

Observational Learning

Lones Smith and Peter Norman Sørensen

Abstract

Observational learning occurs when privately-informed individuals sequentially choose among finitely many actions, after seeing predecessors' choices. We summarize the general theory of this paradigm: *Belief convergence* forces *action convergence*, specifically, copycat “herds” arise. Also, beliefs converge to a point mass on the truth exactly when the private information is not uniformly bounded.

This subsumes two key findings of the original herding literature: With multinomial signals, *cascades* occur, where individuals rationally ignore their private signals, and incorrect herds start with positive probability. The framework is flexible — some individuals may be committed to an action, or individuals may have divergent cardinal or even ordinal preferences.

Keywords

observational learning; social learning; informational herding; Markov process; martingale; information aggregation; stochastic difference equation; informational cascade; limit cascade; action herd; experimentation

Article

Observational Learning. Suppose that an infinite number of individuals each must make an irreversible choice among finitely many actions — 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.

Observational learning occurs specifically when the individuals must decide sequentially, all in some preordained order. 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. This article summarizes

the general framework for the herding model that subsumes all signals, and establishes the correct conclusions. The framework is flexible — eg. some individuals may be committed to an action, or individuals may have divergent preferences.

Banerjee (1992) and Bikhchandani, Hirshleifer, and Welch (1992) (hereafter, BHW) both introduced this framework. Ottaviani and Sørensen (2006) later noted that the same mechanism drives expert herding behavior in the earlier model of Scharfstein and Stein (1990), after dropping their assumption that private signals are conditionally correlated. In BHW's logic, *cascades* eventually start, in which individuals rationally ignore their private signals. Copycat action herds therefore arise *ipso facto*. Also, despite the surfeit of available information, a herd develops on an incorrect action with positive probability: After some point, everyone might just settle on the identical less profitable decision. This result sparked a welcome renaissance in informational economics. Observational learning explains correlation of human behavior in environments without network externalities where one might otherwise expect greater independence. Various twists on the herding phenomenon have been applied in a host of settings from finance to organizational theory, and even lately into experimental and behavioral work.

In this article, we develop and flesh out the general theory of how Bayesian individuals sequentially learn from the actions of posterity, as developed in Smith and Sørensen (2000). Our logical structure is to deduce that almost sure *belief convergence* occurs, which in turn forces *action convergence*, or the action herds. Also, beliefs converge to a point mass on the correct state exactly when the private signal likelihood ratios are not uniformly bounded. For instance, incorrect herds arose in the original herding papers since they assumed finite multinomial signals. We hereby correct a claim by Bikhchandani, Hirshleifer, and Welch (2008), which unfortunately concludes “In other words, in a continuous signals setting herds tend to form in which an individual follows the behaviour of his predecessor with high probability, even though this action is not necessarily correct. Thus, the welfare inefficiencies of the discrete cascades model are also present in continuous settings.”

Multinomial signals also violate a log-concavity condition, and for this reason yield the rather strong form of belief convergence that a cascade is. One recent lesson is the extent to which cascades are the exception rather than rule.

The Model. Assume a completely ordered sequence of individuals $1, 2, \dots$. Each faces an identical binary-choice decision problem, choosing an action $a \in \{1, 2\}$. Individual n 's payoff $u(a_n, \omega)$ depends on the realization of a state of the world, $\omega \in \{H, L\}$, common across n . The high action pays more in the high state: $u(1, L) > u(2, L)$ and $u(1, H) < u(2, H)$. Individuals act as Bayesian expected utility maximizers, choosing action $a = 2$ above a threshold posterior belief \bar{r} , and otherwise action $a = 1$. All share a common prior $q_0 = P(\omega = H)$, and for simplicity, $q_0 = 1/2$.

The decision-making here is partially informed. For exogenous reasons, each individual n privately observes the realization of a noisy signal σ_n , whose distribution depends on the state ω . Conditional on ω , signals are independently and identically distributed. Observational learning is modeled via the assumption that individual i can observe the full history of actions $h_n = (a_1, \dots, a_{n-1})$. While predecessors' private signals cannot be observed directly, they may be partially inferred. The interesting properties of observational learning follow because the private signals are coarsely filtered by coarse public action observations.

The private observation of signal realization σ_n , with no other information, yields an updated *private belief* $p_n \in [0, 1]$ in the state of the world $\omega = H$. The private belief p_n is a sufficient statistic for the private signal σ_n in the n 'th individual's decision problem. Its cumulative distribution $F(p|\omega)$ in state ω is a key primitive of the model. Define the unconditional cumulative distribution $F(p) = [F(p|H) + F(p|L)]/2$. The theory is valid for arbitrary signal distributions, having a combination of discrete and continuous portions. But to simplify the exposition, we assume a continuous distribution, with density f . The state-conditional densities $f(p|\omega)$ obey the Bayesian relation $p = (1/2)f(p|H)/f(p)$ with $f(p) = [f(p|H) + f(p|L)]/2$, implying $f(p|H) = 2pf(p)$ and $f(p|L) = 2(1 - p)f(p)$. The equality $f(p|H)/f(p|L) = p/(1 - p)$ can be usefully reinter-

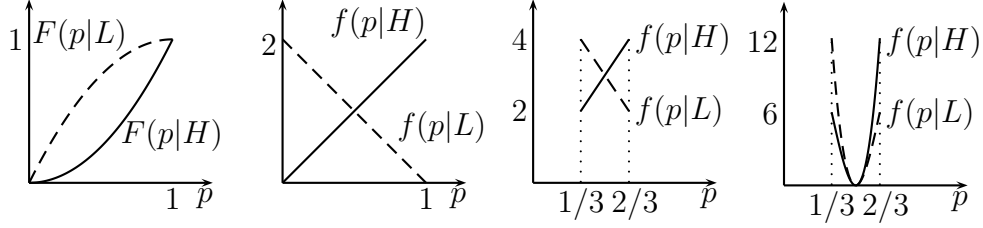


Figure 1: **Private Belief Distributions.** At left are generic private belief distributions in the states L, H , illustrating the stochastic dominance of $F(\cdot|H) \succ F(\cdot|L)$. The three other panels depict the specific densities for the unbounded and bounded private belief signal distributions discussed in the text.

preted as a *no introspection condition*: Understanding the model likelihood ratio of one's private belief p does not allow any further inference. This special ratio ordering implies that the conditional distributions share the same support, but that $F(p|H) < F(p|L)$ for all private beliefs strictly inside the support (Figure 1).

Private beliefs are said to be *bounded* if there exist $p', p'' \in (0, 1)$ with $F(p') = 0$ and $F(p'') = 1$, and *unbounded* if $F(p) \in (0, 1)$ for all $p \in (0, 1)$. For instance, a uniform density $f(p) \equiv 1$ results in the unbounded private belief distributions $F(p|H) = p^2 < 2p - p^2 = F(p|L)$. But if $f(p) \equiv 3$ on the support $[1/3, 2/3]$, then the bounded private belief distributions are $F(p|H) = (3p - 1)(1 + 3p)/3 < (3p - 1)(5 - 3p)/3 = F(p|L)$.

Analysis via Stochastic Processes. Because only the actions are publicly observed with observational learning, the *public belief* q_n in state H is based on the observed history of the first $n - 1$ actions alone. The associated *likelihood ratio* of state L to state H is then $\ell_n = (1 - q_n)/q_n$. And if so desired, we can recover public beliefs from the likelihood ratios using $q_n = 1/(1 + \ell_n)$. Incorporating the most recent private belief p_n yields the posterior belief $r_n = p_n/(p_n + \ell_n(1 - p_n))$ in state H . So indifference prevails at the *private belief threshold* $\bar{p}(\ell)$ defined by

$$\bar{r} \equiv \frac{\bar{p}(\ell)}{\bar{p}(\ell) + \ell(1 - \bar{p}(\ell))} \quad (1)$$

Individual n chooses action $a = 1$ for all private beliefs $p_n \leq \bar{p}(\ell)$, and otherwise

picks $a = 2$. Since higher public beliefs (i.e. lower likelihood ratios) compensate for lower private beliefs in Bayes Rule, the threshold is monotone $\bar{p}'(\ell) > 0$.

We now construct the public stochastic process. Given the likelihood ratio ℓ , action $a = 1, 2$ happens with chance $\rho(a|\ell, \omega)$ in state $\omega \in \{H, L\}$, where

$$\rho(1|\ell, \omega) \equiv F(\bar{p}(\ell)|\omega) \equiv 1 - \rho(2|\ell, \omega) \quad (2)$$

When individual n takes action a_n , the updated public likelihood ratio is

$$\ell_{n+1} = \varphi(a_n, \ell_n) \equiv \ell_n \frac{\rho(a_n|\ell_n, L)}{\rho(a_n|\ell_n, H)} \quad (3)$$

since Bayes' Rule reduces to multiplication in likelihood ratio space due to the conditional independence of private signals. But in light of our stochastic ordering, the binary action choices are informative of the state of the world:

$$\rho(1|\ell_n, L) > \rho(1|\ell_n, H) \quad \text{and} \quad \rho(2|\ell_n, L) < \rho(2|\ell_n, H)$$

Observe what has just happened. Choices have been automated, and what remains is a stochastic process (ℓ_n) that is a *martingale*, conditional on state H .

$$E[\ell_{n+1} | \ell_1, \dots, \ell_n, H] = \sum_m \rho(m|\ell_n, H) \ell_n \frac{\rho(m|\ell_n, L)}{\rho(m|\ell_n, H)} = \ell_n$$

Because the stochastic process (ℓ_n) is a non-negative martingale in state H , the Martingale Convergence Theorem applies. Namely, (ℓ_n) converges almost surely to the (random variable) limit $\ell_\infty = \lim_{n \rightarrow \infty} \ell_n$, namely having (finite) values in $[0, \infty)$. The support of ℓ_∞ contains all candidate limit likelihood ratios. Among the most immediate of implications, *learning cannot result in a fully erroneous belief $\ell = \infty$ with positive probability*. Just as well, this follows from Fatou's Lemma in measure theory, for $E[\liminf_{n \rightarrow \infty} \ell_n | H] \leq \liminf_{n \rightarrow \infty} E[\ell_n | H] = \ell_0$.

Let's continue to trace this logic, by next observing that the sequence of pairs of actions and likelihood ratios (a_n, ℓ_n) is also a *Markov process* on the domain

$\{1, 2\} \times [0, \infty)$. For we can see that each new pair only depends on the last:

$$(a_n, \ell_n) \mapsto (a_{n+1}, \varphi(a_{n+1}, \ell_n)) \quad \text{with chance} \quad \rho(a_{n+1}|\ell_n, H)$$

The big gun for Markov processes is the stationarity condition. While our two-dimensional process (a_n, ℓ_n) is clearly nonstandard, Smith and Sørensen (2000) prove the following version of the Markov stationarity condition: *If the transition functions ρ and φ are continuous in ℓ , then for any $\hat{\ell}$ in the support of ℓ_∞ and for all m , we have either $\rho(m|H, \hat{\ell}) = 0$ or $\varphi(m, \hat{\ell}) = \hat{\ell}$.* In other words, either an action does not occur, or it yields no new information, or both.

The stationary points of the (a_n, ℓ_n) process are therefore the *cascade sets*, namely, those sets of likelihood ratios ℓ indexed by actions m that almost surely repeat action m , namely, $\bar{J}_m = \{\ell \mid \rho(m|\ell, H) = 1\}$. With *bounded private beliefs*, there must exist some high (low) enough likelihood ratios ℓ that pull all private beliefs below (above) the threshold posterior belief \bar{r} . In this case, the cascade sets \bar{J}_1, \bar{J}_2 for the two actions are both non-empty. When private beliefs are unbounded, the cascade sets collapse to the extreme points, $\bar{J}_1 = \{\infty\}$ and $\bar{J}_2 = \{0\}$. And since we have seen that $\ell = \infty$ cannot arise with positive probability, we must converge to a point mass on the truth (or $\ell = 0$).

Next, we claim that convergence of beliefs implies convergence of actions. Whenever someone optimally chooses action m , any successor must optimally follow suit if he bases his decision just on public information. For individual $n - 1$ solves the same decision problem as n faces, but with more information, (a_1, \dots, a_{n-2}) and σ_{n-1} . Contrary actions completely “overturn” the weight of the entire action history, howsoever long. By this *Overturning Principle*, an infinite subsequence of contrary actions precludes belief convergence. By the Martingale Convergence Theorem, this almost surely cannot happen. By the last paragraph, we conclude that *with unbounded private beliefs, a correct herd eventually arises*.

When Only Correct Herds Arise. Consider an illustrative example, with individuals deciding whether to ‘invest’ in or ‘decline’ an investment project of

uncertain value. Investing (action 2) is risky, paying $u > 1$ in state H and -1 in state L ; declining (action 1) is a neutral action with zero payoff in both states. Indifference prevails at the posterior belief $\bar{r} = 1/(1 + u)$. Then equation (1) yields the private belief threshold $\bar{p}(\ell) = \ell/(u + \ell)$.

Assume first the earlier unbounded private beliefs example. Then transition chances are $\rho(1|\ell, H) = \ell^2/(u + \ell)^2$ and $\rho(2|\ell, L) = \ell(\ell + 2u)/(u + \ell)^2$, and continuations

$$\varphi(2|\ell) = \frac{u\ell}{u + 2\ell} < \ell < \ell + 2u \equiv \varphi(1|\ell)$$

by equations (2)–(3). In other words, the likelihood ratio sequence constitutes a stochastic difference equation. Figure 2 shows how $\bar{J}_2 = \{0\}$ is the only stationary finite likelihood ratio in state H : The limit ℓ_∞ is thus concentrated on 0, the truth.

Whenever action 2 is taken, the new likelihood ratio is $\ell_n \geq 2u$. This can only happen finitely many times. So belief convergence implies action convergence, namely, a herd. This example precisely illustrates the logic for one main result: Interestingly, a herd arises despite the fact that a cascade never does, since at each and every stage, a contrary action was possible. Since convergence occurs towards the cascade set but forever lies outside, this is called a *limit cascade*.

When Incorrect Herds Must Sometimes Arise. When private beliefs are bounded, public beliefs still converge, and they result in copycat herds. The main difference now is the positive probability of incorrect herds. Indeed, adjust the last example for the bounded beliefs family. Given the private belief threshold $\bar{p}(\ell) = \ell/(u + \ell)$, the laws of motion (2)–(3) yield transitions

$$\varphi(2|\ell) \equiv \ell \frac{\ell + 4u}{5\ell + 2u} < \ell < \ell \frac{2\ell + 5u}{4\ell + u} \equiv \varphi(1|\ell)$$

with probabilities

$$\rho(1|H, \ell) = \frac{(4\ell + u)(2\ell - u)}{3(u + \ell)^2} \quad \text{and} \quad \rho(2|L, \ell) = \frac{(\ell + 4u)(2u - \ell)}{3(u + \ell)^2}$$

for likelihood ratios $\ell \in (u/2, 2u)$. As seen in Figure 2 (left panel), a cascade can

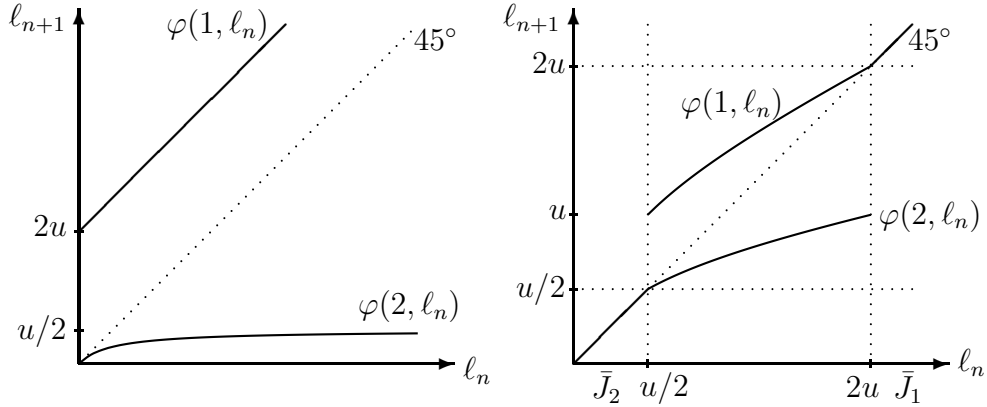


Figure 2: **Transitions and Cascade Sets.** Transition functions for the examples: unbounded private beliefs (left), and bounded private beliefs (right). By the martingale property, the expected continuation in state H lies on the diagonal. The stationary points are where both arms hit the diagonal, or where one arm is taken with zero chance ($\ell = 0$ in the left panel; $\ell \leq 2u/3$ or $\ell \geq 2u$ in the right panel).

never start after the first individual decides. But since the likelihood ratio must converge, a limit cascade starts, towards one of the cascade sets $\bar{\ell}_1$ or $\bar{\ell}_2$. A herd on the corresponding action must then start eventually, lest beliefs fail to converge.

We now explore the easy logic for why *an incorrect herd occurs with strictly positive probability given bounded beliefs*. Again, we appeal to a big gun from measure theory. For if we start at some public likelihood ratio $\ell_0 \in (u/2, 2u)$, then by Figure 2, dynamics are trapped in $(u/2, 2u)$. Since $0 \leq \ell_n \leq 2u$, Lebesgue's Dominated Convergence Theorem allows us to swap the expectation and limit operations, and thus conclude that $E[\ell_\infty | H] = \lim_{n \rightarrow \infty} E[\ell_n | H] = \ell_0$. Write $\ell_0 = \pi(u/2) + (1 - \pi)(2u)$, where $0 < \pi < 1$ whenever $u/2 < \ell_0 < 2u$. Then the random variable ℓ_∞ places weight π on $u/2$ and weight $1 - \pi$ on $2u$. So in state H , a herd arises with chance π on action 2, and with chance $1 - \pi$ on action 1.

Herds Without Cascades. For an interesting contrast to the discrete signal world of BHW, observe that in Figure 2 (right panel), if we do not begin in a cascade, we never enter one — even though a herd eventually starts. Indeed, visually, it is clear that $\ell_n \in (u/2, 2u)$ for all n , provided that initially $\ell_0 \in (u/2, 2u)$. So while the analysis in BHW explicitly depended on cascades ending the dynamics

in finite time, a somewhat subtler dynamic story emerges here: *Herds must arise even though a contrarian has positive probability at every stage.*

This no-cascades result is robust to changes both in the signal distribution and payoffs. For it arises whenever the continuation functions $\varphi(1, \ell)$, $\varphi(2, \ell)$ are monotone increasing in ℓ . Monotonicity asserts the seemingly plausible condition that a higher prior public belief implies a higher posterior public belief after every action. Yet, despite how intuitive this property may seem, it is violated by any multinomial signal distribution (loosely, because it is “lumpy”).

We have shown in Smith and Sørensen (2008) that the continuation functions are monotone under an easily verifiable regularity condition — namely, that the unconditional density of the log-likelihood ratio $\log(p/(1-p))$ be log-concave. Most popular continuous distributions satisfy this condition, for instance, the Gaussian, uniform, or generalized exponential. But the analysis in BHW and a vast number of successor papers was based on the multinomial family — namely, the one main signal family for which the regularity condition *fails*. This discussion hereby corrects the claim by Bikhchandani, Hirshleifer, and Welch (2008), that “In some continuous signal settings cascades do not form (Smith and Sørensen, 2000)”. On the contrary, one really must view cascades as the informationally rare outcome, a case where a tractable example class proved misleading. The true touchstone of this literature is simply the observed phenomenon of action herding.

Cascades with Smooth Signals. To fully flesh out this picture, we offer an example of a continuous signal distribution that violates the monotonicity result. (This example is based on one included in the original working paper of Smith and Sørensen (2000) found in Sørensen (1996)). To this end, we construct a sufficiently heroic violation our log-concavity condition. Suppose that private beliefs p have a quadratic density $f(p) = 324(p - 1/2)^2$ over the bounded support $[1/3, 2/3]$. Then the conditional private belief densities are $f(p|H) = 2pf(p)$ and $f(p|L) = 2(1-p)f(p)$, as depicted in the right panel of Figure 1. Integration yields the (suppressed) polynomial expressions for $F(p|L)$, $F(p|H)$.

Returning to the running investment payoff example, for all likelihood ratios

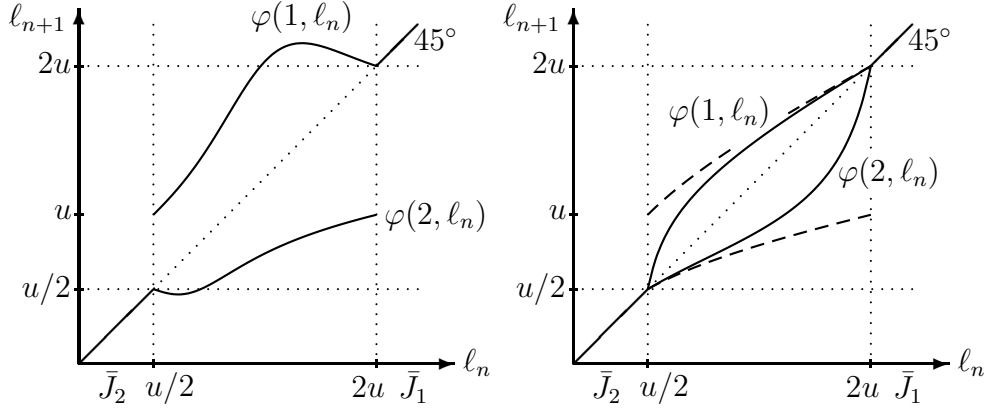


Figure 3: **Modified Transitions.** Transition functions for bounded beliefs with a quadratic density (left panel) and uniform bounded beliefs with and without 20% crazy types (solid and dashed lines in right panel). The non-monotonocities of transition functions (left panel) imply that a cascade on a starts when a is taken where l_n is sufficiently close to \bar{J}_a . The transition function discontinuity vanishes with the addition of crazy types (right panel), corresponding to the failure of the overturning principle.

$\ell \in (u/2, 2u)$, we find the likelihood ratio transitions (left panel of Figure 3):

$$\begin{aligned}\varphi(1, \ell) &= \ell \frac{23(u + \ell)^3 - 93\ell(u + \ell)^2 + 126\ell^2(u + \ell) - 54\ell^3}{3(u + \ell)^3 + 9\ell(u + \ell)^2 - 54\ell^2(u + \ell) + 54\ell^3}, \\ \varphi(2, \ell) &= \ell \frac{12(u + \ell)^3 - 63\ell(u + \ell)^2 + 108\ell^2(u + \ell) - 54\ell^3}{2(u + \ell)^3 + 3\ell(u + \ell)^2 - 36\ell^2(u + \ell) + 54\ell^3}.\end{aligned}$$

A More General Observational Learning Framework. The Overturning Principle may not sound very realistic, a priori. Should we expect that a single deviator from an action herd of one million individuals entirely can by himself change the course of subsequent play? Is the excessive reliance on the assumption of common knowledge of rationality implicit in the overturning principle reasonable? Experimental results on the informational herding model, e.g., Çelen and Kariv (2004), has cast doubt on this. (The review by Anderson and Holt (2008) speaks more broadly to such experimental evidence.)

It turns out that our reduction of the model to a stochastic difference equation in the likelihood ratio obeying a martingale property is robust to a wide array of

economically-inspired modifications that can accommodate deviations from the overturning principle. For instance, suppose that a fraction of ‘crazy’ individuals randomly choose actions. Figure 3 depicts the modified continuation functions in the right panel, for a case where 10% of individuals are committed to action 1 and 10% are committed to action 2. The remaining population is rational. Since all actions occur with a non-vanishing frequency, none can have drastic effects. Yet, the limit beliefs are unaffected by the noise, contrary actions being deemed irrational (and ignored) inside the cascade sets. Of course, the failure of the overturning principle invalidates the argument that limit cascades force herds. But because actions are still informative of beliefs, social learning is productive.

We show more strongly in Smith and Sørensen (2000) that herds nonetheless do arise among all rational (non-crazy) individuals, when beliefs are bounded and have non-zero density near the bounds. Essentially, the public likelihood ratios (ℓ_n) converges so fast that the chance of an infinite string of rational contrarians is zero. (Of course, an outside observer of the action history would hardly be able to detect infrequent rational non-herders, should they occur.)

Alternatively, we may relax the assumption that all individuals solve the same decision problem. Individuals may well have different rational preference types. First, if ordinal preferences are aligned, so that everyone takes action 2 for stronger beliefs in state H , then the limit likelihood ratio ℓ_∞ is focused on the intersection of their respective cascade sets.

Suppose instead that the ordinal preferences differ for some pair of types. Then there arises the possibility of a *confounded learning point*. This is a non-cascade likelihood ratio ℓ^* such that if $\ell_{n-1} = \ell^*$, then individual n 's observation of action a_n is non-informative — the probabilities satisfy $\rho(1|H, \ell^*) = \rho(1|L, \ell^*)$. In this case, $\ell_{n+1} = \ell_n$ following either action of individual n . If such a confounding outcome ℓ^* exists, then it is *locally stochastically stable*: there is positive probability that $\ell_\infty = \ell^*$ provided some ℓ_n is ever sufficiently close to ℓ^* .

Conclusion. This model of observational learning explores a modeling framework to analyze imitation of observed behavior. The model is quite tractable.

Public beliefs based on the ever lengthening action history must converge to a limit, which is among the fixed points of a stochastic difference equation. As long as all ordinal preferences coincide, we eventually settle on an action herd, even though beliefs might never settle down. When private signals sufficiently violate a log-concavity condition, a cascade can arise.

Lee (1993) noted that beliefs can be perfectly revealed when the action space is continuous just like the belief space. The social learning paradigm instead by and large explores when a coarse action set communicates the private beliefs of decision makers. It may sufficiently frustrates the learning dynamics that an incorrect action herd occurs. If individuals seek to help each other by taking more informative actions, and if this signaling is understood by successors, then any cascade sets shrink, and the welfare of later individuals generally rises. As we show in Smith and Sørensen (2008), the analysis is qualitatively similar to that outlined here, although solving for the new, forward-looking transition chances requires dynamic programming.

A greater message of social learning is the self-defeating nature of learning from others. Moving outside the finite action, sequential entry model into a Gaussian world, Vives (1993) found that social learning is slower than private learning in a market setting where individual decisions are obscured by Gaussian noise.

If observations are not made of an ever expanding history, such as simply knowing the number but not order of past action choices, then our approach is less useful. The survey by Gale and Kariv (2008) discuss the problem of learning in networks. In Smith and Sørensen (1994), and chapter 3 of Sørensen (1996), we identified a case where the stochastic difference equation is a useful tool, even when public beliefs do not follow a martingale.

References

ANDERSON, L., AND C. A. HOLT (2008): “Information Cascade Experiments,” in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf, and L. E.

- Blume. Palgrave Macmillan, New York.
- 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.
- (2008): “Information Cascades,” in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf, and L. E. Blume. Palgrave Macmillan, New York.
- ÇELEN, B., AND S. KARIV (2004): “Distinguishing Informational Cascades from Herd Behavior in the Laboratory,” *American Economic Review*, 94, 484–498.
- GALE, D., AND S. KARIV (2008): “Learning and Information Aggregation in Networks,” in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf, and L. E. Blume. Palgrave Macmillan, New York.
- LEE, I. H. (1993): “On the Convergence of Informational Cascades,” *Journal of Economic Theory*, 61, 395–411.
- OTTAVIANI, M., AND P. N. SØRENSEN (2006): “Professional Advice,” *Journal of Economic Theory*, 126, 120–142.
- SCHARFSTEIN, D. S., AND J. C. STEIN (1990): “Herd Behavior and Investment,” *American Economic Review*, 80, 465–479.
- SMITH, L., AND P. SØRENSEN (1994): “An Example of Non-Martingale Learning,” MIT Working Paper.
- (2000): “Pathological Outcomes of Observational Learning,” *Econometrica*, 68, 371–398.
- SMITH, L., AND P. N. SØRENSEN (2008): “Informational Herding and Optimal Experimentation,” University of Copenhagen Working Paper.
- SØRENSEN, P. (1996): “Rational Social Learning,” Ph.D. thesis, MIT.
- VIVES, X. (1993): “How Fast do Rational Agents Learn?,” *Review of Economic Studies*, 60, 329–347.