Social Learning with Rating Model

Yeochang Yoon*

Sunday 6^{th} April, 2014

Abstract

This paper explores rational social learning model in which people can observe ratings for products including the past purchase decisions. In this paper we will demonstrate how rating works in social learning model and investigate whether the additional rating information improves the learning or not. When individuals have heterogeneous preferences over choice, they are less sensitive to the others' decisions and two models eventually reveal the true state almost surely. On the other hand, in homogeneous case there is a positive probability of incorrect herding without rating because people have the same preferences and are more sensitive to others' decisions. Rating can prevent incorrect herding when quality of product is low, but not when it is high.

1 Introduction

Suppose people would like to purchase an experience good that they have not previously purchased. Since they have not experienced the product, they do not have enough information on the quality of the product. In order to know the quality, they can try to collect information they need. One possible way is to learn the information from other people by observing their behavior. For example, when readers choose a book, they can buy a best-seller because high number of sales implies that the book has been purchased by many people who believed that it has high quality. In this process, they can extract others' information by observing the number of sales. For this information, they

*Department of Economics, The Ohio State University, 410 Arps Hall, 1945 North High Street, Columbus, OH, yoon.250@osu.edu.

can visit on-line markets such as Amazon.com or Barnes and Nobles, and sort products by number of sales, referred to as popularity. However, when they visit these websites, they can also see the average ratings of the products. This is another way of collecting information and a type of social learning. The rating can provide information about how the previous buyers feel about the products. However, average rating and popularity provide sometimes different results, i.e., some popular products have the low ratings while less popular ones have the higher ratings. In this case, the readers might be confused about which provides the correct information. The main difference between popularity and rating is that purchase decision is made based on ex-ante utility of potential buyers while rating reflects ex-post utility of previous buyers. The study of observing actions as aggregator of information has been developed extensively, but the study of information conveyed by ratings has been less extensively examined although this is commonly used in real world nowadays.

In the past, it was difficult to observe user-generated ratings in large samples. Only professional reviews and small sample size of word-of-mouth were available. However, as internet service has now grown, we can easily give and observe ratings for products. Many internet retailers provide user-generated ratings for their customers. There are empirical studies that emphasized the importance of rating in purchase decisions. Chevalier and Mayzlin (2006) showed that average ratings have a significant impact on book sales. Li and Hitt (2008) examined that average ratings and product sales have the positive relationship. Dellarocas, Zhang and Awad (2007) found that user-generated reviews are good predictors for box office sales.

The social learning literature has emphasized the possibility of observing other people's actions to collect information. In the basic setting, infinite individuals are exogenously ordered and they take binary choices such as purchase or investment decisions. The decisions are made based on their own private signals and the past actions observed. Banerjee (1992) and Bikhchandani, Hirshleier and Welch (1992) investigated the herding where individuals may take the same actions by observing others' decisions and ignoring their own private information. When agents have the homogeneous preference in our model, the model goes back to Bikchandani et al. Smith and Sorensen (2000) explored how individuals learn sequentially from the discrete actions of others and how the public belief converges over time. When all individuals share the same public belief, it forces action convergence and herds arise. They showed that bounded private belief can cause the incorrect herding while we consider the case where homogeneity of individuals' preferences is the reason of the incorrect herding. Hendrick, Sorensen and Wiseman (2012) introduced costly search in social learning model and investigated the reason why superior products are sometimes ignored. In their model, consumers can purchase product only when they search it.

In this paper we develop a social learning model with rating. We investigate how rating works in a social learning model and whether it improves the learning or not. Moreover, throughout the paper, two different information environments for individuals are considered:

Popularity Information Environment: Individuals can observe the past action history of others' actions without rating.

Rating Information Environment: Individuals can leave ratings for the product based on their ex-post payoffs, and others can observe the history of ratings and past actions.

Moreover, we compare homogeneous and heterogeneous preference cases. When individuals have their own private preferences over a product, their preferences are heterogeneous and there might be lovers or haters of the product. Lovers always purchase the product while haters never purchase it. In homogeneous case, they have the same preference over the product and their payoffs simply depend on the quality of the product. These distinct preference conditions explain when popularity and rating provide the different belief convergences. The belief has the Markov-martingale character and is convergent. We find that beliefs eventually reveal the true quality of the product in both of information environments if individuals have heterogeneous preferences.

We show that incorrect herding may be due to the homogeneity of preference for a product. In heterogeneous preference case, individuals attach their own private values for the product and thus, they are less sensitive to others' actions and ratings. Agents with the extreme private values for the product can be bumpers for the incorrect herding. Learning is complete in the long run and incorrect herding will not occur in the heterogeneous preference case. On the other hand, homogeneous preference might cause incorrect belief convergence and herding. Agents are more sensitive to others' actions and ratings and thus, they are more likely to follow the majority. However, rating can prevent people from choosing undesirable decision when the quality of product is low while popularity cannot. Unfortunately, rating does not always guarantee the better learning. When the quality of the product is high, the incorrect herding is more likely to occur with rating even when rating is more accurate than the private signal.

The remainder of the paper is organized as follows. In Section 2, we builds up basic settings. Section 3 provides results in heterogeneous preference case and Section 4 investigates homogeneous preference case. Finally, extensions are considered in Section 5.

2 Model

A new product is introduced and its quality depends on an underlying state of the world. There are two states of nature which are payoff relevant, indexed by $\omega \in \{H, L\}$ and they are equally likely, $pr(\omega = H) = pr(\omega = L) = \frac{1}{2}$. In state H, the product provides higher quality, $u_H > u_L$. To simplify the notation and the exposition, we assume that $u_H = u > 0$ and $u_L = -u < 0$. An infinite sequence of exogenously ordered individuals, indexed by $n \in N$ sequentially takes one-shot a binary action which is not reversible, $x_n \in \{-1, 1\}$. We can consider $x_n = 1$ as purchase of the new product and $x_n = -1$ as not purchase.

Throughout the paper, we consider two different preference cases. First, individuals have heterogeneous preferences over the product, v_n which follows the cumulative distribution F with $supp(f) \supseteq [-u, u]$ and $E_{-n}(v_n) = 0$. This private value is independent with the state and known to the individual before purchase, but not to others. When we consider homogeneous preference case, we simply assume $v_n = 0$ for all n.

After purchase, individuals might realize that there are some parts which were not considered before purchase. This is captured by idiosyncratic shocks, ϵ_n which follows the cumulative distribution, G. We suppose that the idiosyncratic shocks are i.i.d. across individuals, $supp(g) \supseteq [-u, u]$, and $E(\epsilon_n) = 0$. The idiosyncratic shock is unknown before purchase, so that it does not affect their purchase decisions.

The payoff of individual n depends on the quality of the product he purchased, his own preference over the product and idiosyncratic shock. Hence, if the agent n chooses $x_n = 1$, his ex-post payoff is $u_\omega + v_n + \epsilon_n$ in heterogeneous preference case, and $u_\omega + \epsilon_n$ in homogeneous case, where $\omega \in \{H, L\}$.

On the other hand, if he chooses $x_n = -1$, the payoff is simply zero in both cases and it is always safe choice.

In contrast to the previous literature on social learning, we consider the information environment where individuals can observe ratings from the past buyers. Buyers can leave ratings for the product, $y_n \in \{-1, 1\}$, conditional on purchase of the product, $x_n = 1$. We only investigate the sincere feedback, i.e., $y_n = 1$ if the ex-post payoff of the individual is non-negative while $y_n = -1$ if it is negative. For the agent who did not purchase the item, y_n is zero. In addition, all buyers must leave their ratings.[1]

Throughout the paper two different information environments are considered:

Popularity Information Environment: For individual n, the past action history, $\{x_1, \ldots, x_{n-1}\}$ is the only information aggregator.

Rating Information Environment: Individuals can leave ratings for the product based on their ex-post payoffs, and others can observe ratings and past actions, $\{(x_1, y_1), \ldots, (x_{n-1}, y_{n-1})\}$.

The state ω is not observable. Instead, each individual n receives a private signal, $s_n \in \{-1, 1\}$ and forms a private belief. The accuracy of the private signal is q and the signal is independently generated across individuals conditional on the state, $q = pr(s_n = 1 \mid H) = pr(s_n = -1 \mid L) \in (\frac{1}{2}, 1)$. Hence, we refer the signal $s_n = 1$ as a good signal and $s_n = -1$ as a bad signal for state H.

Moreover, given the period n, individuals observe the history, and update the public belief to p_n. The agent n combines the pubic belief and his own private belief to the posterior belief r_n. Given the public belief p_n,

$$r_n(p_n, s_n = 1) = \frac{p_n q}{p_n q + (1 - p_n)(1 - q)} \text{ and } r_n(p_n, s_n = -1) = \frac{p_n(1 - q)}{p_n(1 - q) + (1 - p_n)q}.$$

Each individual makes his purchase decision, $x_n \in \{-1, 1\}$, depending on his expected payoff based on the posterior belief, r_n. When the expected payoff is zero, the individual chooses $x_n = 1$ with probability $\frac{1}{2}$ and $x_n = -1$ with probability $\frac{1}{2}$. [2]

[1] We will discuss the relaxation of these assumptions in Chapter 5.

[2] This decision is referred to as "flip a coin".

3 Heterogeneous Preference Case

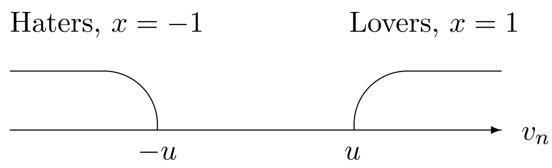
3.1 Convolution Formula

In heterogeneous preference case, individuals are faced on two unknowns when they collect information; others' private values and idiosyncratic shocks. However, distributions of these two variables are known, so that we can define $z_n \equiv v_n + \epsilon_n$, sum of two unknowns, which follows the cumulative distribution Ψ, where $\Psi(z_n \leq Z_n) = F(v_n \leq V)G(\epsilon_n \leq Z - V)$ since v_n and ϵ_n are independent with each other. Individuals' private values affect their purchase decisions because the private signals are known before buyers make their purchase decisions. On the other hand, rating is based on z_n which includes idiosyncratic shocks and prevents the complete extraction of the quality.

$$y_n = \begin{cases} 1 & \text{if } x_n = 1 \text{ and } u_\omega + z_n \geq 0 \\ -1 & \text{if } x_n = 1 \text{ and } u_\omega + z_n < 0 \end{cases}$$

3.2 Lovers and Haters

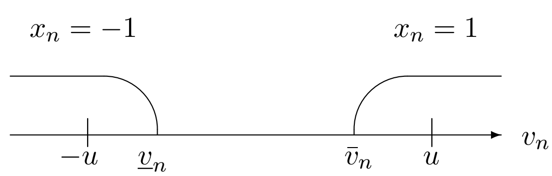
When individuals have heterogeneous preferences over the product and the support range of v_n is greater than quality of product in the state H, $supp(f) \supseteq [-u, u]$, there are lovers and haters of the product. Even though the agent certainly believes that the state is certainly H, $r_n = 1$, the hater never purchases the item whose $v_n < -u$. On the other hand, the agent with $v_n > u$ is lover of the item who always purchase regardless of their beliefs. However, when the negative idiosyncratic shock is significantly large, $\epsilon_n < -u - v_n$, even the lover would leave the negative rating. Lovers and haters do not follow their own signals and beliefs, and take the same actions in both states of nature. This is the reason why they are referred as crazy types in Smith and Sorensen (2000) and Chamley and Gale (2004).



[Figure 1]

3.3 Popularity Information Environment

When individuals decide whether to purchase the product or not, lovers and haters are not the only agents who do not follow their private own signals. When an agent receives low signal, given the public belief, the posterior belief is $r_n(p_n, s_n = -1)$. However, if the expected payoff from purchase is positive, $r_n(p_n,\, s_n = -1)(u + v_n) + [1 - r_n(p_n,\, s_n = -1)](-u + v_n) > 0$, the agent will choose $x_n = 1$ even if he received a bad signal. The cutoff of private value which satisfies this condition is defined as $\bar{v}_n \equiv [1 - 2r_n(p_n,\, s_n = -1)]$. Similarly we can define the lower cutoff, $\underline{v}_n \equiv [1 - 2r_n(p_n,\, s_n = 1)]$. Below the lower cutoff, the agents will choose $x_n = -1$ even after received a good signal. These two cutoffs depends on the current public beliefs while those of lovers and haters do not.



[Figure 2]

Lemma 3.1. *Agents with $v_n \in (\underline{v}_n,\, \bar{v}_n)$ follow their private signals, i.e., $x_n = 1$ if $s_n = 1$ and $x_n = -1$ if $s_n = -1$. Agents with $v_n \in (\bar{v}_n,\, \infty)$ ignore their private signals and choose $x_n = 1$, and those with $v_n \in (-\infty,\, \underline{v}_n)$ choose $x_n = -1$.*

Let $\mu_P(x_n, p_n)$ denote next period's public belief in popularity information environment when the current period's public belief is p_n and the agent chooses action x_n. If $x_n = 1$ is observed, it implies that the agent n's private value for the item is higher than the cutoff $\bar{v}_n$, or it is in $(\underline{v}_n, \bar{v}_n)$ and he received a good signal. Hence,

$$\mu_P(x_n = 1, p_n) = \frac{p_n\{[F(\bar{v}_n) - F(\underline{v}_n)]q + [1 - F(\bar{v}_n)]\}}{[1 - F(\bar{v}_n)] + [F(\bar{v}_n) - F(\underline{v}_n)][p_n q + (1 - p_n)(1 - q)]}.$$

Similarly, if $x_n = -1$, then it implies that the agent n's private value is lower than $\underline{v}_n$, or in $(\underline{v}_n, \bar{v}_n)$ and he received a bad signal. Hence,

$$\mu_P(x_n = -1, p_n) = \frac{p_n\{[F(\bar{v}_n) - F(\underline{v}_n)](1 - q) + F(\underline{v}_n)\}}{F(\underline{v}_n) + [F(\bar{v}_n) - F(\underline{v}_n)][p_n(1 - q) + (1 - p_n)q]}.$$

Using the transitions of public beliefs, we can define the public likelihood ratio $l_n \equiv \dfrac{1 - p_n}{p_n}$ that is the public belief for state L over that for H. When it converges to zero in H, we can say that the learning is complete. Similarly, when the inverse of l_n converges to zero in L, learning is complete. Let $\rho_P(x_n, l_n)$ denote the next period's public likelihood in popularity information environment when the current period's public likelihood ratio is l_n and the agent chooses action x_n, $\rho(x_n, l_n) = \dfrac{1 - \mu(x_n, p_n)}{\mu(x_n, p_n)}$. Hence,

$$\rho_P(x_n = 1, l_n) = l_n \frac{[1 - F(\bar{v}_n)] + [F(\bar{v}_n) - F(\underline{v}_n)](1 - q)}{[1 - F(\bar{v}_n)] + [F(\bar{v}_n) - F(\underline{v}_n)]q},$$

$$\text{and } \rho_P(x_n = -1, l_n) = l_n \frac{F(\underline{v}_n) + [F(\bar{v}_n) - F(\underline{v}_n)]q}{F(\underline{v}_n) + [F(\bar{v}_n) - F(\underline{v}_n)](1 - q)}.$$

These seem quite complicated, but if we see the expected value of the next period's likelihood ratio, then it is simply a martingale given state H. This is the similar result with Smith and Sorensen (2000). Since it is bounded below, $l_n \geq 0$ and a martingale process, l_n is convergent by Martingale Convergence Theorem. Moreover, it does not converge to ∞ by Fatou's lemma, which means that the fully incorrect learning almost surely cannot occur.

Lemma 3.2.

1. *The likelihood ratio process* $< l_n >$ *is a martingale conditional on state* H.

2. *Given state* H, *lim* $l_n < \infty$.

Proof.

$$\begin{aligned} E(l_{n+1}\,|\,l_n, H) &= prob(x_n = 1\,|\,l_n,\,H)\rho(x_n = 1,\,l_n) + prob(x_n = -1\,|\,l_n,\,H)\rho(x_n = -1,\,l_n) \\ &= \{[1 - F(\bar{v}_n)] + [F(\bar{v}_n) - F(\underline{v}_n)]q\}\rho(x_n = 1,\,l_n) \\ &+ \{F(\underline{v}_n) + [F(\bar{v}_n) - F(\underline{v}_n)](1-q)\}\rho(x_n = -1,\,l_n) \\ &= l_n\{[1 - F(\bar{v}_n)] + [F(\bar{v}_n) - F(\underline{v}_n)]q + [F(\bar{v}_n) - F(\underline{v}_n)](1-q)]\} = l_n \end{aligned}$$

□

3.4 Rating Information Environment

In rating information environment, buyers leave their feedbacks based on their ex-post payoffs, $u_\omega + z_n$, while their purchase decisions were made based on the expected value of $u_\omega + v_n$. Hence, for the later agents, both of F and Ψ involve when they update their beliefs in rating information environment. First, an agent whose private value is above $\bar{v}_n$, or in $(\underline{v}_n,\,\bar{v}_n)$ with a good signal purchases the product and leave a positive rating only when his ex-post payoff is positive. Let $\rho_R(y_n,\,l_n)$ denote the next period's public likelihood ratio when the current period's public likelihood ratio is l_n and the agent's rating is y_n. Then

$$\begin{aligned} &\rho_R((x_n = 1,\,y_n = 1),\,l_n) \\ &= l_n \frac{[1 - F(\bar{v}_n)][1 - \Psi(u\,|\,v_n \geq \bar{v}_n)] + [F(\bar{v}_n) - F(\underline{v}_n)](1-q)[1 - \Psi(u\,|\,\underline{v}_n < v_n < \bar{v}_n)]}{[1 - F(\bar{v}_n)][1 - \Psi(-u\,|\,v_n \geq \bar{v}_n)] + [F(\bar{v}_n) - F(\underline{v}_n)]q[1 - \Psi(-u\,|\,\underline{v}_n < v_n < \bar{v}_n)]}, \end{aligned}$$

$$\begin{aligned} &\rho_R((x_n = 1,\,y_n = -1),\,l_n) \\ &= l_n \frac{[1 - F(\bar{v}_n)]\Psi(u\,|\,v_n \geq \bar{v}_n) + [F(\bar{v}_n) - F(\underline{v}_n)](1-q)\Psi(u\,|\,\underline{v}_n < v_n < \bar{v}_n)}{[1 - F(\bar{v}_n)]\Psi(-u\,|\,v_n \geq \bar{v}_n) + [F(\bar{v}_n) - F(\underline{v}_n)]q\Psi(-u\,|\,\underline{v}_n < v_n < \bar{v}_n)}, \end{aligned}$$

and $\rho_R((x_n = -1,\,y_n = 0),\,l_n) = l_n \dfrac{F(\underline{v}_n) + [F(\bar{v}_n) - F(\underline{v}_n)]q}{F(\underline{v}_n) + [F(\bar{v}_n) - F(\underline{v}_n)](1-q)}$.

If the agent n chooses $x_n = -1$, it does not deliver any information about ex-post payoff from the product, so that $\rho_R((x_n = -1,\, y_n = 0),\, l_n) = \rho_P(x_n = -1,\, l_n)$. Using these transitions, we can get the expected value of next period's likelihood ratio given state H, and this is also a martingale and convergent.

Lemma 3.3.

1. *The likelihood ratio process* $< l_n >$ *is a martingale conditional on state* H*.*

2. *Given state* H*, lim* $l_n < \infty$*.*

Proof.

$$
\begin{aligned}
&E(l_{n+1} \,|\, l_{n,} H) = prob((x_n = 1,\, y_n = 1) \,|\, l_n,\, H)\varphi((x_n = 1,\, y_n = 1)\, l_n) \\
&+pr((x_n = 1,\, y_n = -1) \,|\, l_n,\, H)\varphi((x_n = 1,\, y_n = -1)\, l_n) \\
&+pr((x_n = -1,\, y_n = 0) \,|\, l_n,\, H)\varphi((x_n = -1,\, y_n = 0)\, l_n) \\
&= \{[1 - F(\bar{v}_n)][1 - \Psi(u \,|\, v_n \geq \bar{v}_n)] \\
&+[F(\bar{v}_n) - F(\underline{v}_n)]q[1 - \Psi(-u \,|\, \underline{v}_n < v_n < \bar{v}_n)]\}\varphi((x_n = 1,\, y_n = 1)\, l_n) \\
&+\{[1 - F(\bar{v}_n)]\Psi(-u \,|\, v_n \geq \bar{v}_n) + [F(\bar{v}_n) - F(\underline{v}_n)]q\Psi(-u \,|\, \underline{v}_n < v_n < \bar{v}_n)\}\varphi((x_n = 1,\, y_n = -1)\, l_n) \\
&+\{F(\underline{v}_n) + [F(\bar{v}_n) - F(\underline{v}_n)](1 - q)\}\varphi((x_n = -1,\, y_n = 0)\, l_n) = l_n
\end{aligned}
$$

□

Since the public likelihood ratios are convergent in both preference cases, we can find where they converge to.

Proposition 1. *Given state* H*,* $l_n \to 0$ *almost surely in both information environments.*

Proof. Suppose $l \neq 0$ in popularity information environment. If $l \neq 0$, then

$$
\begin{aligned}
\rho_P(x_n = 1,\, l) &= l\frac{[1 - F(\underline{v})] + [F(\bar{v}) - F(\underline{v})](1 - q)}{[1 - F(\underline{v})] + [F(\bar{v}) - F(\underline{v})]q} \neq l. \\
\text{and } \rho_P(x_n = -1,\, l) &= l\frac{F(\underline{v}) + [F(\bar{v}) - F(\underline{v})]q}{F(\underline{v}) + [F(\bar{v}) - F(\underline{v})](1 - q)} \neq l,
\end{aligned}
$$

since we assume $q > \frac{1}{2}$. Hence we can find the contradiction. Similarly, we can prove that $l_n \to 0$ almost surely in rating environment. □

Proposition 1 shows that in both information environments, the true state will be revealed almost surely in the long run because $l_n \to 0$ implies $p_n \to 1$ given state H. Therefore, the complete learning will occur almost surely when there is a private value for the product. This is also true when we consider the inverse of likelihood ratio in state L. Given state L, the inverse of public likelihood ratio is a martingale and it converges to zero, which means the complete learning. Therefore, in heterogeneous preference case, learning is complete and incorrect herding will not occur because the true state will be revealed in the long run.

4 Homogeneous Preference Case

Now, assume that all individuals have the same preference over the product and simplify $v_n = 0$ for all agents. In homogeneous preference case, the payoff from purchase is $u + \epsilon_n$ in H and $-u + \epsilon_n$ in L, where ϵ_n still follows G and $E(\epsilon_n) = 0$.

4.1 Popularity Information Environment

When agents are able to observe the history of past actions only in homogeneous preference case, the model goes back to Bikhchandani et al.(1992). In purchase decisions, the idiosyncratic shock is not considered, so that the expected quality only affects the agents' decisions. Hence, there are no lovers or haters in this case and all agents are in the same situation.

The optimal action for individual is exactly the same with Bikhchandani et al. If the public belief is in the range, $p_n \in (1-q, q)$, then their decisions are affected by the signals that they received. However, if the public belief is out of the range, $p_n > q$ or $p_n < 1-q$, the agents ignore their private signals and follow the majority.

Definition 1. *[Learning Region] If $p_n \in (1-q, q) \equiv \Im$, then x_n can fully convey the agent n's private signal information, and learning can occur.*

In popularity information environment, a good signal and a bad signal have the equal value, which means that when agents extract one good signal and one bad signal from the previous buyers' actions, these two actions are canceled out and thus, the public belief goes back to the initial one.

The order of signals also does not affect the posterior belief. In addition, the expected value of public likelihood ratio is a martingale.

Lemma 4.1.

1. *A good signal and a bad signal are canceled out by each other,*

$$\mu(x_n = 1,\, x_{n+1} = -1,\, p_n,\, p_{n+1} \in \Im) = \mu(x_n = -1,\, x_{n+1} = 1,\, p_n,\, p_{n+1} \in \Im) = p_n.$$

2. *Given state H, $< l_n >$ is a martingale.*

Since a good and a bad signal have the equal value and the order of signals does not affect their beliefs, agents' optimal decisions are determined by difference in number of good and bad signals that they received and extracted from others' actions. When the public belief is in the learning region, the agent's action fully delivers one signal. However, if the belief is q or $1-q$, then it might not deliver the full signal. This is because when $p_n = q$, the posterior belief of agent with a bad private signal is $\frac{1}{2}$ and thus, he chooses $x_n = 1$ with probability $\frac{1}{2}$. Hence, when $x_n = 1$ is observed given $p_n = q$, it is possible that the agent received a good signal, or received a bad signal and flipped a coin and thus, it does not deliver the full signal information. This also happens when $p_n = 1-q$ and $x_n = -1$. In addition, if $p_n > q$ or $p_n < 1-q$, then the agent n's decisions does not depend on the signal he received. For example, if $p_n > q$, then the posterior belief of agent with a bad signal is still greater than $\dfrac{1}{2}$, so that he chooses $x_n = 1$. Hence, if $p_n > q$ or $p_n < 1-q$, his decision does not deliver any information about his signal.

$$\begin{aligned}
\mu_P(x_n = 1,\, p_n \in [1-q,\, q)) &= \frac{p_n q}{p_n q + (1-p_n)(1-q)},\\
\mu_P(x_n = -1,\, p_n \in (1-q,\, q]) &= \frac{p_n(1-q)}{p_n(1-q) + (1-p_n)q},\\
\mu_P(x_n = 1,\, p_n = q) &= \frac{p_n(q + \frac{1}{2}(1-q))}{p_n(q+\frac{1}{2}(1-q)) + (1-p_n)(1-q+\frac{1}{2}q)},\\
\mu_P(x_n = -1,\, p_n = 1-q) &= \frac{p_n(1-q+\frac{1}{2}q)}{p_n(1-q+\frac{1}{2}q) + (1-p_n)(q+\frac{1}{2}(1-q))},\\
\text{and } \mu_P(x_n,\, p_n > q \text{ or } p_n < 1-q) &= p_n
\end{aligned}$$

In homogeneous preference case, whenever there is an imbalance of at least two between purchase and not purchase, it causes the herding on the majority choice and thus, the learning stops and cannot be completed. In figure 3, whenever the public belief is above q or below $1-q$, then it stops at these points. Therefore, the public likelihood ratio cannot converge to zero in this case and learning is not complete. Since given state H, the public likelihood ratio converges to $\frac{(1-q)(2-q)}{q(1+q)}$ or $\frac{q(1+q)}{(1-q)(2-q)}$. The first point is lower than one, which means that the public belief for state H is greater than that for state L. For this convergence, learning is not complete but incorrect herding will not occur. However, the second point is greater than one, which means that the public belief for state H is lower than that for state L. Hence, for this convergence, the incorrect herding will occur and thus, there is a positive probability of incorrect herding in popularity information environment.

In order to get the probability of incorrect herding, we use the recursive method. Let $\lambda_i(p_n, \omega)$ denote probability of incorrect herding in state ω when the current public belief is p_n, given information environment $i \in \{P, R\}$. The probability of incorrect herding is weakly increasing in the current public belief. When agents believe that it is more likely to be state H, they are more likely to purchase the product, so that the incorrect herding is less likely to occur. On the other hand, the incorrect herding is more likely to occur in state L when agents believe that it is more likely to be state H.

Lemma 4.2.

1. *$\lambda_i(p_n, H)$ is weakly decreasing in p_n.*

2. *$\lambda_i(p_n, L)$ is weakly increasing in p_n.*

Proof. See Appendix A. □

With the probability of incorrect herding function, we can see the exact probability of incorrect herding in popularity information environment. In prior, an agent follows the signal he receives and the probability of receiving a good signal in state H is q. After observing a purchase decision, later agents update their public belief to $\mu_P(x_n = 1, p_n = \frac{1}{2}) = q$. On the other hand, the agent receives a bad signal with probability $1-q$ in state H. After observing a not purchase decision, the later agents update their public belief to $\mu(x_n = -1, p_n = \frac{1}{2}) = 1-q$.

$$\lambda_P(p_n = \frac{1}{2},\, H) = q\lambda_P(p_n = q,\, H) + (1-q)\lambda_P(p_n = 1-q,\, H).$$

If we start from the public belief, $\mu_P(x_n = 1,\, p_n = \frac{1}{2}) = q$,

$$\lambda_P(p_n = q,\, H) = \tfrac{1}{2}(1+q)\lambda_P(\mu(x_n = 1, p_n = q),\, H) + \tfrac{1}{2}(1-q)\lambda_P(p_n = \frac{1}{2},\, H).$$

Since $\mu(x_n = 1, p_n = q) > q$, even if $(n+1)th$ agent receives a bad signal, $x_{n+1} = 1$. Hence, $\lambda_P(\mu(x_n = 1, p_n = q),\, H) = 0$. Similarly, $\lambda_P(\mu(x_n = -1, p_n = 1-q),\, H) = 1$.

$$\begin{aligned}
\lambda_P(p_n = q,\, H) &= \tfrac{1}{2}(1-q)\lambda_P(p_n = \frac{1}{2},\, H), \\
\text{and } \lambda_P(p_n = 1-q,\, H) &= (1-\tfrac{1}{2}q)\lambda_P(\mu(x_n = -1, p_n = 1-q),\, H) + \tfrac{1}{2}q\lambda_P(p_n = \frac{1}{2},\, H) \\
&= (1-\tfrac{1}{2}q) + \tfrac{1}{2}q\lambda_P(p_n = \frac{1}{2},\, H).
\end{aligned}$$

$$\text{Hence, } \lambda_P(p_n = \frac{1}{2},\, H) = \frac{(1-q)(2-q)}{2(1-q+q^2)}.$$

Similarly, we can get the probability of incorrect herding in state L in popularity information environment. It is the same with that in state L, because state H and L are symmetric in populiarity information environment.

Lemma 4.3.

1. *Given state H, l_n converges almost surely to a random variable l_P with support in the set* $\{\frac{(1-q)(2-q)}{q(1+q)}\} \cup \{\frac{q(1+q)}{(1-q)(2-q)}\}$.
2. $\lambda_P(p_n = \frac{1}{2},\, H) = \lambda_P(p_n = \frac{1}{2},\, L) = \frac{(1-q)(2-q)}{2(1-q+q^2)}$.

In heterogeneous preference case, agents are less sensitive to others' actions and ratings because all they have different private values. Their actions and ratings might come from different preferences. In homogeneous preference case, however, they are in the same situation, so more sensitive to others' decisions. This distinct preference assumption leads to the different learning dynamics and the possibility of incorrect herding. In popularity information environment, the learning cannot be completed and the incorrect herding might occur with the same probability in state H and L.

4.2 Rating Information Environment

When herding occurs in popularity information environment, their decisions do not deliver any information about their private signals and there is no learning. However, in rating information environment, buyers leave their ratings for the product, so that there is still learning about the quality of the product even when all agents choose $x = 1$ regardless of their private signals. In learning range, the agent n's purchase decision conveys both of his private signal and ex-post payoff information. Moreover, even if the agent simply follows the majority and purchases the item, his rating will deliver ex-post payoff information.

For rating information environment, we need an additional assumption on the distribution of idiosyncratic shock.

Assumption 4.1. *The distribution for idiosyncratic shock, G, is symmetric;*

$$G(x) = 1 - G(-x) \text{ for all } x \in (-M, M).$$

This assumption is not restrictive one because normal and uniform distributions satisfy the assumption. It also implies that $0 < G(-u) < \frac{1}{2} < G(u) < 1$.

In rating information environment, three different information sources are available for individuals; the previous agents' signal information and ratings, and their own private signals. Since the previous agents' signals and their own signals have the same accuracy, the relative accuracy between a private signal and rating determines the posterior public belief.

Definition 2. *[Relative Accuracy Condition] Let γ denote the relative accuracy of private signal over idiosyncratic shock, where*

$$\frac{q}{1-q} = [\frac{G(u)}{1-G(u)}]^{\gamma}.$$

When γ is higher, the private signal is relatively more accurate.

If one positive rating and one negative rating are observed, these two rating information is canceled out. However, since the rating is provided conditional on purchased, the purchase decision information remains and the public belief is the same with that after observing two purchase decisions in popularity information environment.

When there are more good signals extracted and positive ratings, or more bad signals and negative ratings, then the optimal action is clear. However, when there are more good signals and negative ratings, or more bad signals and positive ratings, it depends on the relative accuracy condition of private signal, γ.

Lemma 4.4.

1. *When there are one positive rating and one negative rating, then the opposite ex-post payoffs are canceled out and only signal information remains.*

$$\begin{aligned}&\mu_R((x_n = 1, y_n = 1),\, (x_{n+1} = 1, y_{n+1} = -1),\, p_n,\, p_{n+1} \in \Im)) \\ &= \mu_R((x_n = 1, y_n = -1),\, (x_{n+1} = 1, y_{n+1} = 1),\, p_n,\, p_{n+1} \in \Im)) \\ &= \mu_P((x_n = 1,\, x_{n+1} = 1), p_n, p_n + 1 \in \Im)\end{aligned}$$

2. *Given state H, $< l_n >$ is a martingale and given state L, the inverse is also a martingale.*

Does a negative rating reduce the public belief for state H? It depends on the relative accuracy of private signal, γ. If $\gamma > 1$, i.e., a private signal is relatively accurate, then a purchase decision increases the public belief more than negative ex-post payoff information.

$$\rho_R((x_n = 1,\, y_n = -1),\, p_n \in \Im) > p_n.$$

Even when only negative ratings are observed, the public belief is updated to higher if the purchase decision was made with a good signal. However, if $\gamma < 1$, we can find t which satisfies $\gamma \in (\frac{t-1}{t}, \frac{t}{t+1})$, and then t is the minimum number of negative ratings that occurs herding on $x = -1$ only with negative ratings. For example, when $t = 1$, only one nagative rating leads agents not to purchase the item. Moreover, $(t - 1)$ negative ratings and one bad signal occur herding on $x = -1$.

Lemma 4.5.

1. *Suppose $\gamma > 1$. Even with $y_1 = -1, \ldots, y_t = -1$ for $p_1, ..., p_t \in \Im$, if $s_{t+1} = 1$, then $x_{t+1} = 1$ since $\varphi((x_1 = 1,\, y_1 = -1), \ldots, (x_t = 1,\, y_t = -1),\, p_1 = \frac{1}{2}) > \frac{1}{2}$.*

2. *Consider $\gamma \in (\frac{t-1}{t}, \frac{t}{t+1})$. If $y_1 = -1, \ldots, y_t = -1$, then $x_{t+1} = -1$ even if $s_{t+1} = 1$.*

3. Consider $\gamma \in (\frac{t-1}{t}, 1)$*. If* $y_1 = -1, \ldots, y_{t-1} = -1$*,* $y_t = 0$*, then* $x_{t+1} = -1$ *even if* $s_{t+1} = 1$*.*

In homogeneous preference case, Bayesian updating depends on the current public belief regions. When the current public belief is in learning region, the agent's purchase decision and rating deliver both of his private signal and ex-post payoff information.

$$\begin{aligned}
\mu_R((x_n = 1,\, y_n = 1),\, p_n \in \Im) &= \frac{p_n q G(u)}{p_n q G(u) + (1-p_n)(1-q)G(-u)}, \\
\mu_R((x_n = 1,\, y_n = -1),\, p_n \in \Im) &= \frac{p_n q G(-u)}{p_n q G(-u) + (1-p_n)(1-q)G(u)}, \\
\text{and } \mu_R((x_n = -1,\, y_n = 0),\, p_n \in \Im) &= \frac{p_n(1-q)}{p_n(1-q) + (1-p_n)q}.
\end{aligned}$$

On the other hand, when the current belief is on the boundary of learning region, it does not deliver the full private signal information. When the agent recevies a bad signal given $p_n = q$, he still chooses $x_n = 1$ with probability 1/2.

$$\begin{aligned}
\mu_R((x_n = 1,\, y_n = 1), p_n = q) &= \frac{p_n \frac{1}{2}(1+q)G(u)}{p_n \frac{1}{2}(1+q)G(u) + (1-p_n)(1-\frac{1}{2}q)G(-u)}, \\
\mu_R((x_n = 1,\, y_n = -1),\, p_n = q) &= \frac{p_n \frac{1}{2}(1+q)G(-u)}{p_n \frac{1}{2}(1+q)G(-u) + (1-p_n)(1-\frac{1}{2}q)G(u)}, \\
\text{and } \mu_R((x_n = -1,\, y_n = 0),\, p_n = q) &= \frac{p_n(1-q)}{p_n(1-q) + (1-p_n)q}. \\
\mu_R((x_n = 1,\, y_n = 1), p_n = 1-q) &= \frac{p_n q G(u)}{p_n q G(u) + (1-p_n)(1-q)G(-u)}, \\
\mu_R((x_n = 1,\, y_n = -1),\, p_n = 1-q) &= \frac{p_n q G(-u)}{p_n q G(-u) + (1-p_n)(1-q)G(u)}, \\
\text{and } \mu_R((x_n = -1,\, y_n = 0),\, p_n = 1-q) &= \frac{p_n(1-\frac{1}{2}q)}{p_n(1-\frac{1}{2}q) + (1-p_n)\frac{1}{2}(1+q)}.
\end{aligned}$$

On the other hand, when the current belief is out of the learning region, $p_n > q$ or $p_n < 1-q$, the agent's purchase decision does not provide any private signal information, but ex-post payoff information is still observed conditional on purchased. In popularity information environment, the learning stops at this region while in rating information environment, it does not.

$$\begin{aligned}
\mu_R((x_n = 1,\, y_n = 1),\, p_n > q) &= \frac{p_n G(u)}{p_n G(u) + (1 - p_n) G(-u)}, \\
\mu_R((x_n = 1,\, y_n = -1),\, p_n > q) &= \frac{p_n G(-u)}{p_n G(-u) + (1 - p_n) G(u)}, \\
\mu_R((x_n = -1,\, y_n = 0),\, p_n > q) &= p_n, \\
\text{and } \mu_R((x_n = 1,\, y_n = 1),\, p_n < 1 - q) &= \mu((x_n = 1,\, y_n = -1),\, p_n < 1 - q) \\
= \mu_R((x_n = -1,\, y_n = 0),\, p_n < 1 - q) &= p_n.
\end{aligned}$$

Using the Bayesian updating, we can find the transition functions of public likelihood ratio and it is also a martingale stochastic process. The probability of incorrect herding in rating information environment, is denoted by $\lambda_R(p_n, \omega)$. The convergence is different with the heterogeneous preference case because the Bayesian updating functions of public belief depend on which regions the current beliefs are located in. In some regions, the public beliefs are not changed by any actions, and remain the same. Hence, in these regions, no further learning occurs. In state H, learning might not be complete and there is a positive probability of incorrect herding. However, learning still occurs when agents herd on $x = 1$. In state H, buyers are likely to leave positive ratings and learning can be complete. On the other hand, whenever some agents purchase the product in state L, they are likely to give negative ratings and thus, the incorrect herding can be prevented in state L although learning is not complete.

Proposition 2.

1. *Given state H, l_n converges almost surely to a random variable l_H with support in the set* $\{0\} \cup (\dfrac{q}{1-q}, \infty)$.
2. *Given state L, $1/l_n$ converges almost surely to a random variable l_L with support in the interval* $[0, \dfrac{1-q}{q})$.

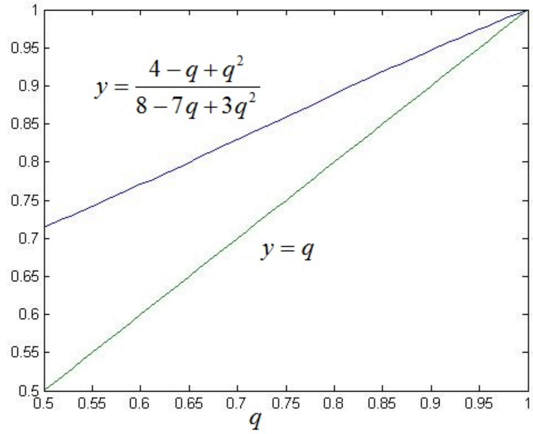
Proof. See Appendix A. □

When herding on $x = 1$ is not correct, it can be stopped when some agents purchase the item and give negative ratings while herding on $x = -1$ does not provide the further information. We find that rating information environment can prevent agents from incorrect herding in state L. However, unfortunately there is still a positive probability of incorrect herding in state H although the learning might be completed by rating. Moreover, the probability of incorrect herding might be even higher in rating information environment under some conditions.

In popularity information environment, the probability of incorrect herding is the same in state H and L. However, in rating information environment, it depends on states. More specifically, the probability of incorrect herding in state L is zero because the inverse of the public likelihood ratio converges to the point which is lower than one. Wrong decisions can be prevented in the long run by rating when the quality of product is low. Even if some buyers purchase the low quality product, they provide their negative ex-post payoff information and it can prevent wrong decisions of later agents. However, in state H, the public likelihood ratio converges to either zero or the point which is greater than one. This implies that learning might be complete in state H, but if it is not complete, the incorrect herding might occur.

In state H, there is a positive probability of incorrect herding in both of popularity information environment and rating information environment. If rating information is noisy and not informative, then agents tend to ignore rating information and use the previous agents' purchase decision and their own private signals only. Since they use the same information with popularity information environment, the probability of incorrect herding is the same in both information environments. On the other hand, if rating information is perfectly accurate, the previous agents' rating are simply able to reveal the state and learning is complete and the probability of incorrect herding is lower with rating. However, since there is still chance of not receiving rating information when the previous agents did not purchase the product, the probability of incorrect herding is not zero.

More interestingly, the probability of incorrect herding might be higher with rating even when rating information is quite accurate. When $\gamma < 1$, people tend to rely on rating information more than on other information sources. In this case, a single negative rating might have more value than others' purchase decisions or their own private signals, and it leads agents to safe choice. However, if rating is not significantly accurate, this situation frequently occurs. Hence, $G(u) \in (q, \dfrac{4-q+q^2}{8-7q+3q^2})$ is the sufficient condition for that the incorrect herding is more likely to



[Figure 3] $G(u) \in (q, \, \dfrac{4-q+q^2}{8-7q+3q^2})$

occur in rating information environment, given state H.

Proposition 3.

1. $\lambda_P(p_n = \frac{1}{2}, L) \geq \lambda_R(p_n = \frac{1}{2}, L) = 0.$
2. *If* $G(u) = 1$, *then* $\lambda_P(p_n = \frac{1}{2}, H) \geq \lambda_R(p_n = \frac{1}{2}, H)$
3. *If* $G(u) = \frac{1}{2}$, *then* $\lambda_P(p_n = \frac{1}{2}, H) = \lambda_R(p_n = \frac{1}{2}, H)$
4. *If* $G(u) \in (q, \, \frac{4-q+q^2}{8-7q+3q^2})$, *then* $\lambda_P(p_n = \frac{1}{2}, H) < \lambda_R(p_n = \frac{1}{2}, H)$.

Proof. See Appendix A. □

5 Extensions

A large social learning literature has focused on observing others' actions as aggregation of information. Observing actions allows agents to collect more information and learn the state of the world. However, as technology is now well-developed, buyers can share their ex-post payoff information. In this paper, we studied a social learning model where agents can observe not only others' actions, but also their ratings. This is another aggregator to provide further information. We compared two different preference cases over a product. First, when agents have heterogeneous preferences over the product, the true state is revealed in both information environments. However, when they have the homogeneous preferences, the public beliefs might converge to different points in two information environments. Rating tends to lead agents to choose the safe one when the quality of product is low. We found that the additional rating information does not guarantee better learning in a homogeneous preference case.

The analysis can be extended in three directions; first to relax the sincere rating assumption, second to allow buyers not to leave their ratings, and third to endogenize agents' decision timing. The first extension is allowing manipulation in rating. Throughout the paper, we assumed all buyers leave positive ratings whenever their ex-post payoffs are non-negative. However, there is a possibility of manipulation or opinion spams in real world, and it is difficult for agents to distinguish between manipulation and sincere rating. The manipulation will be a kind of noise for agents and reduce the accuracy of rating information. Second, the model assumed that all buyers leave their ratings, but it might be unrealistic. However, even if we relax the assumption, we will get similar results in the long run. The relaxation causes the delay in convergence, but not the different convergent point. Finally, the analysis also can be extended to endogenous timing model. In this paper, agents are exogenously ordered and delays are not allowed. If endogenous timing is introduced and the number of agents is finite, then we expect to see the dynamics of purchase decisions. Li and Hitt (2008) show that the average rating tends to decrease over time. This trend can be explained in an endogenous timing model. In the heterogeneous preference case, there are lovers and haters who take same actions in both states. If the endogenous timing is allowed and there is a discounting factor, then lovers tend to purchase the item earlier and haters never purchase it. The existence of lovers and haters leads higher ratings at the beginning. The rating, however, will decrease over time as agents in middle range make their decisions, and this can support the empirical data which is described in Li and Hitt (2008).

References

[1] Banerjee, A. V. (1992): "A Simple Model of Herd Behavior," Quarterly Journal of Economics, 107, 797-817.

[2] Banerjee, A. V., and D. Fudenberg (2004): "Word-of-mouth Learning," Games and Economic Behavior, 46, 1-22.

[3] 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.

[4] Dellarocas, C., X. Zhang, and N. F. Awad (2007): "Exploring the Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures," Journal of Interactive Marketing, 21, 23-45.

[5] Doob, J. L. (1953): Stochastic Procss, New York: Wiley.

[6] Harper, F. M., X. Li, Y. Chen, and J, Konstan (2005): "An Economic Model of User Rating in an Online Recommender System," Springer's LNAI series, UM 2005 User Modeling: Proceedings of the Tenth International Conference.

[7] Hendrick, K., A. Sorensen, and T. Wiseman (2012): "Observational Learning and Demand for Search Goods," American Economic Journal: Microeconomics, 4, 1-31.

[8] Lafky, J. (2010): "Why Do People Write Reviews? Theory and Evidence on Online Ratings," Ph.D. thesis, University of Pittsburgh.

[9] Li, X., and L. M. Hitt (2008): "Self-Selection and Information Role of Online Product Reviews," Information Systems Research, 1-19.

[10] Smith, L. and P. Sorensen (1998): "Pathological Outcomes of Observational Learning," Econometrica, 68, 371-398.

Appendix

Lemma 4.2

Assume that $\lambda(p_n, H)$ is not weakly decreasing in p_n. Then, we can find p_0 and p_1 s.t. $p_0 > p_1$ and $\lambda(p_0, H) > \lambda(p_1, H)$. Since $p_0 > p_1$, we can define a signal with accuracy $\sigma_1 > \dfrac{1}{2}$ s.t.

$$p_0 = \frac{p_1\sigma_1}{p_1\sigma_1 + (1-p_1)(1-\sigma_1)}.$$

Let $p_2 \equiv \dfrac{p_1(1-\sigma_1)}{p_1(1-\sigma_1)+(1-p_1)\sigma_1} < p_1$. Hence, $\lambda(p_1, H) = \sigma_1\lambda(p_0, H) + (1-\sigma_1)\lambda(p_2, H)$. Since $\sigma_1 > \dfrac{1}{2}$ and $\lambda(p_0, H) > \lambda(p_1, H)$, $\lambda(p_1, H) > \lambda(p_2, H)$ and $p_1 > p_2$. Continuously, $p_1 > p_2 > p_3 > \ldots$ and $\lambda(p_1, H) > \lambda(p_2, H) > \lambda(p_3, H) > \ldots$. Ultimately, $\lambda(p_0, H) > \lambda(0, H)$. But, when $p = 0$, all agents choose $x = -1$. Hence, $\lambda(0, H) = 1$. There is a contradiction.

Proof for Proposition 2

1. Given $p_n \in \Im$,

   $\rho_R((x_n = 1, y_n = 1), p_n \in \Im) = l_n \dfrac{1-q}{q}\dfrac{G(-u)}{G(u)}$.

   $\rho_R((x_n = 1, y_n = -1), p_n \in \Im) = l_n \dfrac{1-q}{q}\dfrac{G(u)}{G(-u)}$.

   $\rho_R((x_n = -1, y_n = 0), p_n \in \Im) = l_n \dfrac{q}{1-q}$.

2. Given $p_n = q$,

   $\rho_R((x_n = 1, y_n = 1), p_n = q) = l_n \dfrac{1-\frac{1}{2}q}{\frac{1}{2}(1+q)}\dfrac{G(-u)}{G(u)}$.

   $\rho_R((x_n = 1, y_n = -1), p_n = q) = l_n \dfrac{1-\frac{1}{2}q}{\frac{1}{2}(1+q)}\dfrac{G(u)}{G(-u)}$.

   $\rho_R((x_n = -1, y_n = 0), p_n = q) = l_n \dfrac{q}{1-q}$.

3. Given $p_n = 1-q$,

   $\rho_R((x_n = 1, y_n = 1), p_n = 1-q) = l_n \dfrac{1-q}{q}\dfrac{G(-u)}{G(u)}$.

   $\rho_R((x_n = 1, y_n = -1), p_n = 1-q) = l_n \dfrac{1-q}{q}\dfrac{G(u)}{G(-u)}$.

   $\rho_R((x_n = 1, y_n = 0), p_n = 1-q) = l_n \dfrac{1+q}{1-\frac{1}{2}q}$.

4. Given $p_n > q$,

   $\rho_R((x_n = 1,\ y_n = 1),\ p_n > q) = l_n \dfrac{G(-u)}{G(u)}$.

   $\rho_R((x_n = 1,\ y_n = -1),\ p_n > q) = l_n \dfrac{G(u)}{G(-u)}$.

   $\rho_R((x_n = -1,\ y_n = 0),\ p_n > q) = l_n$.

5. Given $p_n < 1 - q$,

   $\rho_R(x_n = 1,\ y_n = 1),\ p_n < 1 - q) = \rho_R(x_n = 1,\ y_n = -1),\ p_n < 1 - q)$

   $= \rho_R(x_n = -1,\ y_n = 0),\ p_n < 1 - q) = l_n$

Let us consider state H first. In the closure of learning region, l_{n+1} cannot be the same with l_n unless $l_n = 0$. Since $p_n \in [1 - q,\ q]$ in the closure of learning region, the public likelihood ratio does not converge to the closure of learning region. On the other hand, if $p_n > q$, the public likelihood ratio converges to zero. If $p_n < 1-q$, learning can be also stopped. When $x_n = -1$ and $x_{n+1} = -1$ given $p_n = 1/2$, $l_{n+2} = \dfrac{q(1+q)}{(1-q)(2-q)}$ and the public likelihood ratio will not be changed by any actions. In state L, the inverse of public likelihood ratio does not converge to the closure of learning region neither. If $p_n > q$, the inverse cannot converge to zero since it means that $p_n = 0$. Hence, the inverse can converge only to the region where $p_n < 1 - q$.

Proof for Proposition 3. 4

Suppose that $G(u) > q$. Then $\mu_R(x_n = 1, y_n = -1) < \dfrac{1}{2}$.

$$\begin{aligned}\lambda_R(H) \;&= qG(u)\lambda_R(\mu_R(x_n = 1, y_n = 1),\ H) + qG(-u)\lambda_R(\mu_R(x_n = 1, y_n = -1),\ H)\\ &+(1-q)\lambda_R(\mu_R(x_n = -1, y_n = 0),\ H),\end{aligned}$$

where

1. $\lambda_R(\mu_R(x_n = 1, y_n = 1),\ H) \geq 0$.

2. $\lambda_R(\mu_R(x_n = 1, y_n = -1),\ H) > \lambda_R(H)$, since $\mu_R(x_n = 1, y_n = -1) < \dfrac{1}{2}$.

3. $\lambda_R(\mu_R(x_n = -1, y_n = 0),\ H) = (1-\dfrac{1}{2}q)\lambda_R(\mu_R((x_n = -1, y_n = 0), (x_{n+1} = -1, y_{n+1} = 0),\ H)$ $+\ \dfrac{1}{2}q[G(u)\lambda_R(\mu_R((x_n = -1, y_n = 0), (x_{n+1} = -1, y_{n+1} = 0),\ H) + G(-u)\lambda_R(\mu_R((x_n =$

$-1, y_n = 0), (x_{n+1} = 1, y_{n+1} = -1),\ H)]$.

4. $\lambda_R(\mu_R((x_n = -1, y_n = 0), (x_{n+1} = -1, y_{n+1} = 0)),\ H) = 1$.

5. $\lambda_R(\mu_R((x_n = -1, y_n = 0), (x_{n+1} = 1, y_{n+1} = 1)),\ H),\ H) \geq 0$.

6. $\lambda_R(\mu_R((x_n = -1, y_n = 0), (x_{n+1} = 1, y_{n+1} = -1)),\ H),\ H) \geq \lambda_R(H)$.

Hence, $[1 - qG(-u) - \frac{1}{2}q(1-q)G(-u)]\lambda_R(H) > (1-q)(1-\frac{1}{2}q) + (1-q)\frac{1}{2}qG(-u)$, which implies

$$\lambda_R(H) > \frac{(1-q)(1-\frac{1}{2}q) + (1-q)\frac{1}{2}qG(-u)}{[1 - qG(-u) - \frac{1}{2}q(1-q)G(-u)]}.$$

Since $\lambda_P(H) = \frac{(1-q)(1-\frac{1}{2}q)}{1-q+q^2}$, if $\frac{(1-q)(1-\frac{1}{2}q) + (1-q)\frac{1}{2}qG(-u)}{[1 - qG(-u) - \frac{1}{2}q(1-q)G(-u)]} > \frac{(1-q)(1-\frac{1}{2})}{1-q+q^2}$, then

$$\lambda_R(H) > \lambda_P(H).$$

Therefore, if $G(u) \in (q, \frac{4-q+q^2}{8-7q+3q^2})$, then $\lambda_R(H) > \lambda_P(H)$. Since $q \leq \frac{4-q+q^2}{8-7q+3q^2}$, the region always exists.