martingale Processes

We state this theoretical finding in some generality. Let $\langle (m_n, x_n) \rangle$ be a discrete-time Markov process on $\mathcal{M} \times \mathbb{R}$, for some finite set $\mathcal{M} = \{1, 2, \ldots, M\}$, with transitions given by

$$x_n = \varphi(m, x_{n-1}) \quad \text{with probability } \psi(m|x_{n-1}) \quad (m \in \mathcal{M}) \tag{7}$$

Further assume that $\langle x_n \rangle$ is a martingale: that is, $x \equiv \sum_{m=1}^{M} \psi(m|x)\varphi(m,x)$. Let $\hat{x}$ be a *fixed point* of (7), so that for all m: either $\hat{x} = \varphi(m, \hat{x})$ or $\psi(m|\hat{x}) = 0$. Our focus will be on functions φ and ψ that are C^1 (once continuously differentiable) at $\hat{x}$. If $x_n \to \hat{x}$, then θ is a *convergence rate* provided $\hat{\theta}^{-n}(x_n - \hat{x}) \to 0$ at all $\hat{\theta} > \theta$. Observe that if θ is a convergence rate of $\langle x_n \rangle$, then so is any $\theta' > \theta$. Also, if $x_{n_0} = \hat{x}$ for some n_0, then $\theta = 0$.

Appendix C develops a local stability theory for such Markov processes. For an intuitive overview, recall that near the fixed point $\hat{x}$, $\langle x_n \rangle$ is well approximated by the following linearized stochastic difference equation: starting at $(m_n, x_n - \hat{x})$, the continuation is $(m_{n+1}, x_{n+1} - \hat{x}) = (m, \varphi_x(m, \hat{x})(x_n - \hat{x}))$ with chance $\psi(m|\hat{x})$. Now, for a linear process $\langle y_n \rangle$, where $y_{n+1} = a_m y_n$ with chance p_m $(m = 1, \ldots, M)$, we have $y_n = a_1^{\chi_1(n)} \cdots a_M^{\chi_M(n)} y_0$, where $\chi_m(n)$ counts the m-continuations in the first n steps. Since $\chi_m(n)/n \to p_m$ almost surely by the Strong Law of Large Numbers, the product $a_1^{p_1} \cdots a_M^{p_M}$ fixes the long-run stability of the stochastic system $\langle y_n \rangle$ near 0. Accordingly, the product $\prod_{m=1}^M \varphi_x(m, \hat{x})^{\psi(m|\hat{x})}$ determines the local stability of the original non-linear system (7) near $\hat{x}$. Rigorously,

Theorem 4 *Assume that at a fixed point $\hat{x}$ of the Markov-martingale process* (7), $\psi(m|\cdot)$ *and $\varphi(m, \cdot)$ are C^1 (resp. left or right C^1), for all m. Assume $\varphi(m, \cdot)$ is weakly increasing near $\hat{x}$ for all m, and $\varphi_x(m, \hat{x}) \neq 1$ for some m with $\psi(m|\hat{x}) > 0$. Then $\hat{x}$ is locally stable: with positive chance, $x_n \to \hat{x}$ (resp. $x_n \uparrow \hat{x}$ or $x_n \downarrow \hat{x}$) for x_0 near $\hat{x}$ (resp. below or above $\hat{x}$). Whenever $x_n \to \hat{x}$, convergence is almost surely at the rate $\theta \equiv \prod_{m=1}^M \varphi_x(m, \hat{x})^{\psi(m|\hat{x})} < 1$.*

Proof: For the in-text proof here, we make the simplifying assumption that $\hat{x} = \varphi(m, \hat{x})$ for all m. By Corollary C.1, given a frequency-weighted geometric mean $\theta < 1$ of the continuation derivatives, $x_n \to \hat{x}$ with positive chance, and at rate θ.

Differentiate the martingale identity $x \equiv \sum_{m=1}^M \psi(m|x)\varphi(m, x)$ to get:

$$1 = \sum_{m=1}^M \psi(m|x)\varphi_x(m, x) + \sum_{m=1}^M \psi_x(m|x)\varphi(m, x) \tag{8}$$

Differentiating the probability sum $\sum_{m=1}^M \psi(m|x) \equiv 1$ yields implies $\sum_{m=1}^M \psi_x(m|x) = 0$. Because $\varphi(m, \hat{x}) = \hat{x}$ for all m at the fixed point $\hat{x}$, the second sum in (8) vanishes at $\hat{x}$, and we are left with $\sum_{m=1}^M \psi(m|x)\varphi_x(m, x) = 1$. The continuation slopes $\varphi_x(m, x) \geq 0$ are not all equal since we have assumed $\varphi_x(m, \hat{x}) \neq 1$ for some m. Hence, the *arithmetic mean - geometric mean (AM-GM) inequality* holds strictly. This yields $\theta < 1$.

The proof admitting the event that $\varphi(m, \hat{x}) \neq \hat{x}$ and $\psi(m, \hat{x}) = 0$ is appendicized. ◇

5.2 Cascade Sets Attract Herds, and Confounded Learning Arises

We now apply Theorem 4 to $\langle \ell_n, m_n \rangle$, and prove that both incomplete learning outcomes do arise. Theorem 5-a, b below shows that a limit cascade develops with positive chance if beliefs initially lie near a cascade set. Absent belief atoms, and with monotonic continuation functions $\varphi(m, \cdot)$, this follows from Theorem 4 provided $\varphi_\ell(m_1, \ell^*) \neq \varphi_\ell(m_2, \ell^*)$, some $m_1 \neq m_2$. A positive private belief tail density guarantees this inequality. The details, as well as consideration of nonmonotonic $\varphi(m, \cdot)$, are appendicized.

Next, part (c) proves that $\langle \ell_n \rangle$ tends with positive chance to the confounding outcome ℓ^*, even though it is an isolated interior point. The reason for this is that ℓ^* can be locally stochastically stable. This occurs provided $\varphi_\ell(m_1, \ell^*) \neq \varphi_\ell(m_2, \ell^*)$, where actions m_1 and m_2 are taken with positive probability at ℓ^*. Not being an identity, this inequality is generically satisfied. For nondegenerate parameters, both derivatives are positive, as the example shows. We let *confounded learning* denote convergence to the confounding point, namely the event where $\ell_n \to \ell^*$.

Theorem 5 *(a) Assume bounded private beliefs and let F^H, F^L have C^1 tails, with $f^H(\underline{b}), f^L(\bar{b}) > 0$. If ℓ_0 is close enough to the cascade set component $\hat{J} = \bar{J}^1_{m_1} \cap \cdots \cap \bar{J}^T_{m_T} \subset \bar{J}$, and $\hat{J}$ is not a single point, then $\ell_n \to \hat{\ell} \in \hat{J}$ with positive chance, and at some rate $\theta < 1$. Whenever $\ell_\infty \in \hat{J}$, a herd develops: eventually type t takes $a^t_{m_t}$.*
(b) Assume unbounded private beliefs. If $\inf \bar{K} > 0$, then $\Pr(\ell_\infty = 0) \to 1$ as $\ell_0 \to 0$.
(c) Confounded learning occurs: For nondegenerate data, points $\ell^* \in K$ are locally stable.

With atomic tails of the private belief distributions — as in BHW's analysis with discrete private signals — each active continuation $\varphi(m, \cdot)$ is discontinuous near its associated cascade set $\bar{J}_m$; therefore, $\langle \ell_n \rangle$ might well leap over a small enough cascade set $\bar{J}_m$. By graphical reasoning, when F^H, F^L have C^1 tails, dynamics can jump into the cascade set $[2u, \infty)$ in Figure 4 with a left derivative $\varphi_{\ell-}(1, 2u) < 0$. More generally, a single action can toss everyone into a cascade with a nonmonotonic continuation; and a confounding outcome need not be stable. While we know of no simple sufficient conditions that guarantee monotonic continuation functions, we can show how a nonmonotonicity may arise: Since $\varphi(1, \ell) = \ell F^L(\bar{p}(\ell))/F^H(\bar{p}(\ell))$, the private belief odds $F^L(\bar{p}(\ell))/F^H(\bar{p}(\ell))$, decreasing by Lemma A.1-e, might be more than unit-elastic in the prior likelihood ratio ℓ. This arises with near-atomic private beliefs (see our MIT working paper with the same title).

Finally, part (a) provides a direct explanation as to why herds arise with a single type. Our current logic is indirect: Martingale methods force belief convergence, and a failure to herd precludes belief convergence. Yet a herd arises iff eventually the public belief swamps all private beliefs. Here we see more directly that belief convergence is exponentially fast, and *ipso facto* the sequence of contrarian chances is summable. Namely, if A_k denotes the event that Mr. k deviates once a herd has begun, then on almost surely all approach paths to the fixed point, $\Pr(A_k)$ vanishes geometrically fast given the exponential stability, so that $\sum \Pr(A_k) < \infty$. By the first Borel-Cantelli Lemma, $\Pr(A_k \text{ infinitely often}) = 0$ — even though the events $\{A_k\}$ are not independent. In other words, an infinite sequence of contrarians is impossible, and a herd must eventually start.

6. CONCLUSION AND EXTENSIONS

6.1 Multiple States

(a) *The Revised Model.* We can admit any finite number S of states. Rather than partition $[0,1]$ into closed subintervals, optimal decision rules instead slice the unit simplex in $\mathbb{R}^{S-1}$ into closed convex polytopes. Belief thresholds become hyperplanes. Fixing a reference state, the likelihood ratios $\ell^1, \ldots, \ell^{S-1}$ are each a convergent conditional martingale.

(b) *Revised Convergence Notions.* The extreme interval cascade sets for each rational type generalize to convex sets around each corner of the belief simplex. With unbounded beliefs, all cascade sets lie on the boundary of the simplex; with bounded beliefs, interior public beliefs near the corners lie in cascade sets, and there can exist insurance action cascade sets away from the boundary. Limit cascades must arise with a single rational type.

Long-run ties where two or more actions are optimal in a given state may non-generically occur, as BHW note. Barring this possibility, a herd occurs eventually with probability one with a single rational type and no noise. If some action is optimal in several states of the world, then its cascade set will contain the simplex face spanned by these states. Even with unbounded beliefs, complete learning need not obtain, as the limit belief may lie on this simplex face but not at a corner. Yet there is *adequate learning,* in the terminology of Aghion, Bolton, Harris, and Jullien (1991): Eventually an optimal action is chosen.

(c) *Robustness of Confounding Outcomes.* At a confounding outcome, for each ratio ℓ^s, all M possible continuations must coincide. This produces $(M-1)(S-1)$ generically independent equations to be solved in the $S-1$ likelihood ratios; this is possible for nondegenerate data if $M=2$, while $M>2$ yields an over-identified system. Thus, the existence of confounding outcomes is robust: Theorem 2 reads exactly as before.

(d) *Robustness of Local Stability.* The extension of Theorem 5-a (stable interior cascade sets) eludes us as it is unclear which limit point to focus on as we near the hyperplane-frontier. The proof of Theorem 5-b (stable extreme cascade sets) is robust to several states, since we may simply consider the likelihood ratio of all other states to the reference state.

For Theorem 5-c (stable confounding outcomes), we assume the generic property that in a neighborhood of the confounding outcome $\hat{\ell}$, only two actions m_1, m_2 are active (taken by rational types with positive chance). For each inactive action m, we have $\varphi(m,\ell)=\ell$, and thus the Jacobian $D_\ell\varphi(m,\ell)=I$. Differentiating the martingale property yields $\sum_{m=1}^{M}\psi(m|S,\ell)D_\ell\varphi(m,\ell)=I$. Since inactive actions leave ℓ unchanged in this neighborhood, we may as well focus entirely on the occurrences of m_1 and m_2, now with chances $\hat{\psi}(m_i|S,\ell)=\psi(m_i|S,\ell)/(\psi(m_1|S,\ell)+\psi(m_2|S,\ell))$. Then $\hat{\psi}(m_1|S,\ell)D_\ell\varphi(m_1,\hat{\ell})+$

$\hat{\psi}(m_2|S,\ell)D_\ell\varphi(m_2,\hat{\ell}) = I$. By Theorem C.2, $\hat{\ell}$ is locally stable under the weak condition that the Jacobians $D_\ell\varphi(m_i,\hat{\ell})$ both have real, distinct, positive, and non-unit eigenvalues. Since eigenvalues are zeros of the characteristic polynomial, this requirement is nondegenerate. It is the natural generalization of non-negative derivatives of φ.

6.2 Links to the Experimentation Literature

Our (1998) companion paper draws a formal parallel between bad herds and the well-known phenomenon of incomplete learning in optimal experimentation. It also shows that even a patient social planner (unable to observe private signals) who controls action choices will with positive probability succumb to bad herds, assuming bounded private beliefs. The confounded learning that we find is formally similar to McLennan's (1984) confounded learning in experimentation, in the precise sense that probability density functions of the observables coincide across all states at an interior belief. McLennan's finding is not robust to more than two observable signals, nor is ours to more than two active actions. Thus, if his buyers desire to buy more than one unit of the good, confounded learning is a nondegenerate outcome. However, as noted, there may be three or more actions in our informational herding setting, even though only two are active at some public belief.

APPENDICES

A. ON BAYESIAN UPDATING OF DIVERSE SIGNALS

To justify the private belief structure of §2, let (Σ,μ) be an arbitrary probability measure space. Partition the signal measure as $\mu = (\mu^H + \mu^L)/2$, the average of two state-specific signal measures. Conditional on the state, signals $\{\sigma_n\}$ are i.i.d., and drawn according to the probability measure μ^s in state $s \in \{H, L\}$. Assume μ^H, μ^L are mutually absolutely continuous, so that there is a Radon-Nikodym derivative $g = d\mu^L/d\mu^H : \Sigma \to (0,\infty)$. Given signal $\sigma \in \Sigma$, Bayes' rule yields the private belief $p(\sigma) = 1/(g(\sigma)+1) \in (0,1)$ that the state is H. The private belief c.d.f. F^s is the distribution of $p(\sigma)$ under the measure μ^s. Then F^H and F^L are mutually absolutely continuous, as μ^H and μ^L are.

Lemma A.1 *Consider the c.d.f.'s F^H, F^L resulting from (Σ,μ) and a fair prior on H, L.*
(a) The derivative $f \equiv dF^L/dF^H$ of private belief c.d.f.s F^H, F^L satisfies $f(p) = (1-p)/p$ almost surely in $p \in (0,1)$. Conversely, if $f(p) = (1-p)/p$ then F^H, F^L arise from updating a common prior with some signal measures μ^H, μ^L.
(b) $F^L(p) - F^H(p)$ is nondecreasing or nonincreasing as $p \lesseqgtr 1/2$.

(*c*) $F^H(p) < F^L(p)$ *except when both terms are* 0 *or* 1.

(*d*) $F^L(p)/F^H(p) \geq (1-p)/p$ *for* $p \in (0,1)$, *strictly so if* $F^H(p) > 0$.

(*e*) *The likelihood ratio* F^H/F^L *is weakly increasing, and strictly so on* $\text{supp}(F)$.

Proof: If the individual further updates his private belief p by asking of its likelihood in the two states of the world, he must learn nothing more. So, $p = 1/[1+f(p)]$, as desired. Conversely, given $f(p) = (1-p)/p$, let σ have distribution F^s in state s, $s \in \{H, L\}$.

Part (a) implies part (b), and is strict when $p \in \text{supp}(F) \setminus \{1/2\}$. Hence (c) follows. Since $f = dF^L/dF^H$ is a strictly decreasing function, we can prove part (*d*):

$$F^L(p) = \int_{r \leq p} f(r)\, dF^H(r) > f(p) \int_{r \leq p} dF^H(r) = F^H(p) f(p) = F^H(p)(1-p)/p \tag{9}$$

for any p with $F^H(p) > 0$. For (*e*), whenever $F^L(p) > F^L(q) > 0$, (9) implies

$$F^L(p) - F^L(q) = \int_q^p f(r)\, dF^H(r) < [F^H(p) - F^H(q)] f(q) < [F^H(p) - F^H(q)] F^L(q)/F^H(q)$$

Expanding the RHS above, it immediately follows that $F^H(p)/F^L(p) > F^H(q)/F^L(q)$. $\diamondsuit$

B. FIXED POINTS OF MARKOV-MARTINGALE SYSTEMS

This appendix establishes results needed to understand limiting behavior of the Markov-martingale process of beliefs and actions. Despite a countable state space, standard convergence results for discrete Markov chains have no bite, as states are in general transitory.

Given is a finite set $\mathcal{M}$, and Borel measurable functions $\varphi(\cdot\,,\cdot) : \mathcal{M} \times \mathbb{R}_+ \to \mathbb{R}_+$, and $\psi(\cdot\,|\,\cdot) : \mathcal{M} \times \mathbb{R}_+ \to [0,1]$ satisfying:

- $\psi(\cdot\,|\,x)$ is a probability measure on $\mathcal{M}$ for all $x \in \mathbb{R}_+$, or $\sum_{m \in \mathcal{M}} \psi(m|x) = 1$.
- ϕ and ψ jointly satisfy the following 'martingale property' for all $x \in \mathbb{R}_+$:

$$\textstyle\sum_{m \in \mathcal{M}} \psi(m|x)\ \varphi(m,x) = x \tag{10}$$

For any set B in the Borel σ-algebra $\mathcal{B}$ on $\mathbb{R}_+ = [0,\infty)$, define a transition probability $P : \mathbb{R}_+ \times \mathcal{B} \to [0,1]$:

$$P(x,B) = \textstyle\sum_{m\,|\,\varphi(m,x) \in B} \psi(m|x)$$

Let $\langle x_n \rangle_{n=1}^{\infty}$ be a Markov process with transition from $x_n \mapsto x_{n+1}$ governed by P, and $Ex_1 < \infty$. Then $\langle x_n \rangle$ is a martingale, true to the above casual label of (10):

$$E[x_{n+1}|x_1,\ldots,x_n] = E[x_{n+1}|x_n] = \int_{\mathbb{R}_+} tP(x_n,dt) = \textstyle\sum_{m \in \mathcal{M}} \psi(m|x_n)\varphi(m,x_n) = x_n$$

By the Martingale Convergence Theorem, there exists a real, non-negative stochastic variable x_∞ such that $x_n \to x_\infty$ a.s. Since $\langle x_n \rangle$ is a Markov chain, the distribution of x_∞ is intuitively invariant for the transition P, as in Futia (1982). The a.s. convergence then suggests that the invariant limit must be pointwise invariant. While Theorem B.2 below can be proved along these lines, some continuity assumptions will be used. Doing away with continuity, we establish an even stronger result.

Theorem B.1 *Assume that the open interval $I \subseteq \mathbb{R}_+$ satisfies*

$$\exists \varepsilon > 0 \; \forall x \in I \; \exists m \in \mathcal{M} : \; \psi(m|x) > \varepsilon \text{ and } |\varphi(m,x) - x| > \varepsilon \tag{11}$$

Then I cannot contain any point from the support of the limit x_∞.

Proof: Let I be an open interval satisfying (11) for $\varepsilon > 0$, and suppose for a contradiction that there exists $\bar{x} \in I \cap \operatorname{supp}(x_\infty)$. Let $J = (\bar{x} - \varepsilon/2, \bar{x} + \varepsilon/2) \cap I$. By (11), for all $x \in J$, there exists $m \in \mathcal{M}$ with $\psi(m|x) > \varepsilon$ and $\varphi(m,x) \notin J$. Since $\bar{x} \in \operatorname{supp}(x_\infty)$, $x_n \in J$ eventually with positive probability. But whenever $x_n \in J$, $x_{n+1} \notin J$ with chance at least ε. That is, the conditional chance that the process stays in J in the next period is at most $1 - \varepsilon$. So the process $\langle x_n \rangle$ almost surely eventually exits J. This contradicts the claim that with positive chance $\langle x_n \rangle$ is eventually in J. Hence, $\bar{x}$ cannot exist. ♢

Theorem B.2 *If $x \mapsto \varphi(m,x)$ and $x \mapsto \psi(m|x)$ are continuous for all $m \in \mathcal{M}$, then for all $\bar{x} \in \operatorname{supp}(x_\infty)$, stationarity $P(\bar{x}, \{\bar{x}\}) = 1$ obtains, i.e.*

$$\psi(m|\bar{x}) = 0 \quad \textit{or} \quad \varphi(m,\bar{x}) = \bar{x} \qquad \textit{for all } m \in \mathcal{M} \tag{12}$$

Proof: If there is an m such that $\bar{x}$ does not satisfy (12), and both $x \mapsto \varphi(m, x)$ and $x \mapsto \psi(m|\, x)$ are continuous, then there is an open interval I around $\bar{x}$ in which $\psi(m|x)$ and $\varphi(m,x) - x$ are both bounded away from 0. This implies that (11) obtains, and so Theorem B.1 yields an immediate contradiction. ♢

C. STABLE STOCHASTIC DIFFERENCE EQUATIONS

This appendix derives results on the local stability of nonlinear stochastic difference equations. There is a very abstract related literature (see Kifer (1986)), but appendix 1 of Ellison and Fudenberg (1995) treats a model closer to ours. We generalize their stability result to cover state-dependent transition chances (below, ψ may depend on x), and multi-dimensional states (x is $(S-1)$-dimensional), and analyze convergence rates.

Given is a finite set $\mathcal{M} = \{1, \ldots, M\}$, and Borel measurable functions $\varphi(\cdot\,,\cdot) : \mathcal{M} \times \mathbb{R}_+ \to \mathbb{R}$, and $\psi(\cdot\,|\,\cdot) : \mathcal{M} \times \mathbb{R}_+ \to [0,1]$ satisfying $\sum_{m \in \mathcal{M}} \psi(m|x) = 1$. Let $x_0 \in \mathbb{R}$. Then

(7) defines a Markov process $\langle x_n \rangle$. This can be recast as: Let $\langle \sigma_n \rangle$ be a sequence of i.i.d. uniform-$(0,1)$ random variables. Let $\langle y_n \rangle$ with values in $\mathcal{M}$ be defined by $y_n = m$ when $\sigma_n \in (\sum_{i=1}^{m-1} \psi(i, x_{n-1}), \sum_{i=1}^{m} \psi(i, x_{n-1})]$. Then $x_n \equiv \varphi(y_n, x_{n-1})$.

Stability of Linear Equations. Consider this special case of (7): $\psi(m|x_n) = p_m$ and $\varphi(m, x_n) = a_m x_n$. Here $a_1, \ldots, a_M \in \mathbb{R}$ and $p_1, \ldots, p_M \in [0,1]$ satisfy $\sum_{m \in \mathcal{M}} p_m = 1$.

Lemma C.1 *Define* $\bar{\theta} = \prod_{m=1}^{M} |a_m|^{p_m}$.
(a) Almost surely, $\theta^{-n} x_n \to 0$ *for all* $\theta > \bar{\theta}$. *In particular,* $x_n \to 0$ *almost surely if* $\bar{\theta} < 1$.
(b) If $\bar{\theta} < 1$ *and* $\mathcal{N}_0$ *is any open ball around 0, then there is a positive probability that* $x_n \in \mathcal{N}_0$ *for all* n*, provided* $x_0 \in \mathcal{N}_0$.
(c) With $\theta > \bar{\theta}$ *and* $\mathcal{N}_0$ *an open ball around* 0, $\Pr(\forall n \in \mathbb{N} : \ \theta^{-n} x_n \in \mathcal{N}_0 \mid x_0 \in \mathcal{N}_0) > 0$.

Proof: (*a*) Let $Y_n^m \equiv \sum_{k=1}^{n} 1_{\{y_k = m\}}$, so $|x_n| = \left(\prod_{m=1}^{M} |a_m|^{\frac{Y_n^m}{n}} \right)^n |x_0|$. Since $Y_n^m / n \to p_m$ a.s. by the Strong Law of Large Numbers, the result follows from $\prod_{m=1}^{M} |a_m|^{\frac{Y_n^m}{n}} \to \bar{\theta}$ a.s.

(*b*) Since $x_n \to 0$ a.s., $\Pr\left(\bigcup_{k \in \mathbb{N}} \bigcap_{n \geq k} \{x_n \in \mathcal{N}_0\} \right) = 1$. So $\Pr\left(\forall n \geq k, x_n \in \mathcal{N}_0\right) > 0$ for some k. So with positive chance, $\langle x_n \rangle$ stays inside $\mathcal{N}_0$ starting at that x_k. WLOG $k = 0$ since dynamics are time invariant. With linear dynamics, any $x_0 \in \mathcal{N}_0$ will do.

(*c*) Use the result in (*b*) on the modification of (7) with constants a_m / θ. ◇

Local Stability of Nonlinear Equations. We care about the *fixed points* $\hat{x}$ of (7): namely, where $\varphi(m, \hat{x}) = \hat{x}$ for all $m \in \mathcal{M}$.

Theorem C.1 *At a fixed point* $\hat{x}$ *of* (7)*, assume that each* $\psi(m|\cdot) > 0$ *is continuous and* $\varphi(m, \cdot)$ *has a Lipschitz constant*[2] L_m. *If the stability criterion* $\bar{\theta} \equiv \prod_{m=1}^{M} L_m^{\psi(m|\hat{x})} < 1$ *holds, then for all* $\theta \in (\bar{\theta}, 1)$ *there exists an open ball* $\mathcal{N}_0$ *around* 0*, such that* $x_0 - \hat{x} \in \mathcal{N}_0 \Rightarrow$ $\Pr(x_n \to \hat{x}) \geq \Pr\left(\forall n \in \mathbb{N} : \ \theta^{-n} |x_n - \hat{x}| \in \mathcal{N}_0\right) > 0$. *If* $x_n \to \hat{x}$*, then it converges at rate* $\bar{\theta}$.

Proof: First, we majorize (7) locally around $\hat{x}$ by a linear system, and then argue that Lemma C.1's conclusion applies to our original non-linear system.

WLOG, assume that $0 \leq L_1 \leq \cdots \leq L_M$. By continuity of $\psi(m, \cdot)$ we may choose $\mathcal{N}_0$ small enough and constants p_m close enough to $\psi(m|\hat{x})$ so that,

$$\prod_{m=1}^{M} L_m^{p_m} < \theta, \ \sum_{i=1}^{m} \psi(i|x) \geq \sum_{i=1}^{m} p_i, \text{ and } |\varphi(m, x) - \varphi(m, \hat{x})| \leq L_m |x - \hat{x}|, \ \forall \ m \in \mathcal{M}$$

for all $x - \hat{x} \in \mathcal{N}_0$. Fix x_0 with $x_0 - \hat{x} \in \mathcal{N}_0$. Define a new process $\langle \tilde{x}_n \rangle$ with $\tilde{x}_0 = x_0$ given, and $\tilde{x}_n - \hat{x} = L_m(\tilde{x}_{n-1} - \hat{x})$ when $\sigma_n \in (\sum_{i=1}^{m-1} p_i, \sum_{i=1}^{m} p_i]$ where $\langle \sigma_n \rangle$ is our earlier i.i.d. uniform sequence. Lemma C.1 then asserts $\theta^{-n}(\tilde{x}_n - \hat{x}) \in \mathcal{N}_0$ for all n with positive chance.

[2] $f : \mathbb{R}^m \to \mathbb{R}^k$ is *Lipschitz* at $\hat{x}$ with *Lipschitz constant* $L \geq 0$ if $\forall x \in \mathcal{N}(\hat{x}) : \ \|f(x) - f(\hat{x})\| \leq L\|x - \hat{x}\|$ for some *neighborhood* $\mathcal{N}(\hat{x})$. If f is C^1 at $\hat{x}$, then it is Lipschitz with any constant $L > \|Df(\hat{x})\|$.

In any realization of $\langle \sigma_n \rangle$ yielding $\theta^{-n}(\tilde{x}_n - \hat{x}) \in \mathcal{N}_0$ for all n, the resulting deterministic linear process $\langle \tilde{x}_n \rangle$ majorizes the non-linear one $\langle x_n \rangle$: On $\mathcal{N}_0$ we have inductively in n,

$$\sigma_n \in \left(\sum_{i=1}^{m-1} p_i, \sum_{i=1}^{m} p_i \right] \Rightarrow |\tilde{x}_n - \hat{x}| = L_m |\tilde{x}_{n-1} - \hat{x}| \geq L_m |x_{n-1} - \hat{x}| \geq |x_n - \hat{x}|$$

So $\theta^{-n}(x_n - \hat{x}) \to 0$, and with positive probability.

Finally, the rate of convergence is $\bar{\theta}$, since for any $\theta > \bar{\theta}$, a small enough neighborhood exists for which the linear system converges at a rate less than θ. Whenever $x_n \to \hat{x}$, $\langle x_n \rangle$ eventually stays in that neighborhood, wherein it is dominated by $\langle \tilde{x}_n \rangle$. $\diamondsuit$

Corollary C.1 *If each $\varphi(m, \cdot)$ is also continuously differentiable, then Theorem C.1 is true with $\bar{\theta} = \prod_{m \in \mathcal{M}} |\varphi_x(m, \hat{x})|^{\psi(m|\hat{x})} < 1$.*

We must generalize Theorem C.1 for $x_n \in \mathbb{R}^{S-1}$, $S > 2$. We focus on the Markov-martingale context relevant for our model, and restrict attention to $M = 2$. A weaker, more generally applicable, result could be obtained using the sup-norm of matrices; see our MIT working paper.

Theorem C.2 *Let $\hat{x}$ be a fixed point of* (7) *in $\mathbb{R}^{S-1}$ with $M = 2$. Assume that each $\psi(m|\cdot)$ is continuous at $\hat{x}$, with $0 < \psi(1|\hat{x}) < 1$, and that each $\varphi(m, \cdot)$ is C^1 at $\hat{x}$. Assume that each $D_x\varphi(m, \hat{x})$ has distinct, real, positive, non-unit eigenvalues, and that $\psi(1|\hat{x})D_x\varphi(1, \hat{x}) + \psi(2|\hat{x})D_x\varphi(2, \hat{x}) = I$. For any open ball $\mathcal{O} \ni \hat{x}$ there exists $\theta < 1$ and an open ball $\mathcal{N} \subseteq \mathcal{O}$ around $\hat{x}$, such that $x_0 \in \mathcal{N} \Rightarrow$* $\Pr(\theta^{-n}||x_n - \hat{x}|| \to 0) \geq \Pr(\forall n \in \mathbb{N} : \ x_n \in \mathcal{O}) > 0$.

Proof: The proof directly extends the methods used in the uni-dimensional case, by considering $A = D_x\varphi(1, \hat{x})$, $B = D_x\varphi(2, \hat{x})$, and $\psi = \psi(1|\hat{x})$. Thus, $\psi A + (1 - \psi)B = I$.

First, by basic linear algebra, if A has distinct real eigenvalues, then it can be diagonalized as $J_A = QAQ^{-1}$, where Q is an invertible matrix. Likewise, because $J_B = QBQ^{-1} = Q(I - \psi A)Q^{-1}/(1 - \psi) = (I - \psi J_A)/(1 - \psi)$, the matrix Q also diagonalizes B. Rearranging terms, $\psi J_A + (1 - \psi)J_B = I$. Since all eigenvalues are positive and not one, $J_A^{\psi} J_B^{1-\psi}$ has all diagonal entries inside $(0, 1)$, by the earlier scalar AM-GM inequality. Let $\bar{\theta} < 1$ be the maximal element in $J_A^{\psi} J_B^{1-\psi}$. Put $J_C = \max\{J_A, J_B\}$ (componentwise), and $C = Q^{-1}J_C Q$.

On $\mathbb{R}^{S-1}$, we use the inner product $\langle x, y \rangle = x'Q'Qy$ and the norm $\|x\|^2 = \langle x, x \rangle$. For this yields $\langle Ax, Ax \rangle = x'A'Q'QAx = x'Q'J_A'J_AQx \leq x'Q'J_C'J_CQx = \langle Cx, Cx \rangle$, and so $\|Ax\| \leq \|Cx\|$. By continuous differentiability of φ, for any $\delta > 0$, there exists an open ball $\mathcal{N}_1(\delta)$ around $\hat{x}$ such that $x \in \mathcal{N}_1(\delta)$ implies $\varphi(1, x) - \hat{x} = A(x - \hat{x}) + \gamma(x - \hat{x})$, where $\|\gamma(x)\| < \delta\|x\|$. If the maximal eigenvalue of A is $\bar{\lambda}$, then we have

$$\|\varphi(1, x) - \hat{x}\|^2 \leq \|A(x - \hat{x})\|^2 + \delta^2\|x - \hat{x}\|^2 + 2\delta\bar{\lambda}\|x - \hat{x}\|^2$$

Since no eigenvalue of A or B is 0, for any $\eta > 1$, there exists an open ball $\mathcal{N}_2(\eta)$ around $\hat{x}$ so small that $x \in \mathcal{N}_2(\eta)$ implies $\|\varphi(1,x)-\hat{x}\| \leq \|\eta A(x-\hat{x})\|$ and $\|\varphi(2,x)-\hat{x}\| \leq \|\eta B(x-\hat{x})\|$. Clearly then, $\|\varphi(i,x) - \hat{x}\| \leq \|\eta C(x - \hat{x})\|$ for both $i = 1,2$, for all $x \in \mathcal{N}_2(\eta)$.

For any $\varepsilon > 0$, there exists an open ball $\mathcal{N}_3(\varepsilon)$ around $\hat{x}$ in which $\psi(1|\cdot) \in (\psi-\varepsilon, \psi+\varepsilon)$. When $\eta - 1 > 0$ and $\varepsilon > 0$ are both small enough, all diagonal entries of the diagonal matrix $(\eta J_A)^{\psi-\varepsilon}(\eta J_B)^{1-\psi-\varepsilon}(\eta J_C)^{2\varepsilon}$ lie inside $(0,1)$. Put $\mathcal{N}_0 = \mathcal{N}_2(\eta) \cap \mathcal{N}_3(\varepsilon) \cap \mathcal{O}$.

Consider now

$$\tilde{x}_n - \hat{x} = \begin{cases} \eta A(\tilde{x}_{n-1} - \hat{x}) & \text{if } \sigma_n \leq \psi - \varepsilon \\ \eta B(\tilde{x}_{n-1} - \hat{x}) & \text{if } \sigma_n > \psi + \varepsilon \\ \eta C(\tilde{x}_{n-1} - \hat{x}) & \text{else} \end{cases}$$

This linear system is stable. Namely, $y_n = Q(\tilde{x}_{n-1} - \hat{x})Q^{-1}$ follows the stochastic difference equation $y_{n+1} = \eta J_A y_n$ or $y_{n+1} = \eta J_B y_n$ or $y_{n+1} = \eta J_C y_n$ with chances $\psi-\varepsilon$, $1-\psi-\varepsilon$, and 2ε. In this diagonal system, each coordinate follows a scalar equation. By Lemma C.1, each individual coordinate converges a.s. upon zero at rate $\eta\bar{\theta}$. The intersection of a finite number of probability one events has probability one, so a.s. $y_n \to 0$. Thus, a.s. $\tilde{x}_n \to \hat{x}$ at rate $\eta\bar{\theta}$. Extending the proof of Lemma C.1-*b*, there exists an open ball $\mathcal{N} \subseteq \mathcal{N}_0$ around zero, such that $x_0 \in \mathcal{N}$ implies $\tilde{x}_n$ remains in $\mathcal{N}_0$ with positive probability.

We have already shown that the linear system $\langle \tilde{x}_n - \hat{x} \rangle$ dominates in norm the non-linear one $\langle x_n - \hat{x} \rangle$ on $\mathcal{N}_0$. Hence, $x_n \to \hat{x}$ with positive probability, and just as in the proof of Theorem C.1, the convergence is at rate $\bar{\theta}$. $\diamondsuit$

D. OMITTED PROOFS

Cascade Set Characterization: Proof of Lemma 2. Since $p_m^t(q)$ is increasing in m by (1), $[p_{m-1}^t(q), p_m^t(q)]$ is an interval for all q. Then J_m^t is the closed interval of all q that fulfill

$$p_{m-1}^t(q) \leq \underline{b} \quad \text{and} \quad p_m^t(q) \geq \bar{b} \tag{13}$$

Interior disjointness is obvious. Next, if $\text{int}\,(J_m^t) \neq \varnothing$ then $F^H(p_{m-1}^t(q)) = 0$ and $F^H(p_m^t(q)) = 1$ for all $q \in \text{int}\,(J_m^t)$. The individual will choose action m a.s., and so no updating occurs; therefore, the continuation belief is a.s. q, as required.

With bounded beliefs, one of the inequalities in (13) holds for some q, but no q might simultaneously satisfy both. As (1) yields $p_0^t(q) \equiv 0$ and $p_{M_t}^t(q) \equiv 1$ for all q, we must have $J_1^t = [0, \underline{q}^t]$ and $J_{M_t}^t = [\bar{q}^t, 1]$, where $p_1^t(\underline{q}^t) \equiv \bar{b}$ and $p_{M_t}^t(\bar{q}^t) \equiv \underline{b}$ define $0 < \underline{q}^t < \bar{q}^t < 1$.

Finally, let $m_2 > m_1$, with $q_1 \in J_{m_1}^t$ and $q_2 \in J_{m_2}^t$. Then $p_{m_2-1}^t(q_1) \geq p_{m_1}^t(q_1) \geq \bar{b} > \underline{b} \geq p_{m_2-1}^t(q_2)$; and so $q_2 > q_1$ because $p_{m_2-1}^t$ is strictly decreasing in q.

With unbounded beliefs, $\underline{b} = 0$ and $\bar{b} = 1$. Hence, $p_{m-1}^t = 0$ and $p_m^t = 1$ for $q \in J_m^t$

by (13). By (1), this only happens for $m = 1$ and $q = 0$, or $m = M_t$ and $q = 1$.

With bounded beliefs, type t takes only two actions with positive chance in a neighborhood of the non-empty cascade set J^t_m. This follows from (13), since all p^t_m are continuous, and $p^t_{m-1}(q) < p^t_m(q)$ for all m and $q \in (0,1)$, absent a weakly dominated action for type t.

Limit Cascades Occur: Proof of Theorem 1 (a). We first proceed here under the simplifying assumption that ψ and φ are continuous in ℓ. By Theorem B.2, stationarity at the point $\hat{\ell}$ yields $\psi(m|\hat{\ell}) = 0$ or $\varphi(m, \hat{\ell}) = \hat{\ell}$. Assume $\hat{\ell}$ meets this criterion, and consider the smallest m such that $\rho(m|\hat{\ell}) > 0$, so $F^H(\bar{p}_{m-1}(\hat{\ell})) = F^L(\bar{p}_{m-1}(\hat{\ell})) = 0$. Then $\varphi(m, \hat{\ell}) = \hat{\ell}$ implies $F^H(\bar{p}_m(\hat{\ell})) = F^L(\bar{p}_m(\hat{\ell})) > 0$. Since $F^H \succ_{FSD} F^L$ by Lemma A.1(c), this equality is only possible if $F^H(\bar{p}_m(\hat{\ell})) = F^L(\bar{p}_m(\hat{\ell})) = 1$. Thus, $\hat{\ell} \in \bar{J}_m$, as required.

Next abandon continuity. Suppose by way of contradiction that there exist a point $\hat{\ell} \in \text{supp}(\ell_\infty)$ with $\hat{\ell} \notin \bar{J}$. Then for some m we have $0 < F^H(\bar{p}_m(\hat{\ell})-) < 1$, so that individuals will strictly prefer to choose action m for some private beliefs and $m+1$ for others. Consequently, $\bar{p}_m(\hat{\ell}) > \underline{b}$, and since $\bar{p}_0(\hat{\ell}) = 0 \leq \underline{b}$, the least such m satisfying $\bar{p}_m(\hat{\ell}) > \underline{b}$ is well-defined. So we may assume $F^H(\bar{p}_{m-1}(\hat{\ell})-) = 0$.

Next, $F^H(\bar{p}_m(\ell)) > 0$ in a neighborhood of $\hat{\ell}$. There are two possibilities:

Case 1. $F^H(\bar{p}_m(\hat{\ell})) > F^H(\bar{p}_{m-1}(\hat{\ell}))$.

Here, there will be a neighborhood around $\hat{\ell}$ where $F^H(\bar{p}_m(\ell)) - F^H(\bar{p}_{m-1}(\ell)) > \varepsilon$ for some $\varepsilon > 0$. From (3), $\psi(m|\ell) = \psi(m|H, \ell)$ is bounded away from 0 in this neighborhood, while (5) reduces to $\varphi(m, \ell) = \ell F^L(\bar{p}_m(\ell))/F^H(\bar{p}_m(\ell))$, which is also bounded away from $\hat{\ell}$ for ℓ near $\hat{\ell}$. Indeed, $\bar{p}_m(\hat{\ell})$ is in the interior of co(supp(F)), and so Lemma A.1 guarantees us that $F^L(\bar{p}_m(\ell))$ exceeds and is bounded away from $F^H(\bar{p}_m(\ell))$ for ℓ near $\hat{\ell}$ (recall that $\bar{p}_m$ is continuous). By Theorem B.1, $\hat{\ell} \in \text{supp}(\ell_\infty)$ therefore cannot occur.

Case 2. $F^H(\bar{p}_m(\hat{\ell})) = F^H(\bar{p}_{m-1}(\hat{\ell}))$.

This can only occur if F^H has an atom at $\bar{p}_{m-1}(\hat{\ell}) = \underline{b}$, and places no weight on $(\underline{b}, \bar{p}_m(\hat{\ell})]$. It follows from $F^H(\bar{p}_{m-1}(\hat{\ell})-) = 0$ and $\bar{p}_{m-2} < \bar{p}_{m-1}$, that $F^H(\bar{p}_{m-2}(\ell)) = 0$ for all ℓ in a neighborhood of $\hat{\ell}$. Therefore, $\psi(m-1|\ell)$ and $\varphi(m-1, \ell) - \ell$ are bounded away from 0 on an interval $[\hat{\ell}, \hat{\ell} + \eta)$, for some $\eta > 0$. On the other hand, the choice of m ensures that $\psi(m|\ell)$ and $\varphi(m, \ell) - \ell$ are boundedly positive on an interval $(\hat{\ell} - \eta', \hat{\ell}]$, for some $\eta' > 0$. So once again Theorem B.1 (observe the order of the quantifiers!) proves that $\hat{\ell} \notin \text{supp}(\ell_\infty)$.

Confounding Outcomes are Nondegenerate: Rest of Proof of Theorem 2(g). Let types (U, i) prefer action 1 to 2 in state H, and types (V, j) prefer action 1 to 2 in state L. By a rescaling, we may assume that the payoff vector of type (U, i) is (b^U_i, c^U_i) in state H and $(0, 1)$ in state L, with $b^U_i > c^U_i$; type (V, j) respectively earns $(0, 1)$ and (b^V_j, c^V_j), with $b^V_j > c^V_j$. These engender *posterior* belief thresholds $1/(1 + u_i)$ and $1/(1 + v_j)$, where $u_i = b^U_i - c^U_i$ and $v_j = b^V_j - c^V_j$. As in the example of section 3.1, we have private belief thresholds $\bar{p}_i^U(\ell) = \ell/(u_i + \ell)$ and $\bar{p}_j^V(\ell) = \ell/(v_j + \ell)$.

Case 1. Bounded beliefs.

Assume $T = 2$. Generically, $u \neq v$, so assume WLOG $u < v$. The cascade set for type t is $\bar{J}^t = [0, \underline{\ell}^t] \cup [\bar{\ell}^t, \infty]$. With $u < v$ and yet a common likelihood interval of activity, we have $\underline{\ell}^U < \underline{\ell}^V < \bar{\ell}^U < \bar{\ell}^V$. For $\ell \in [\underline{\ell}^U, \underline{\ell}^V)$ only type U is active. Thus $\psi(1|L,\ell)$ rises above $\psi(1|H,\ell)$ in this interval, so $\psi(1|L,\underline{\ell}^V) > \psi(1|H,\underline{\ell}^V)$. Near $\bar{\ell}^V$, only type V is active, so $\psi(1|L,\bar{\ell}^U) < \psi(1|H,\bar{\ell}^U)$. By continuity, $\psi(1|L,\ell^*) = \psi(1|H,\ell^*)$ for some $\ell^* \in (\underline{\ell}^V, \bar{\ell}^U)$.

Case 2. Unbounded beliefs.

Let λ_i^U and λ_j^V denote the population weights.

$$\psi(1|s,\ell) = \kappa_1 + \kappa \textstyle\sum_i \lambda_i^U F^s(\ell/(u_i + \ell)) + \kappa \sum_j \lambda_j^V [1 - F^s(\ell/(v_j + \ell))]$$
$$\Rightarrow \psi_\ell(1|s,0) = \kappa f^s(0) \left(\textstyle\sum_i \lambda_i^U u_i/(u_i + 0)^2 - \sum_j \lambda_j^V v_j/(v_j + 0)^2 \right)$$

Since $f^L(0) > f^H(0)$ by Lemma A.1, $\psi_\ell(1|H,0) < \psi_\ell(1|L,0)$ when $\sum_i \lambda_i^U/u_i > \sum_j \lambda_j^V/v_j$. Since $f^L(1) < f^H(1)$, $\psi_\ell(1|H,\infty) > \psi_\ell(1|L,\infty)$ likewise ensues from $\sum_i \lambda_i^U u_i < \sum_j \lambda_j^V v_j$. Finally, both inequalities hold (or both fail) for sufficiently different (and opposed) preferences — namely u_i small enough and/or v_i big enough, in the case above; we get the reverse inequality in each case for u_i big enough and/or v_i small enough.

Proof of Overturning Principle used in Theorem 3. If n optimally takes m, his belief p_n satisfies

$$\frac{1 - r_{m-1}}{r_{m-1}} > \ell(h)\frac{1 - p_n}{p_n} \geq \frac{1 - r_m}{r_m} \tag{14}$$

Let $\Sigma(h)$ denote the set of all beliefs p_n that satisfy (14). Then individual n chooses action m with probability $\int_{\Sigma(h)} dF^H$ (resp. $\int_{\Sigma(h)} f\, dF^H$) in state H (resp. state L). This yields the continuation

$$\ell(h,m) = \ell(h)\frac{\int_{\Sigma(h)} f\, dF^H}{\int_{\Sigma(h)} dF^H}$$

Cross-multiply and use (14) with Lemma A.1(a) to bound the right hand integral.

Stability of Markov-Martingale Processes: Rest of Proof of Theorem 4. Next suppose that for m in some subset $\mathcal{M}_0 \subseteq \mathcal{M}$, $\varphi(m,\bar{x}) \neq \bar{x}$ and thus $\psi(m,\bar{x}) = 0$. Then $\mathcal{M} \setminus \mathcal{M}_0 \neq \varnothing$ since $\sum_m \psi(m|\cdot) = 1$. Note that $\theta \equiv \prod_{m \notin \mathcal{M}_0} |\varphi_x(m,\bar{x})|^{\psi(m|\bar{x})} < 1$. Choose $\hat{\theta} \in (\theta, 1)$. By Corollary C.1, there is an open ball $\mathcal{N}_0$ around 0 such that $x_0 - \bar{x} \in \mathcal{N}_0 \Rightarrow \Pr(\forall n \in \mathbb{N} : \hat{\theta}^{-n}|x_n - \bar{x}| \in \mathcal{N}_0) > 0$ when only actions in $\mathcal{M} \backslash \mathcal{M}_0$ are taken.

Define events $E_n = \{m_n \notin \mathcal{M}_0\}$, $F_n = \{\hat{\theta}^{-n}|x_n - \bar{x}| < |x_0 - \bar{x}|\}$, $G_n = \cap_{k=0}^n (E_k \cap F_k)$ and $G_\infty = \cap_{k=0}^\infty (E_k \cap F_k)$. Then,

$$\Pr(G_\infty) = \Pr(G_0) \textstyle\prod_{n=0}^\infty \Pr(G_{n+1}|G_n) = \Pr(G_0) \prod_{n=0}^\infty \Pr(F_{n+1}|E_{n+1}, G_n) \Pr(E_{n+1}|G_n)$$

Corollary C.1 implies $0 < \prod_{n=0}^{\infty} \Pr(F_{n+1}|E_{n+1}, G_n)$. When $x_n \to \bar{x}$ exponentially fast and ψ is C^1, the sequence $\sum_{m \in \mathcal{M}_0} \psi(m|x_n)$ vanishes exponentially fast; therefore $0 < \prod_{n=1}^{\infty} \Pr(E_{n+1}|G_n)$. Collecting pieces, $0 < \Pr(G_\infty) \leq \Pr(x_n \text{ converges to } \bar{x} \text{ at rate } \theta)$.

Herds and Confounded Learning Arise: Proof of Theorem 5.

Part (a): Let ℓ_0 be close to, but just below, the fixed point $\hat{\ell} = \inf \hat{J}$. Generically, only one rational type t is active in a neighborhood of $\hat{\ell}$. The other rational types are then equivalent to noise in this neighborhood, and we continue WLOG as if there is only a single rational type. Let m denote the action which this rational type takes with probability one on $\hat{J}$. Since private beliefs have C^1 tails, $\rho(m|\cdot)$, $\psi(m|\cdot)$, and $\varphi(m, \cdot)$ are C^1 near $\hat{\ell}$.

Suppose that $\varphi_\ell(m, \hat{\ell}) < 0$. As $\ell \uparrow \hat{\ell}$, both C^1 functions $\rho(m|H, \ell), \rho(m|L, \ell) \to 1$, and thus $\varphi(m, \ell) - \ell \to 0$. Then $\ell_{n+1} \in \hat{J}$ for ℓ_n close enough to $\hat{\ell}$. So after finitely many steps, and thus with a positive probability, $\langle \ell_n \rangle$ jumps into $\hat{J}$, and convergence is at rate $\theta = 0$.

Next posit $\varphi_\ell(m, \hat{\ell}) \geq 0$. Let $\beta(m, \ell) \equiv \ell \rho(m|L, \ell)/\rho(m|H, \ell)$ be the continuation without noise. Then $\beta_\ell(m, \hat{\ell}) = 1 + \hat{\ell}\,[f^H(\underline{b}) - f^L(\underline{b})] < 1$ by Lemma A.1(a), for bounded and not all uninformative beliefs ($0 < \underline{b} < 1/2$). So $\varphi_\ell(m, \hat{\ell}) = (\kappa_m + \kappa \beta_\ell(m, \hat{\ell}))/(\kappa_m + \kappa) < 1$, and thus $\theta < 1$ by Theorem 4, provided $\varphi(m', \hat{\ell}) \geq 0$ for all actions m'. Then $\varphi_\ell(m', \ell) = 1$ for ℓ near $\hat{\ell}$ for all actions except $m' = m + 1$, the only other one rationally taken for ℓ close enough to $\bar{J}_m$. But $\beta_\ell(m, \hat{\ell}) < 1$ implies $\beta_\ell(m+1, \hat{\ell}) > 1$, and thus $\varphi_\ell(m+1, \hat{\ell}) > 1$.

Finally, assume $\ell_\infty \in \hat{J}$, and that the final approach was from the left. Eventually, all but type t must be inside their cascade sets, and thus herding. The convergence of ℓ_n to ℓ_∞ is exponentially fast. With $f^H(\underline{b}) > 0$, $\rho^t(m|H, \ell_n)$ then converges exponentially fast to one. By the first Borel-Cantelli lemma, eventually type t takes action m.

Part (b): By Fatou's lemma, $E[\ell_\infty] \leq \ell_0$. Let $\ell^* = \inf \bar{K} > 0$. Since the limit is concentrated on $\{0\} \cup \bar{K}$, it must be 0 with probability at least $(\ell^* - \ell_0)/\ell^*$.

Part (c): We only need robustness to noise since the example establishes the noiseless case. Let $\tilde{\varphi}$, $\tilde{\psi}$ denote the analogues of φ, ψ when noise is added. Then $\tilde{\psi}(m|H, \ell^*) = \kappa_m + \kappa \psi(m|H, \ell^*)$. Differentiating $\tilde{\varphi}(m, \ell) = \ell \tilde{\psi}(m|L, \ell)/\tilde{\psi}(m|H, \ell)$ then yields

$$\tilde{\varphi}_\ell(m, \ell^*) = 1 + \ell^* \frac{\tilde{\psi}_\ell(m|L, \ell^*) - \tilde{\psi}_\ell(m|H, \ell^*)}{\tilde{\psi}(m|H, \ell^*)} = 1 + \kappa \ell^* \frac{\psi_\ell(m|L, \ell^*) - \psi_\ell(m|H, \ell^*)}{\kappa \psi(m|H, \ell^*) + \kappa_m}$$

If $\psi(m|H, \ell^*) > 0$, then $\tilde{\varphi}_\ell$ approaches φ_ℓ as $\kappa_m \to 0$. If $\psi(m|H, \ell^*) = 0$, then φ_ℓ does not affect the stability criterion for ℓ^*, but does when $\kappa_m > 0$. For any $\varepsilon > 0$, $\tilde{\varphi}_\ell(m, \ell^*) < (1 + \varepsilon)/\tilde{\psi}(m|H, \ell^*)$ for small enough noise, since $\lim_{\kappa_m \to 0} \tilde{\psi}(m|H, \ell^*) \tilde{\varphi}_\ell(m, \ell^*) = 1$. As $\kappa_m \to 0$, $\tilde{\psi}(m|H, \ell^*)$ tends to $\psi(m|H, \ell^*) = 0$, and so for all $\varepsilon > 0$, eventually $\tilde{\varphi}_\ell(m, \ell^*)^{\tilde{\psi}(m|H, \ell^*)} < ((1+\varepsilon)/\tilde{\psi}(m|H, \ell^*))^{\tilde{\psi}(m|H, \ell^*)} < 1 + \varepsilon$. So if $\Pi_m \varphi_\ell(m, \ell^*)^{\psi(m|H, \ell^*)} < 1$, then ℓ^* remains stable for small enough κ_m.

References

Aghion, P., P. Bolton, C. Harris, and B. Jullien (1991): "Optimal Learning by Experimentation," *Review of Economic Studies*, 58(196), 621–654.

Banerjee, A. V. (1992): "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, 107, 797–817.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992): "A Theory of Fads, Fashion, Custom, and Cultural Change as Information Cascades," *Journal of Political Economy*, 100, 992–1026.

Bray, M., and D. Kreps (1987): "Rational Learning and Rational Expectations," in *Kenneth Arrow and the Ascent of Economic Theory*, ed. by G. Feiwel, pp. 597–625. Macmillan, London.

Breiman, L. (1968): *Probability*. Addison-Wesley, Reading, Mass.

Doob, J. L. (1953): *Stochastic Processes*. Wiley, New York.

Ellison, G., and D. Fudenberg (1995): "Word of Mouth Communication and Social Learning," *Quarterly Journal of Economics*, 110, 93–125.

Futia, C. A. (1982): "Invariant Distributions and the Limiting Behavior of Markovian Economic Models," *Econometrica*, 50, 377–408.

Kifer, Y. (1986): *Ergodic Theory Of Random Transformations*. Birkhauser, Boston.

McLennan, A. (1984): "Price Dispersion and Incomplete Learning in the Long Run," *Journal of Economic Dynamics and Control*, 7, 331–347.

Milgrom, P. (1979): "A Convergence Theorem for Competitive Bidding," *Econometrica*, 47, 670–688.

Piketty, T. (1995): "Social Mobility and Redistributive Politics," *Quarterly Journal of Economics*, 110, 551–584.

Smith, L., and P. Sørensen (1998): "Informational Herding and Optimal Experimentation," MIT Working Paper.