Observational Learning and Demand for Search Goods

Ken Hendricks
University of Texas at Austin

Alan Sorensen
Stanford University & NBER

Tom Wiseman
University of Texas at Austin

November 22, 2008
Very Preliminary

Abstract

In many differentiated good markets like music, books, and movies, the choice set of available products is overwhelmingly large and growing as new products flow into the market each month. Consumers are not aware or poorly informed about many of the available products. They learn about products and their preferences for them from the purchasing decisions of other consumers and through costly search. We use a variant of the sequential search models of Banerjee [2], Bikhchandani, Hirshleifer, and Welch [5] and Smith and Sorensen [18] to study market demand in these kinds of markets. The option to search prior to purchase leads to different dynamics and outcomes than the standard herding models. The results explain both the unpredictability of sales conditional on quality and the inequality of sales across products. The model also yields testable predictions regarding the impact of product quality, search costs, and price on the likelihood of a high-quality product ending up with low sales (i.e., a "bad" herd). We validate the model using data from an experimental study by Salganik, Dodds, and Watts [15].

1 Introduction

In many differentiated good markets, the number of available products is overwhelmingly large. The sheer size of the choice set implies that consumers are often unaware or poorly informed about many products, especially new products. According to the marketing literature, the response of consumers is often to reduce their choice problem by selecting a much smaller set of products. Eliaz and Spiegel [9] refer to such sets as "consideration sets". After selecting consideration sets, consumers decide which products to purchase based upon their information about the products and their preferences. When the products are search goods, consumers can learn their preferences for the goods prior to purchase by acquiring informative signals about the goods. Market demand in these markets then depends not only upon consumers' knowledge of the product space and their preferences, but also upon the process by which they select a consideration set. This process is driven in part by observational learning, which occurs when the behavior of individuals is influenced by observing the choices of other people. Consumers tend to consider the products they hear about, and they hear about the products that other consumers buy. As a result, a product's success or lack of success reinforces itself, causing market demand to be unpredictable and sales across products to be substantially more skewed than in a world where consumers costlessly know about all available products and their preferences for them.

Our main goal in this paper is to examine the impact of observational learning on market demand for search goods. Search goods are products whose quality can be ascertained by consumers prior to purchase. We use a variant of the learning models introduced by Banerjee [2] and Bikhchandi, Hirshliefer and Welch [5] and subsequently generalized by Smith and Sorensen [18]. An infinite number of consumers with heterogenous preferences arrive sequentially and have to decide whether or not to buy a new product whose quality can either be high or low. Consumers do not know the quality of the product or their idiosyncratic preferences for it.[1] Prior to making their purchasing decision, consumers observe the purchasing history of previous consumers or some summary statistic of that history. They can also engage in costly search, which informs them about their preferences for the product. For example, in the market for recorded music, search would involve listening to songs on the internet or at the listening post in the store. A positive search

[1] Smith and Sorensen [18] also admit heterogeneous preferences but assume that consumer know the idiosyncratic component.

cost provides a rationale for consumers *not* to consider the product if they believe they are unlikely to buy it. More generally, it captures the idea that, even though cost per product may be quite small, the number of products is too large for a consumer to search every product.[2]

When consumers search prior to purchase, the purchasing decision of each consumer is informative to subsequent consumers and influences their search decisions. However, it does not perfectly reveal which search actions were taken. If a consumer does not buy a product, it may be because she decided not to consider it or because she did consider it and did not like it. We show that, in the long-run, two outcomes are possible. One is that a herd forms on the "search" action. In this case, the market learns the quality of the product and demand converges to its "true" market share (i.e., the share that it would obtain if product quality is known). This outcome can only occur if product quality is high. The other outcome is that a herd forms on the "no search" action. In this case, the market does not learn the true quality of the product and sales converge to zero. This outcome is certain to occur if product quality is low and occurs with positive probability if product quality is high. Thus, observational learning with search prevents a population of consumers from considering low quality products but can lead them not to consider, and therefore not to buy, high quality products. We refer to this event as a "bad" herd. These results differ from those of the standard herding model where the market never learns the true quality of products and market shares never converge to their correct market shares since eventually everybody either buys or does not buy with positive probability in each state.

The fact that observational learning can generate "bad" herds is not surprising. The more interesting issue is how the probability of a "bad" herd depends upon such factors as product quality, price, signal quality, and search costs. Robust comparative statics on the likelihood of "bad" herds are surprisingly difficult to obtain in herding models. The key restriction is the posterior likelihood ratio that the product is low rather than high has to be monotone increasing in the prior likelihood ratio following each action. We provide sufficient conditions on the joint distribution of signals and preferences for the monotonicity condition to hold. Given this condition, the likelihood ratio for a high quality product is forever trapped between two stationary points and we are able to obtain a closed form solution for the probability that the ratio does not converge to zero (i.e., market learns true

[2]We are working on a model with more than one product. The choice dynamics in a multi-product are more complex because of the interactions in the learning process across products.

quality). We then show that the probability of a "bad" herd increases with product price and search costs, and decreases with product quality.

Rosen [13] argued that the reward function to quality is convex because, in equilibrium, more talented artists can sell more units at higher unit prices. Our results suggests another source of convexity: in the long-run, consumers are more likely to learn about higher quality products and are more likely to buy them. Thus, small differences in product quality can lead to large differences in expected sales even when prices do not vary with quality. It can also explain why prices of products like albums, books and videos do not vary with quality. A small increase in price can have a disproportionate effect on expected sales since it decreases market share *and* increases the probability of a "bad" herd. We also show that the impact of a decrease in search costs is smaller on higher quality products. This result implies that a decline in search costs due to the Internet has a larger impact on the sales of niche products than on hit products.

There is a large empirical literature that tries to quantify the effect that the choices of others have on an individual's choices and identify the source of the effect (see Cai, Chen, and Fang [8] for discussion and references). One possible source is observational learning but another plausible mechanism is that individuals want to conform.[3] However, none of these studies have studied the implications of social influence on choice dynamics and outcomes. The exception is an ingenious experimental study conducted by Salganik, Dodds, and Watts [15] (hereafter referred to as SDW). SDW created an artificial music market in which hundreds of participants arrived sequentially at a website to listen to 48 songs by unknown artists. After listening to a song, participants could choose to download it. Listening to a song is analogous to product search in our model, and downloading is analogous to purchasing the product. In the control experiment, the songs were shown in random order, with no information about the previous participants' listening or downloading choices. The fraction of consumers who download a song conditional on listening to it (SDW called this fraction the song's "batting average") can be interpreted as a measure of the song's intrinsic quality. In the treatment experiments, of which they were eight, the songs were listed by download rank with download counts. The authors find that download rates in the treatment experiments were more highly skewed than in the control experiment, and quite unpredictable, particularly for songs with higher batting averages. The worst songs

[3] Cai, Chen, and Fang [8] also discuss a third mechanism, the saliency effect. The choices of other consumers may matter simply because those choices are highlighted.

never did well.

We use data from the SDW study to validate our model. The authors interpreted their results as evidence of social influence on download decisions but did not directly address the issue of the source of the social influence: did the download information affect choices primarily through observational learning (i.e., information) or through a desire to conform (i.e., preferences)? The observational learning hypothesis implies that download rate should affect listening probabilities but not the download probabilities conditional on listening. The conformity hypotheis implies that download rates should effect both probabilities. We examine these hypotheses using a probit analysis. We find that a higher download rank and count causes the song's listening probability to increase but the download information has no impact on the conditional download probability after controlling the song's intrinsic quality (i.e., its batting average in the control experiment). The download rates in the control experiment allows us to resolve the identification problem that Manski [11] calls the "reflection problem". Our analysis also controls for "framing" effect that can arise from ordering the songs by download rates. Thus, we conclude that participants download the songs that others download primarily because they tend to listen to the songs that others download. We also examine the relationship between song quality and long-run outcomes. The herding model predicts that the listening and downloading rates for the better songs will either converge to zero or to some positive fraction bounded away from zero. The outcomes in the eight treatment experiments are largely consistent with this prediction.

Hendricks and Sorensen [10] provide indirect evidence of the importance of observational learning on market demand for albums. Using detailed albums sales data for 355 artists, they show that many albums that flopped when they were released succeeded after the artists released another album that was a hit. If consumer preferences do not change over time, then the spillover reflects the arrival of new information that led many consumers to consider buying the debut album. In other words, the debut album suffered from a "bad herd", and the artist's new hit album caused consumer beliefs about the debut album to change enough that many of them were willing to consider buying it. Hendricks and Sorensen estimated a model of album demand based on this simple idea and found that it fits the data remarkably well. The parameter estimates indicate that almost all consumers know their preferences for a debut album that is a major hit, but only 32% of consumers know about their preferences for a debut album that achieves the median level of sales. This finding implies that album sales would have been substantially less skewed in a world where all consumers knew their preferences for the debut albums. For example, the authors estimated that the sales of the top artist in the sample would have exceeded the median artist's sales by a factor of 30 instead of the observed factor of 90.

The paper is organized as follows. In Section II, we describe our model. In Section III, we characterize the equilibrium dynamics and outcomes. In Section IV we present the comparative results. Section V provides an extension of the model in which the consumers have heterogeneous search costs. Section VI studies the application of the model to data from the experimental study by SDW. Section VII concludes.

2 The Model

In this section we present a sequential choice model in which heterogenous consumers arrive randomly and have the option of searching before deciding whether or not to purchase a product of unknown quality. Search involves acquiring a costly, private signal about preferences for the product. Each consumer's purchasing decision is observable to later consumers but her search decision is not observable.

An infinite sequence of consumers indexed by t enter in exogenous order. Each consumer makes an irreversible decision on whether or not to purchase the product. Consumer t's utility for the product is given by

$$V_t = X + U_t$$

where X denotes the mean utility or quality of the product and U_t is the idiosyncratic component. Here U_t is identically and independently distributed across consumers with zero mean. Let F_U denote the distribution of U. There are two quality levels: $X = H$ and $X = L$, where $H > L$. We will refer to H as the high quality state and L as the low quality state. We normalize $L = 0$. Consumer t does not know X or U_t. There is a common prior belief that assigns a probability μ_0 to the event that $X = H$. The price of the product is p. Consumers' utility is quasilinear in wealth, so consumer t's net payoff from purchasing the product is $V_t - p$.

Consumer t has two available actions. Buying the product involves risk since the ex post payoff may be negative. She can reduce the likelihood of this event by choosing to *Search (S)* before making her purchasing decision. Search involves paying a cost c to obtain a private, informative signal about V_t, and then purchasing if the expectation of V_t conditional on the signal exceeds p and not purchasing otherwise. For notational simplicity, it will be convenient to assume that the signal is perfectly informative and reveals V_t precisely.[4] Note that search remains a valuable option for consumer t even if she has learned X. The other action that she can choose is to *Not Search and Not Buy (N)*. Let $a_t \in \{N, S\}$ denote the action chosen by consumer t.

Given the consumer's purchasing rule following search, the expected value of search conditional on state X is

$$w(X) = \int_{P-X}^{\infty} (X - p + u) dF_U(u).$$

Hence, the payoff to a consumer from action S in state X is $w(X) - c$. We impose the following restrictions on the payoffs from search in each state.

A1: (a) $w(H) - c > 0$ and $w(H) - c > H - p$; (b) $w(0) - c < 0$.

Condition (a) of Assumption A1 states that, conditional on H, the payoff to S is positive and exceeds the payoff from buying without search. It implies that consumers never purchase without search even when they know that the state is H. The second inequality states that the consumer's payoff to S is negative if she knows that the state is L. It implies that the consumer's optimal action in state L is N.

Consumer t's action generates a purchasing outcome $b_t \in \{0, 1\}$. Here $b_t = 0$ is the outcome in which consumer t does not purchase the good and $b_t = 1$ is the outcome in which consumer t purchases the product. Outcome 0 occurs if consumer t chooses N or if she chooses S and obtains a realization of V_t such that her net payoff from purchase is negative. Outcome 1 arises if consumer t chooses S and obtains a realization of V_t such that her net payoff from purchase is positive.

Before taking her action, consumer t observes a private signal about the quality of the product.[5] Smith and Sorensen [18] have shown that there is no loss in generality in defining the private signal that a consumer receives as her *private belief*. Here we denote the signal by σ and define it as the probability that the state is H. Conditional on the state, the signals are identically and independently distributed across consumers and drawn from a distribution F_X, $X = H, L$. We assume that F_L and F_H are continuous and differentiable with densities f_L and f_H. Under the assumption that both states are equally likely, the

[4] The important restriction is that the signal is informative about X and not just U.

[5] It is not yet clear whether we can generalize the model to allow the signals to be informative about X and U.

unconditional distribution of σ is $F = (F_L + F_H)/2$ with density f. Smith and Sorensen (2000) show that defining the private signal in this way implies that the joint distribution of X and σ possesses the monotone likelihood ratio property. As is well known, this property implies that the conditional distributions F_L and F_H as well as their hazard and reverse hazard rates are ordered. Private beliefs are *bounded* if the convex hull of the common support of F_L and F_H consists of an interval $[\underline{d}, \overline{d}]$ where $\underline{d} > 0$ and $\overline{d} < 1$.

In addition to the private signal, consumer t also observes the purchasing decisions of consumers 1 through $t-1$. She does not observe the private signals they received, the actions they chose or, if they searched, the information they obtained, and if they purchased, the payoffs they realized. The private signals imply that the consumers' search decisions are also private information. This feature of the model has important implications for outcomes and learning dynamics. The space of possible t-period purchase histories is given by $\Omega_t = \{0, 1\}^{t-1}$ and a particular history is denoted by ω_t. The initial history is defined as $\omega_1 = \varnothing$. The assumption that consumers observe the entire ordered action history ω is obviously quite strong. In most markets, consumers are likely to know only the number of past purchases. Later we consider situations in which consumers only observe some summary statistics of the purchasing history. Note that the assumption that the private signals are not informative about U implies that t-period history is also not informative about U_t.

Given any history ω_t, consumer t updates her beliefs about X using Bayes rule. Let $\mu_t(\omega_t)$ represent her posterior belief that the state is H conditional on history ω_t. Since ω_t is publicly observable, μ_t is also the *public* belief in period t. Given public belief μ_t and private signal σ_t, consumer t's *private* belief that the state is H is

$$r(\sigma_t, \mu_t) = \frac{\sigma_t \mu_t}{\sigma_t \mu_t + (1 - \sigma_t)(1 - \mu_t)}. \tag{1}$$

In studying the dynamics of beliefs and actions, we follow Smith and Sorensen [18] and work with the public likelihood ratio that the state is L versus H rather than public beliefs. Define

$$l_t = \frac{1 - \mu_t}{\mu_t}.$$

and let l_0 denote the prior likelihood ratio. Using this transformation of variables in equation (1), consumer t's private belief that the state is H becomes $r(\sigma_t, l_t)$. Her expected payoff to S is

$$\pi(S; \sigma_t, l_t) = r(\sigma_t, l_t) w(H) + (1 - r(\sigma_t, l_t)) w(0) - c. \tag{2}$$

Recall that L is normalized to zero. Therefore, if the consumer chooses N, her payoff is zero. We look for a Bayesian equilibrium where everyone computes posterior beliefs using Bayes rule, knows the decision rules of all consumers and knows the probability laws determining outcomes under those rules.

A *cascade* on action $a \in \{S, N\}$ occurs when a consumer chooses a regardless of the realization of her private signal σ. Because of the distinction between actions and outcomes, we have to careful in defining a herd. We say that a *herd* on action a occurs at time n if each consumer $t \geq n$ chooses action a. Note that while a herd on N implies that all future outcomes are the same (all 1's and 0's, respectively), a herd on S does not. The outcome for a consumer who chooses S depends not only on X (which is common across consumers) but also on the realization of the idiosyncratic component U. In fact, a herd on S precludes the event that all future outcomes are the same (almost surely) - if the outcome does not vary with the realization of U, then it is not worthwhile paying c to search.

A related concept is outcome convergence. Let $\lambda_t \in [0, 1]$ be the fraction of the first $t-1$ consumers whose outcome was 1 (purchase). Outcome convergence is the event that λ_t converges to some limit $\lambda \in [0, 1]$. A herd implies outcome convergence. A herd on N leads to $\lambda = 0$; and a herd on S leads to $\lambda = 1 - F_U(p - X)$.

How does our model differ from the standard herding model? In the standard herding model, actions are observable; in our model, they are not. Instead, consumers observe outcomes, which are signals about the actions taken. More precisely, when consumer t does not purchase the product, subsequent consumers do not know whether it is because she chose N or because she chose S and obtained a realization on V_t such that her net payoff from purchase was negative. If consumer t purchases the product, then subsequent consumers can infer that she chose S. As we shall see, the endogeneity of the signal generated by the action taken has important implications for outcomes and for the learning dynamics.

3 Outcomes and Learning Dynamics

In this section we characterize the equilibrium outcomes and dynamics. We begin by defining thresholds. Let $\widehat{r}$ represent the private belief at which a consumer is indifferent between S and N. From equation (2),

$$\widehat{r} = \frac{c - w(0)}{w(H) - w(0)}. \tag{3}$$

Assumption A1 implies that $\widehat{r} \in (0,1)$. Using equations (1) and (3), we can then define the private signal at which a consumer is indifferent between S and N (assuming it is interior) as

$$\widehat{\sigma}(l) = \frac{(c - w(0))l}{w(H) - c + (c - w(0))l}. \tag{4}$$

Thus, given l, the consumer's optimal action is to choose S if $\sigma \geq \widehat{\sigma}$ and to choose N if $\sigma < \widehat{\sigma}$. We will refer to $\widehat{\sigma}$ as the *search threshold*.

Next we define the cascade regions. Let $\underline{l}$ denote the largest value of the public likelihood ratio such that a consumer is certain to choose S. From equation (4), $\underline{l}$ satisfies $\widehat{\sigma}(\underline{l}) = \underline{d}$. Solving this equation for $\underline{l}$ yields

$$\underline{l} = \frac{\underline{d}(w(H) - c)}{(1 - \underline{d})(c - w(0))}. \tag{5}$$

Let $\overline{l}$ denote the lowest value of the public likelihood ratio such that a consumer is certain to choose N. Once again, using equation (4), $\overline{l}$ satisfies $\widehat{\sigma}(\overline{l}) = \overline{d}$. Solving this equation for $\overline{l}$ yields

$$\overline{l} = \frac{\overline{d}(w(H) - c)}{(1 - \overline{d})(c - w(0))}. \tag{6}$$

Thus, we can partition the values of the public likelihood ratio into three intervals. When $l < \underline{l}$, there is a cascade on S; when $\underline{l} \leq l \leq \overline{l}$, the consumer searches with probability $1 - F(\widehat{\sigma}(l))$ and does not search with probability $F(\widehat{\sigma}(l))$; and when $l > \overline{l}$, there is a cascade on N.

We now characterize the dynamics of the public likelihood ratio. Suppose $\underline{l} < l_t < \overline{l}$. Then the probability that consumer t buys the product in state X is

$$\Pr\{b_t = 1|X, l_t\} = (1 - F_X(\widehat{\sigma}(l_t)))(1 - F_U(p - X)).$$

It is the probability that consumer t searches in state X times the probability that she gets a realization of U that lies above the purchasing threshold $p - X$. The probability that consumer t does not buy the product in state X is

$$\Pr\{b_t = 0|X, l_t\} = F_X(\widehat{\sigma}(l_t)) + (1 - F_X(\widehat{\sigma}(l_t)))F_U(p - X).$$

The first term is the probability of the event that consumer t does not search in state X; the second term is the probability of the event that consumer t searches in state X and gets a value of U that lies below the purchasing threshold. Using Bayes' rule, the public likelihood ratio in period $t+1$ is given by

$$l_{t+1}(b_t) = \begin{cases} \left[\dfrac{(1-F_L(\widehat{\sigma}(l_t)))(1-F_U(p))}{(1-F_H(\widehat{\sigma}(l_t)))(1-F_U(p-H))}\right] l_t & \text{if } b_t = 1 \\ \left[\dfrac{F_L(\widehat{\sigma}(l_t)) + (1-F_L(\widehat{\sigma}(l_t)))F_U(p)}{F_H(\widehat{\sigma}(l_t)) + (1-F_H(\widehat{\sigma}(l_t)))F_U(p-H)}\right] l_t & \text{if } b_t = 0 \end{cases}. \tag{7}$$

The dynamic system consists of a pair of non-linear, first-order difference equations with initial condition l_0. When $l_t < \underline{l}$, $F_L(\widehat{\sigma}) = F_H(\widehat{\sigma}) = 0$, and the dynamic system reduces to a pair of *linear,* first-order difference equations. When $l_t > \bar{l}$, $F_L(\widehat{\sigma}) = F_H(\widehat{\sigma}) = 1$ and $l_{t+1} = l_t$.

Smith and Sorensen [18] show the following:

Lemma 1 *Conditional on state H, the public likelihood ratio is a martingale. It converges to a random variable with support in $[0,\infty)$ so fully wrong learning has probability zero. Conditional on state L, the inverse public likelihood ratio is a martingale. It converges to a random variable with support in $[0,\infty)$.*

The martingale property rules out convergence to nonstationary limit beliefs such as cycles or to incorrect point beliefs.

For notational convenience, define

$$\psi_0(l_t) = \frac{F_L(\widehat{\sigma}(l_t)) + (1-F_L(\widehat{\sigma}(l_t)))F_U(p)}{F_H(\widehat{\sigma}(l_t)) + (1-F_H(\widehat{\sigma}(l_t)))F_U(p-H)}$$

$$\psi_1(l_t) = \frac{(1-F_L(\widehat{\sigma}(l_t)))(1-F_U(p))}{(1-F_H(\widehat{\sigma}(l_t)))(1-F_U(p-H))}$$

With this notation, equation (7) becomes

$$l_{t+1}(b_t) = \begin{cases} \psi_1(l_t)l_t & \text{if } b_t = 1 \\ \psi_0(l_t)l_t & \text{if } b_t = 0 \end{cases}. \tag{8}$$

The following lemma follows from our assumptions on the conditional distributions F_H and F_L.

Lemma 2 *(i) ψ_0 is continuous, $\psi_0(l) > 1$ on $[0,\bar{l})$, and $\psi_0(l) = 1$ on $[\bar{l},\infty)$. (ii) $\psi_1(l) < 1$ and continuous on $[0,\bar{l})$ with $lim_{l\uparrow\bar{l}}\psi(l) > 0$.*

A particular useful property for comparative static purposes is when the posterior likelihood ratio following each action is monotone increasing in the prior likelihood ratio. Smith and Sorensen [19] find that their herding model has this property if the density of the private belief log-likelihood ratio is log-concave. In our model, log-concavity is sufficient to establish that the "buy" difference equation is strictly increasing in l_t for $l_t < \bar{l}$ but it is not sufficient for the "no buy" difference equation. In the former case, there is no distinction between outcome and action: when consumer $t+1$ observes the outcome "buy", she knows that consumer t chose action S. This is not true when consumer $t+1$ observes the outcome "no buy": it is consistent with consumer t choosing either action N or S. As a result, we need to impose an additional restriction on the distribution of U [and σ] to ensure monotonicity.

For any private belief σ, let γ denote the natural log of the corresponding likelihood ratio:

$$\gamma = \ln\left(\frac{1-\sigma}{\sigma}\right).$$

Denote the unconditional distribution of γ by F', with density f'.

A2: (i) The density of the log likelihood ratio, f', is log-concave and (ii) $E[\gamma|\gamma \geq g] - E[\gamma|\gamma < g] > \ln F_U(p) - \ln F_U(p-H)$ for all g in the support of γ.

Part (ii) of Assumption A2 ensures that, at any public belief, a decision not to search (after receiving the private signal) is more suggestive of state L than deciding to search but then not buying.

Lemma 3 *Assumption A2 implies that $\psi_0(l_t)l_t$ and $\psi_1(l_t)l_t$ are increasing in $l_t \in [0,\bar{l}]$.*

Figure 1 illustrates a dynamic system that satisfies Assumption A2. The cascade set for S is the interval $[0,\underline{l}]$ and the cascade set for N is the interval $[\bar{l},\infty)$. The "no buy" difference equation intersects the diagonal at 0 and at $\bar{l}$ and lies everywhere above the diagonal in between these two values. It is linear on the cascade set for S and strictly increasing on the interval $(\underline{l},\bar{l})$. The "buy" difference equation intersects the diagonal at 0 and lies below the diagonal for positive values of l_t.It is linear on the cascade set for S and increasing on the interval $(\underline{l},\bar{l})$. Active dynamics occur when the prior is such that $l_0 \in (0,\bar{l})$.

We can adapt Smith and Sorensen's [18] arguments to derive the following results.

Proposition 4 *Suppose $0 < l_0 < \bar{l}$. (a) Outcome convergence occurs almost surely. In state H, $\lambda = 1 - F_U(p-H)$ with positive probability and $\lambda = 0$ with positive probability;*

in state L, $\lambda = 0$. (b) In state H, beliefs converge to the truth when $\lambda = 1 - F_U(p - H)$; otherwise, the limit belief converges to $\bar{l}$ and learning is incomplete. Beliefs can never enter the cascade set for N from outside. (c) Actions in state H converge almost surely to S when $\lambda = 1 - F_U(p - H)$; otherwise they converge almost surely to N.

The key feature of the dynamics is that the public likelihood ratio never enters the cascade set for N from outside. For any $l_0 < \bar{l}$, the dynamics are forever trapped between the stationary points 0 and $\bar{l}$ and convergence is always asymptotic. A sequence of "buy" outcomes causes the likelihood ratio to decrease in ever smaller increments towards 0; a sequence of "no buy" outcomes causes the likelihood ratio to increase in ever smaller increments to $\bar{l}$. A herd always eventually starts. In state H, the herd can form with positive probability on either N or on S. If it forms on S, then the market eventually learns the true state from the frequency of purchases. In state L, the herd can only form on N. Intuitively, if a herd formed on S, then the purchase frequency reveals the state is L, which contradicts the assumption that it is not optimal to search in state L.

To illustrate the role of Assumption A2, we consider a special case in which the private signal is completely uninformative about X. In this case, the distributions F_L and F_H are degenerate at $\sigma_t = 1/2$, the thresholds $\underline{l} = \bar{l}$ where

$$\bar{l} = \frac{(w(H) - c)}{(c - w(0))},$$

and each consumer's private belief is equal to the public belief. Hence, consumer t is certain to search when $l_t < \bar{l}$ and is certain not to search when $l_t > \bar{l}$. Given any $l_t < \bar{l}$, equation (7) reduces to

$$l_{t+1}(b_t) = \begin{cases} \dfrac{1 - F_U(p)}{1 - F_U(p - H)} l_t \text{ if } b_t = 1 \\ \dfrac{F_U(p)}{F_U(p - H)} l_t \text{ if } b_t = 0. \end{cases}$$

Otherwise, $l_{t+1} = l_t$.

Figure 2 presents the phase diagram for the two linear difference equations. The slope of the "buy" equation is less than 1 which implies that, given any $l_t < \bar{l}$, a sequence of "buy" outcomes causes the public likelihood ratio to converge in ever smaller increments to 0. Thus, conditional on H, there is a positive probability that beliefs converge to the truth. The slope of the "no buy" equation exceeds 1. The downward discontinuity in $\psi_0(l_t)l_t$ at $\bar{l}$

arises from the fact that probability of search is not continuous: it is equal to 1 if $l_t < \bar{l}$ and zero otherwise. The important property is not the discontinuity but the fact that $\psi_0(l_t)l_t$ is not monotone increasing at $\bar{l}$. When this is the case, a sequence of "no buy" outcomes causes the likelihood ratio to increase in ever larger increments and, after a finite number of periods, "jump" beyond $\bar{l}$. Convergence is not asymptotic and, given any $l_t < \bar{l}$, the set of rest points for the likelihood ratio consists of 0 and a nondegenerate subset of the stationary points $\{l \geq \bar{l}\}$. As we shall see in the next section, models with these kinds of dynamics yield relatively few testable predictions.

The above analysis shows how search modifies the results of herding models. In the standard herding model, the market never learns the true quality of products since eventually everybody either buys or does not buy and both outcomes can occur with positive probability in each state. Thus, the market share of high quality products never converges to their correct market shares, and the market share of low quality products converges to the wrong market share with positive probability. By contrast, when consumers can search prior to purchase, the market can learn the true quality of high quality products and outcomes can converge to the correct market shares. They are certain to do so for low quality products and with positive probability for high quality products. High quality products can still fail in our model due to the herding effect: outcomes converge to 0 with positive probability. Thus, our model generates a stochastic mapping between quality and market outcomes that allows market shares to vary with quality but explains why high quality products can still fail.

4 Comparative Statics

In this section we study the comparative static properties of our model. The goal is to derive testable predictions that can be applied to data. We will be interested in two sets of questions: how does cumulative sales vary with product price, quality, and search costs and how does long run sales vary with these parameters? Throughout this section, we will assume that the number of consumers is very large but finite.

Cumulative sales of a product depends upon two factors: the number of consumers who search and the fraction of these consumers who purchase the product. Let M denote the equilibrium number of consumers who search. It is a random variable whose realizations depend upon whether and when a herd forms on N. Let $F_{M|X}$ denote the distribution of

M conditional on state X. Applying the law of large numbers, the fraction of M consumers that purchase the product is given by the probability of purchase, $1 - F_U(p - X)$. This probability is clearly decreasing in price and increasing in product quality. Unfortunately, similar comparative static results on $F_{M|X}$ (and therefore on cumulative sales) are difficult to obtain.

When price increases or, equivalently, H decreases, the probability of purchase falls, which in turn reduces the value of search. Thus, given any sequence of realizations of idiosyncratic shocks, zero outcomes are more likely to occur because search is less likely and purchasing conditional on search is less likely. However, more zero outcomes do not necessarily cause public beliefs to be more pessimistic. The change in purchasing and search rules affects the informativeness of the signal generated by the purchasing outcomes, and consumers take these changes into account when they update their beliefs about the state. It is not difficult to construct examples of sequences in which the number of consumers who search is actually higher at higher prices or lower quality. As a result, it is not possible to determine the effects of changes in price or product quality on the distribution of M or its moments.

The one case in which we are able to obtain a comparative static result on M is when private signals are completely uninformative. In this model, consumers are certain to search until public beliefs enter the cascade set for N. The purchasing rule also does not depend upon c. Consequently, given any sequence of realizations of idiosyncratic shocks, outcomes are unaffected by an increase in c as long as public beliefs are less than $\bar{l}$. Thus, the impact of an increase in search costs on M is completely determined by its impact on $\bar{l}$. The following proposition establishes that distribution of M is stochastically decreasing in c.

Proposition 5 *Suppose private signals are completely uninformative. Then* $F_{M|X}(m; c) \leq F_{M|X}(m; c')$ *for* $c > c'$, $X = L, H$.

We turn next to long run sales. The main prediction of our model (which does not depend upon Assumption A2) is that, in the long run, sales of high quality products either converge to 0 or to their true market shares. The question of interest is how the probability of a "bad" herd varies with the parameters of the model. Recall that, since the likelihood ratio process $\langle l_t \rangle$ is a bounded martingale conditional on state H, the $E[l_\infty] = l_0$. This property of martingales is quite useful when public beliefs cannot enter the cascade set for

N from outside. When this is the case, $\text{supp}l_\infty = \{0, \bar{l}\}$, from which it follows that

$$\Pr\{l_\infty = \bar{l}\} = \frac{l_0}{\bar{l}}. \tag{9}$$

Thus, the sign of the changes in product quality, search costs, and price on the probability of a "bad" herd (i.e., long-run sales are 0) are determined by the sign of their impact on the value of $\bar{l}$. Differentiating $\bar{l}$ with respect to H, c and p yields the following results:

Proposition 6 *Suppose the state is H. Then the probability of a limit cascade on N is (a) strictly decreasing in H; (b) increasing and convex in c; (c) increasing in p.*

The proposition yields several predictions. An increase in search costs increases the likelihood that long-run sales of a high quality product is zero and hence reduces its expected long-run sales. The impact of the increase is larger at higher cost levels. An increase in price also increases the probability of zero long-run sales but the sign of the second derivative of $\bar{l}$ with respect to p depends upon the distribution of the private signal. This issue is important because, if $\bar{l}$ is convex, then the model could explain why search goods like music albums and books sell for the same prices even when the distributors know that the quality of their products vary.

A number of papers (e.g., [7]) have argued that the decline in search costs due to the Internet has disproportionately increased sales of niche products and reduced the concentration of sales. The next proposition provides support for this claim.

Proposition 7 *Suppose the state is H. Then the impact of an increase in c (or an increase in p) on the probability of a limit cascade on N is smaller (in absolute value) for higher quality products.*

The results follows from differentiating $\bar{l}$ with respect to H and c (and H and p). Proposition 7 implies that a decrease in search costs has a larger impact on long-run sales of niche products (i.e., medium quality products) than on high quality products.

Finally, the probability of a limit cascade on N does not depend upon the precision of the private signal. The only property of F that matters (aside from continuity) is $\bar{d}$, the upper bound of its support. This is a striking result, which has important implications for the kind of information that the market should reveal about consumers' purchasing decisions.

The comparative static results depend critically upon Assumption A2. When public beliefs can enter the cascade set on N from outside, the support of l_∞ consists of 0 and a nondegenerate subset of the cascade set. We can use the martingale property to bound the probability of a "bad" herd but cannot predict how this bound will vary with the parameters of the model. Pastine and Pastine [12] obtained similar "perverse" results when they studied the effect of changing the accuracy of signals on the probability of incorrect herds in a herding model where beliefs can enter the cascade set from outside.

The main prediction of our model (which does not depend upon Assumption A2) is that, in the long run, sales of high quality products either converge to 0 or to their true market shares. When sales converge to 0, a seller has an incentive to invest in a signal (e.g., advertisements) that can change public beliefs and lead the market to correct its mistake.

Proposition 8 *The introduction of a sufficiently positive, public signal after beliefs have converged (a) leads to an increase in short-run sales and, with positive probability, a herd on S if the initial herd is on N; (b) has no effect on sales if the initial herd is on S.*

In their study of recorded music, Hendricks and Sorensen [10] show that the release of a hit new album by an artist can substantially raise sales of the artist's catalog albums. They attribute the spillover to consumer learning. The new release causes some consumers to discover the artist or to revise their beliefs about the artist's catalog albums and purchase them. In the context of our model, the release of a hit new album can be interpreted as a strongly positive, public signal about the catalog album. If public beliefs about catalog albums have previously converged to $\bar{l}$, then the release of a new album that is a hit can cause public beliefs to move away from $\bar{l}$ and sales can re-converge with some probability to the true market share.

5 Extension

In some applications, the primary source of consumer heterogeneity in search is likely to be differences in search costs rather than differences in signals about the unknown quality of the product. We consider such an application in the next section. Here we show how the model needs to be modified to consider heterogeneity in search costs.

Consumer t's private seach cost c_t is an independent draw from a distribution F_C with support $[\underline{c}, \bar{c}]$. We shall impose the following restrictions on consumer payoffs.

Assumption 3: (i) $\underline{c} > w(0)$ and (ii) $H - p < 0$.

Condition (i) states that the value of search in state L is less than the lowest possible search cost. It implies that optimal action for every consumer is not to search if they are certain that the state is L. Condition (ii) states that the expected payoff to buying the product in state H is always less than the price. It implies that action "buy without search" is always dominated by S or N.

Let $\widehat{c}$ denote the cost at which a consumer is indifferent between searching and not searching. We will refer to $\widehat{c}$ as the *cost threshold*. It is defined by

$$\widehat{c} = \mu w(H) + (1-\mu)w(0),$$

or equivalently,

$$\widehat{c}(l) = (1+l)^{-1}(w(H) - w(0)) + w(0).$$

Note that the cost threshold is strictly decreasing l. In other words, when the likelihood of state L increases relative to state H, then consumers are less likely to search.

We now define the cascade sets. Here $\bar{l}$ is given by

$$\widehat{c}(\bar{l}) = \underline{c} \Longrightarrow \bar{l} = \frac{w(H) - \underline{c}}{\underline{c} - w(0)}.$$

Similarly, $\underline{l}$ is given by

$$\widehat{c}(\underline{l}) = \bar{c} \Longrightarrow \underline{l} = \frac{w(H) - \bar{c}}{\bar{c} - w(0)}$$

if $w(H) > \bar{c}$ and zero otherwise. In the latter case, a positive fraction of consumers will not search and hence not buy the product even when they are certain that the state is H.

We now characterize the dynamics of the public likelihood ratio. Suppose $l_t < \bar{l}$. Then the probability that consumer t buys the product in state X is

$$\Pr\{b_t = 1|X, l_t\} = F_C(\widehat{c}(l_t)))(1 - F_U(p - X)).$$

It is the probability that consumer t's search cost is less than the threshold cost times the probability that she gets a realization of U that lies above the purchasing threshold $p - X$. The probability that consumer t does not buy the product in state X is

$$\Pr\{b_t = 0|X, l_t\} = 1 - F_C(\widehat{c}(l_t)) + F_C(\widehat{c}(l_t))F_U(p - X).$$

The first term is the probability of the event that consumer t does not search in state X; the second term is the probability of the event that consumer t searches in state X and gets a value of U that lies below the purchasing threshold. Using Bayes' rule, the public likelihood ratio in period $t+1$ is given by

$$l_{t+1}(b_t) = \begin{cases} \left[\dfrac{(1-F_U(p))}{1-F_U(p-H))}\right] l_t \quad \text{if } b_t = 1 \\ \left[\dfrac{1-F_C(\widehat{c}(l_t)) + F_C(\widehat{c}(l_t)))F_U(p))}{1-F_C(\widehat{c}(l_t)) + F_C(\widehat{c}(l_t)))F_U(p-H)}\right] l_t \quad \text{if } b_t = 0 \end{cases}. \tag{10}$$

If $l_t > \bar{l}$, then $F_C(\widehat{c}) = 0$ and $l_{t+1} = l_t$. Note that if $l_t < \underline{l}$, then $F_C(\widehat{c}) = 1$, and the "no buy" difference equation becomes linear with slope greater than 1.

To obtain comparative static results, we need to find conditions under which the posterior likelihood ratio following each action is monotone increasing in the prior likelihood ratio. As before, define

$$\begin{aligned} \psi_0(l_t) &= \frac{1-F_C(\widehat{c}(l_t)) + F_C(\widehat{c}(l_t)))F_U(p))}{1-F_C(\widehat{c}(l_t)) + F_C(\widehat{c}(l_t)))F_U(p-H)} \\ \psi_1(l_t) &= \frac{(1-F_U(p))}{1-F_U(p-H))}. \end{aligned}$$

Here ψ_1 does not depend upon F_C and therefore is invariant with respect to l_t. The reason is that consumers only purchase if they search, and search by itself is not informative about the state. Thus, the posterior likelihood ratio when the action is "buy" is strictly increasing (with slope less than 1) in the prior likelihood ratio. The monotonicity property does not generally hold when the action is "no buy". Observing this action is more discouraging when the consumer is more likely to search, and the consumer is more likely to search when l_t is low. This effect tends to make l_{t+1}decrease in l_t. On the other hand, starting from a higher initial belief results in a higher posterior belief, as else equal. That effect tends to make l_{t+1} increase in l_t. If we can ensure that the first effect is small, then overall l_{t+1} will increase in l_t. The following condition on the distribution of search costs can be shown to be sufficient for $\psi_0(l_t)l_t$ to be monotone increasing in $l_t \in (\underline{l}, \bar{l})$.

$$\text{A2'}: \; f(c) \leq \frac{(1-F_U(p))(1-F_U(p-H)}{(w(H)-w(0))(F_U(p)-F_U(p-H))}.$$

Assumption A2' imposes a bound on the increase in the probability of search that results from a given increase in l_t. The magnitude of the bound is determined by how informative it is when a consumer searches and then does not buy; that is, by the difference in the probability of purchase conditional on search in states H and L. Those probabilities are

$(1 - F_U(p))$ and $(1 - F_U(p - H)$, respectively. Note that the bound is decreasing in the first probability and increasing in the second (and thus decreasing in the difference). As H approaches L, the two probabilities are equal (no informativeness), the bound approaches infinity, and thus the condition is always satisfied.

6 Application

The basic setup of our model is mirrored nicely in a recent online experiment conducted by Salganik *et al.*[6] In the experiment, thousands of subjects were recruited to participate in artificial online music markets. Participants arrived sequentially and were presented with a list of 48 songs, which they could listen to, rate, and then download (for free) if they so chose. In real time, each participant was randomly assigned to one of nine "worlds." In the treatment worlds, of which there were eight, songs were listed by download rank: the first song listed was the one with the most downloads by previous participants in that same world, the second song listed had the second most downloads, and so on. In the control world, the 48 songs were shown in a random order, with no information about previous participants' listening, rating, or downloading behavior. The eight treatment worlds operated independently of one another, so that the researchers could observe eight separate realizations of the stochastic process.

As in our model, the products in these experiments were search goods. The songs were carefully screened to ensure that they would be unknown to the participants.[7] Choosing whether to sample a song is analogous to the decision of whether to search in our model. The cost of search in the experiment (i.e., the opportunity cost of the time spent listening to a song) was large enough that most participants listened to very few songs. Downloading a song (after listening to it) is analogous to the purchase decision in our model. Also, since participants assigned to treatment worlds were shown the number of downloads by previous participants, the information they received is essentially the same as in our model.

Table 1 summarizes the behavior of the participants in the experiment. On average, participants listened to fewer than four songs and downloaded fewer than two. The median

[6] See Salganik, Dodds, and Watts [15] for a brief but insightful analysis of the experiments; Salganik [14] provides a more thorough description and analysis.

[7] They were obtained from the music website *purevolume.com*, a website where aspiring bands can create homepages and post music for download. Bands that had played too many concerts or received too many hits on their homepages were excluded.

number of listens was 1 and the median number of downloads is zero.[8] Over 90% of participants listened to 10 songs or fewer, and roughly 40% exited the experiment without listening to any songs at all. Overall, the listening and downloading behavior of the participants in the treatment worlds is similar to that of the participants in the control world.

Table 2 summarizes the outcomes for the 48 songs. In the control world, the fraction of participants who listened to a given song was roughly equal across songs, ranging from 6% to 11%. In the treatment worlds, listening probabilities varied widely, with some songs in some worlds being listened to by more than 40% of participants. The higher variance of listening probabilities naturally translated to a higher variance in downloading probabilities. Hence, downloads were much more "skewed" in the treatment worlds than in the control world.

Table 2 also describes the distribution of the songs' *conditional* download probabilities—i.e., the probability that a participant downloaded the song conditional on listening to it. (SDW refer to these conditional probabilities as "batting averages".) These conditional probabilities provide the best measure of the songs' relative qualities. Unconditional download probabilities do not accurately reflect song quality because they conflate the probability of listening (which was highly variable across songs, and across worlds for a given song) with the conditional probability of downloading. On the other hand, conditional on listening to a song, the probability of downloading is clearly higher for songs with greater appeal. As shown in the table, quality varied substantially across songs: the conditional download probability was nearly 60% for the highest-quality song, and only 11% for the lowest-quality song.[9]

The distinction between listening and downloading is an important one. In our model, consumers are influenced by others' purchases only insofar as those purchases affect the decision to search. There are no social effects in the traditional sense: preferences are unaffected by previous consumers' purchases. We find clear support for this assumption

[8] It was not possible in the experiment to download a song without first listening to it.

[9] We use conditional download probabilities from the control world as our measure of song quality, because the probabilities from the treatment worlds may be tainted by a selection effect. For example, if in a treatment world a participant listens to a song with a very low download rank, it may indicate that something in the song's title (or the artist's name) was idiosyncratically appealing to the participant, which may increase the conditional probability of download. This possibility does not seem too important, however, as the conditional probabilities from the treatment worlds are generally very close to those from the control world.

in the experimental data. Participants in the treatment worlds were roughly 10 times more likely to listen to the top-ranked (i.e., most downloaded) song than any song ranked below 30. More formally, letting L_{jt} be an indicator variable equal to 1 if participant t listened to song j, we can ask whether L_{jt} is a function of song j's download share among participants $1, \ldots, t-1$. Table 3 reports the results from a probit regression in which L_{jt} is assumed to depend on song j's current download share (i.e., song j's share of total downloads by previous participants). Because some song titles and/or artist names might be more appealing than others on average, we include the song's listening share from the independent world as a control. The coefficient on download share is positive and highly significant, indicating that participants' decisions to listen to a song were sensitive to the information provided about previous participants' downloads.

In interpreting this result, one might argue that it partially reflects a framing effect. Because people generally tend to choose the first item when selecting from a list, the apparent influence of download information could be conflated with the impact of list position itself. Indeed, even in the control world, where songs were ordered randomly for each participant, participants were much more likely to listen to the first listed song. However, the download information is easily shown to have an effect above and beyond the effect that comes from list position. For example, if we include dummies for list position in the probit regression reported in Table 3, the coefficient on download share remains positive and significant. Even more tellingly, Salganik *et al* ran separate experiments in which download information was provided but songs were still randomly ordered; in these experiments, the provided information still had a substantial impact on listening and downloading probabilities (albeit not as large as in the experiments we analyze here).

By contrast, participants' decisions about downloading songs (after listening to them) did *not* appear to be influenced by the information provided. Letting D_{jt} be an indicator equal to 1 if participant t downloaded song j (conditional on listening to it), we can ask whether D_{jt} is a function of song j's download share among participants $1, \ldots, t-1$. This is a trickier question, however, because a probit regression of D_{jt} on download share involves an obvious reflection problem. (Songs with the most downloads will naturally have higher average download probabilities.) Fortunately, the conditional download probability from the control world is a natural control variable. When included, it forces the coefficient on download share to be identified from time variation in the download share relative to what it "ought" to be (as indicated by its download probability in the treatment world). Estimates of this model are reported in the second column of Table 3. The coefficient on download share is statistically indistinguishable from zero. We conclude that the information provided in the treatment worlds affected participants' listening decisions, but not their "preferences". In other words, participants tended to sample the songs that were most heavily downloaded by previous participants, but once they had listened to a song, the decision of whether to download it was essentially independent of previous participants' downloads.

Table 4 reports the distribution of listening probabilities among the last 100 participants in each of the eight worlds. We classified songs into three quality categories based on their batting averages in the control experiment. For each song, we simply calculated the fraction of participants who listened to the song, among the last 100 participants to arrive. The results for the best 16 songs indicate that outcomes in this category were highly unpredictable. In 53% of the cases, the listening probabilities were less than 5%, but in 13% of the cases they exceeded 20%. Outcomes were more predictable for the lower quality songs. The listening probabilities for most of these songs were less than 5%, with only 3% of songs obtaining listening probabilities above 10%. Recall that roughly 2% of the participants had essentially zero search costs and listened to all 48 songs. Consequently, in contrast to our model, we would not expect to see listening probabilities converge to zero in the experiment. Overall, therefore, we interpret the patterns as being broadly consistent with the predictions of our model.

Although the stochastic nature of the learning process makes it so that the market sometimes converges to the "wrong" outcome, in general the provision of information on previous consumer's decisions should make search more efficient. To test this in the experimental data, we can simply compare the listen rates and download rates for three different groups of participants: (1) those who were randomly assigned to the control world; (2) those who were assigned to a treatment world, and were among the *first* 100 participants to arrive; (3) those who were assigned to a treatment world, and were among the *last* 100 participants to arrive. Listening and downloading behavior among group 2 should be similar to group 1, since not much information has yet accumulated. However, we should expect search to be noticeably more efficient for the third group, since they observe substantial information on previous participants' downloads, and that information should in most cases guide them toward higher quality songs.

Table 5 reports listens per participant (total number of listens divided by total number of participants), and downloads per listen (total number of downloads divided by total number of listens) for each group. A comparison of the listening rates for the three groups suggests that consumers listen to fewer songs when they are more informed. The listening rates for the first 100 participants in the treatment world are lower than in the control world but higher than the listening rates for the last 100 participants. The download rate for the first 100 participants in the treatment group are essentially the same as the download rate in the control group but substantially lower than the download rate for the last 100 participants. The download rates increased by approximately 16%.

7 Conclusion

We have studied a simple choice problem in which consumers have to decide whether or not to consider a product of unknown utility. Consumers only purchase products in their consideration sets, but including a product in the set involves a small search cost. Consumers would prefer not to pay this cost if they believe they are unlikely to buy the product. The purchasing decisions of other consumers influences their beliefs about the gains from search. A poor purchasing record can feed on itself and lead consumers to wrongfully omit high quality products from their consideration sets. On the other hand, a good purchasing record can also feed on itself and lead consumers to include the product in their consideration sets. In this case, the market learns the quality of the product and the product obtains its true market share. The experimental study by Salganik et al provides evidence on the feedback mechanism and shows how it can affect outcomes. The results are largely consistent with our model.

However, their study also suggests that the dynamics of product demand are more complicated when the choice set consists of multiple products. Consumers typically search products sequentially according to the order in which they listed, downloading the songs they like, and stopping when the expected benefits of search exceed the cost. But preferences do not appear to be additive since the listening probability of a song that makes the considerations sets of most of the last 100 participants varies across the experiments. The implication is that long-run sales of high quality products that succeed and the probability that it will succeed are likely to depend upon the number of products that consumers want to purchase and the set of other high quality products that survive (i.e., have herds on S). It will also depend upon the kind of information the market provides. For example, in the market for recorded music, the market reports the sales ranks of albums but not their actual sales. We hope to explore the dynamic interactions that can occur when the choice set consists of more than one product in subsequent research.

We have largely ignored the issue of how firms may want to increase the likelihood that consumers will include their products in their consideration sets. Eliaz and Spiegler explore this issue in a static model in which firms employ costly marketing devices such as advertising to influence the formation of consumer consideration sets. They use the model to study whether firms can profitably exploit the bounded rationality of consumers and the impact of their rational behavior on product variety and marketing.

References

[1] Adler, M. "Stardom and Talent", *American Economic Review*, 75 (1), 1985, pp. 208-12.

[2] Banerjee, A. "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, v. 107 n. 3, August 1992, pp. 797-817.

[3] Becker, G., Grossman, M., and K. Murphy, "An Empirical Analysis of Cigarette Addiction," *American Economic Review*, v. 84 n. 3, June 1994, pp. 396-418.

[4] Becker, G., and Murphy, K., *Social Economics: Market Behavior in a Social Environment*, Harvard University Press, 2000.

[5] Bikhchandani, S., Hirshleifer, D. and Welch, I., "A Theory of Fads, Fashions, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy*, 100, October 1992, pp. 992-1026.

[6] Brock, W. and S. Durlauf, "Interactions-Based Models," in *Handbook of Econometrics*, vol. 5., J. Heckman and E. Leamer eds., Amsterdam: North-Holland, 2001.

[7] Brynjolfsson, E., Hu, Y., and D. Simester, "Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales," working paper, MIT Sloan School of Management, December 2006.

[8] Hongbin Cai, Y. Chen, and H. Fang, "Observational Learning: Evidence from a Randomized Natural Field Experiment", forthcoming, *American Economic Review*.

[9] Eliaz, K. and R. Spiegler, "Consideration Sets and Competitive Marketing," working paper, August 31, 2008.

[10] Hendricks, K. and A. Sorensen, "Information and Skewness of Music Sales," working paper, September, 2008.

[11] Manski, C. "Identification of Exogenous Social Effects: The Reflection Problem", Review of Economic Studies, 1993, 60(1), pp. 31-42.

[12] Pastine, I. and T. Pastine, "Signal Accuracy and Informational Cascades", CEPR Discussion Paper Series #5219, 2006.

[13] Rosen, S., "The Economics of Superstars", The American Economic Review, 71 (5), 1981, pp. 845-58.

[14] Salganik, M., "Success and Failure in Cultural Markets", Ph.D. thesis, 2007, Department of Sociology, Columbia University, New York.

[15] Salganik, M., P. Dodds, and D. Watts (2006), "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market," *Science*, 311, pp. 854-856.

[16] Salganik, M.and D. Watts, "Leading the Herd Astray: An Experimental Study of Self-Fulfilling Prophecies in an Artificial Cultural Market", working paper, 2008.

[17] Sorensen, A., "Bestseller Lists and Product Variety: The Case of Book Sales," forthcoming, *Journal of Industrial Economics.*

[18] Smith, L. and P. Sorensen, "Pathological Outcomes of Observational Learning", *Econometrica,* 68, 2000, pp. 371-398.

[19] Smith, L. and P. Sorensen, "Informational Herding and Optimal Experimentation", working paper, 2005.

Appendix

Proof of Lemma 2:

The continuity of ψ_0 and ψ_1 follows from the continuity of F_L, F_H, and $\widehat{\sigma}$. Since F_L and F_H satisfy the monotone likelihood ratio property, F_H first order stochastically dominates F_L, so $\psi_0(l) > 1$ and $\psi_1(l) < 1$ for $l \in [0, \bar{l})$. For $l \geq \bar{l}, F_H(\widehat{\sigma}(l)) = F_L(\widehat{\sigma}(l)) = 1$, and so $\psi_0(l) = 1$. Finally,

$$\begin{aligned}
\lim_{l \uparrow \bar{l}} \psi_1(l) &= \frac{1 - F_U(p)}{1 - F_U(p - H)} \lim_{l \uparrow \bar{l}} \left[\frac{1 - F_L(\widehat{\sigma}(l))}{1 - F_H(\widehat{\sigma}(l))} \right] \\
&= \frac{1 - F_U(p)}{1 - F_U(p - H)} \lim_{l \uparrow \bar{l}} \left[\frac{f_L(\widehat{\sigma}(l))\widehat{\sigma}'(l)}{f_H(\widehat{\sigma}(l))\widehat{\sigma}'(l)} \right] \\
&= \frac{1 - F_U(p)}{1 - F_U(p - H)} \frac{f_L(\widehat{\sigma}(\bar{l}))}{f_H(\widehat{\sigma}(\bar{l}))} > 0.
\end{aligned}$$

The last inequality follows from the fact that

$$\widehat{\sigma}'(l) = \frac{(w(H) - c)(c - w(0))}{[w(H) - c + ((c - w(0))l)]^2} \neq 0.$$

Q.E.D.

Proof of Lemma 3:

Smith and Sorensen's [19] Lemma 6 establishes that if f^γ is log-concave, then

$$\left[\frac{F_L(\widehat{\sigma}(l_t))}{F_H(\widehat{\sigma}(l_t))} \right] l_t$$

is increasing in l_t and

$$\left[\frac{1 - F_L(\widehat{\sigma}(l_t))}{1 - F_H(\widehat{\sigma}(l_t))} \right] l_t$$

is increasing in l_t. (That is, the posterior likelihood ratio after observing only whether or not consumer t searches increases with l_t.)

In our setting, the posterior likelihood ratio after observing $b_t = 1$ is

$$\psi_1(l_t) l_t = \frac{1 - F_U(p)}{1 - F_U(p - H)} \left[\frac{1 - F_L(\widehat{\sigma}(l_t))}{1 - F_H(\widehat{\sigma}(l_t))} \right] l_t.$$

Thus, $\psi_1(l_t) l_t$ is a scalar multiple of

$$\left[\frac{1 - F_L(\widehat{\sigma}(l_t))}{1 - F_H(\widehat{\sigma}(l_t))} \right] l_t,$$

and so is also increasing in l_t.

The posterior likelihood ratio after observing $b_t = 0$ is

$$\psi_0(l_t)l_t = \left[\frac{F_L(\widehat{\sigma}(l_t)) + (1 - F_L(\widehat{\sigma}(l_t)))F_U(p)}{F_H(\widehat{\sigma}(l_t)) + (1 - F_H(\widehat{\sigma}(l_t)))F_U(p-H)}\right] l_t.$$

Defining

$$\rho(l_t) = \frac{F_L(\widehat{\sigma}(l_t))}{F_H(\widehat{\sigma}(l_t)) + (1 - F_H(\widehat{\sigma}(l_t)))F_U(p-H)},$$

we can rewrite the previous equation as

$$\psi_0(l_t)l_t = \rho(l_t)\left[\frac{F_L(\widehat{\sigma}(l_t))l_t}{F_H(\widehat{\sigma}(l_t))}\right] + (1-\rho(l_t))\left[\frac{(1 - F_L(\widehat{\sigma}(l_t)))F_U(p)l_t}{(1 - F_H(\widehat{\sigma}(l_t)))F_U(p-H)}\right].$$

We know that the terms in brackets are increasing in l_t. Thus, if we can show that

$$(i) \quad \frac{F_L(\widehat{\sigma}(l_t))l_t}{F_H(\widehat{\sigma}(l_t))} > \frac{(1 - F_L(\widehat{\sigma}(l_t)))F_U(p)l_t}{(1 - F_H(\widehat{\sigma}(l_t)))F_U(p-H)}$$

and that (ii) $\rho(l_t)$ is increasing in l_t, then we have established that $\psi_0(l_t)l_t$ is increasing in l_t. To verify claim (i), note that because $f_L(\sigma) = 2(1-\sigma)f(\sigma)$ and $f_H(\sigma) = 2\sigma f(\sigma)$, we can rewrite

$$\frac{F_L(\widehat{\sigma}(l_t))}{F_H(\widehat{\sigma}(l_t))} = \frac{\int_0^{\widehat{\sigma}(l_t)} f_L(\sigma)d\sigma}{\int_0^{\widehat{\sigma}(l_t)} f_H(\sigma)d\sigma} = \frac{\int_0^{\widehat{\sigma}(l_t)} (1-\sigma)f(\sigma)d\sigma}{\int_0^{\widehat{\sigma}(l_t)} \sigma f(\sigma)d\sigma} = \frac{1 - E[\sigma|\sigma \le \widehat{\sigma}(l_t)]}{E[\sigma|\sigma \le \widehat{\sigma}(l_t)]}$$

and

$$\frac{1 - F_L(\widehat{\sigma}(l_t))}{1 - F_H(\widehat{\sigma}(l_t))} = \frac{\int_{\widehat{\sigma}(l_t)}^{\infty} f_L(\sigma)d\sigma}{\int_{\widehat{\sigma}(l_t)}^{\infty} f_H(\sigma)d\sigma} = \frac{\int_{\widehat{\sigma}(l_t)}^{\infty} (1-\sigma)f(\sigma)d\sigma}{\int_{\widehat{\sigma}(l_t)}^{\infty} \sigma f(\sigma)d\sigma} = \frac{1 - E[\sigma|\sigma > \widehat{\sigma}(l_t)]}{E[\sigma|\sigma > \widehat{\sigma}(l_t)]}.$$

Assumption A2(ii) then implies the claim. To check claim (ii), differentiating $\rho(l_t)$ and simplifying yields

$$\rho'(l_t) = \frac{F_U(p-H)f_H(\widehat{\sigma}(l_t))\widehat{\sigma}'(l_t)}{[F_H(\widehat{\sigma}(l_t)) + (1 - F_H(\widehat{\sigma}(l_t))F_U(p-H)]} > 0,$$

since

$$\widehat{\sigma}'(l_t) = \frac{(w(H)-c)(c-w(0))}{[w(H)-c+((c-w(0))l_t]} > 0.$$

Q.E.D.

Proof of Proposition 4:

Lemma 1 shows that the public likelihood ratio converges almost surely to a random variable l_∞. The only fixed points of the Markov process on the public likelihood rato (and thus the only possible values of l_∞) are $l = 0$ and $l \geq \bar{l}$. Given belief convergence and the monotonicity of the Markov process (Lemma 3 implies monotonicity for l between $\underline{l}$ and $\bar{l}$, the process is clearly monotonic below $\underline{l}$ and above $\bar{l}$, and the process is continuous at $\underline{l}$ and $\bar{l}$), Smith and Sorensen's [19] Lemma 12a establishes that the public likelihood ratio cannot enter the cascade set $l \geq \bar{l}$ from outside, and so if $l_0 < \bar{l}$, then either $l_\infty = 0$ or $l_\infty = \bar{l}$. Because l_∞ almost surely cannot be fully wrong, in state L, $l_\infty = \bar{l}$ with probability one. Consequently, in state L actions converge to N and so $\lambda = 0$. In state H, the argument of Smith and Sorensen's [18] Theorem 1d shows that the events $l_\infty = 0$ and $l_\infty = \bar{l}$ both have positive probability. In the former case, learning is complete, actions converge to S, and $\lambda = 1 - F_U(p - H)$. In the latter case, learning is incomplete, actions converge to N, and $\lambda = 0$. Q.E.D.

Proof of Propositions 6 and 7:

A limit cascade on N is the event that $l_\infty = \bar{l}$, which has probability

$$\Pr\{l_\infty = \bar{l}\} = \frac{l_0}{\bar{l}} = \frac{(c - w(0))(1 - \underline{d})}{(w(H) - c)\underline{d}} l_0 = l_0 \frac{(1 - \underline{d})\left(c - \int_p^\infty (-p + u) dF_U(u)\right)}{\underline{d}\left(\int_{p-H}^\infty (H - p + u) dF_U(u) - c\right)}.$$

Differentiating with respect to H, c, and p yields, respectively,

$$\frac{\partial \Pr\{l_\infty = \bar{l}\}}{\partial H} = \frac{-(1 - F_U(p - H))(c - w(0))(1 - \underline{d})}{(w(H) - c)^2 \underline{d}} l_0 < 0,$$

$$\frac{\partial \Pr\{l_\infty = \bar{l}\}}{\partial c} = \frac{(w(H) - w(0))(1 - \underline{d})}{(w(H) - c)^2 \underline{d}} l_0 > 0,$$

and

$$\frac{\partial \Pr\{l_\infty = \bar{l}\}}{\partial p} = \frac{(1 - F_U(p))(w(H) - c) + (1 - F_U(p - H)(c - w(0))(1 - \underline{d})}{(w(H) - c)^2 \underline{d}} l_0 > 0.$$

Further differentiating $\frac{\partial \Pr\{l_\infty = \bar{l}\}}{\partial c}$ with respect to c and H yields, respectively,

$$\frac{\partial^2 \Pr\{l_\infty = \bar{l}\}}{(\partial c)^2} = \frac{2(w(H) - c)(w(H) - w(0))(1 - \underline{d})}{(w(H) - c)^3 \underline{d}} l_0 > 0$$

and

$$\frac{\partial \Pr\{l_\infty = \bar{l}\}}{\partial c \partial H} = \frac{-(1 - F_U(p-H))[(c - w(0)) + (w(H) - w(0))](1 - \underline{d})}{(w(H) - c)^3 \underline{d}} l_0 < 0.$$

Proof of Proposition 8:

If the initial herd is on S (corresponding to $l_\infty = 0$), then subsequent buyers will choose action S (and then purchase with probability $1 - F_U(p - H)$ with or without the positive public signal. If the initial herd is on N (corresponding to $l_\infty = \bar{l}$ and $\lambda = 0$), then a positive public signal pushes teh public likelihood ratio below $\bar{l}$. The public likelihood ratio will reconverge, to $\bar{l}$ in state L but with positive probability to 0 (and positive long-run sales) in state H. Thus, in state H, a positive public signal raises expected sales when the initial herd is on N. Q.E.D.

Figure 1

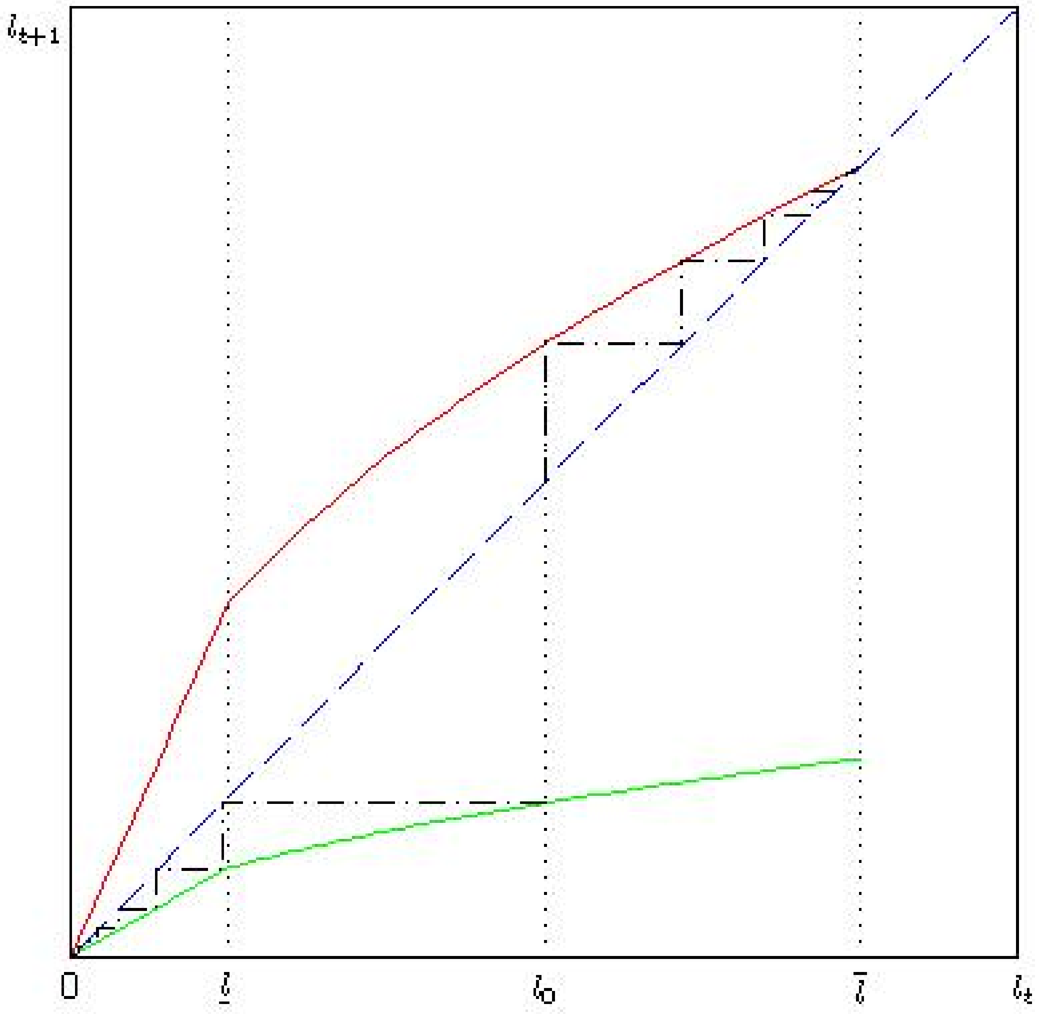

Figure 2