

# Fragility of Confounded Learning

Xuanye Wang \*

Institute for Advanced Economics Research, Dongbei University of Finance and Economics

June 9, 2021

## Abstract

We consider an observational learning model with exogenous public payoff shock. We show that confounded learning doesn't arise for almost all private signals and almost all shocks, even if players have sufficiently divergent preferences.

Keywords: Social Learning, Information Aggregation, Confounded Learning.

JEL Classification: C11, D83

## 1 Introduction

People herd in daily lives. Everyone goes to their friends' favorite restaurants. College students are more likely to register the same classes as their familiar senior students do. Intuitively, one may think such herding behavior as blind and irrational.

In contrast to intuition, Banerjee (1992) and Bikhchandani et al. (1992) point out that herding behavior can be rational. In many circumstances, people's payoff are affected by an unknown underlying state. For example, the payoff of choosing one restaurant depends on the underlying quality of the chef. Decisions made by previous choice-makers reveal valuable private information about the underlying state. Thus, one could rationally herd as long as one thinks the observed previous choices contain enough information.

These two seminal works start a large literature named observational learning. This literature focuses on circumstances where learning could only happen through observing

---

\*Email: xuanye.wang@dufe.edu.cn.

others' actions. In these circumstances, information about the underlying state is distributed among a large group of people. Each person's private information cannot be credibly or precisely transmitted. People guess others' private information behind observed actions, and then combine his/her own private information with these guesses to make decisions. One learns from others' actions, and one's own action reveals one's private information that others can learn. The key question is whether long run efficiency is achieved, whether people eventually assign all the weight to the true underlying state among all alternatives.

The answer depends on several important primitives of modeling: (1) the observing structure-whose actions one person could observe; (2) the signal structure-what kind of private information one person could get; (3) the payoff structure-what payoff one person could receive. Lots of important papers study different variations of these primitives and provide excellent answers. To name a few, Smith and Sørensen (2000) study the classic observing structure that each player observes all previous actions in order. They find that long run efficiency critically depend on whether players can get arbitrarily precise signal about the underlying states. If they do, then any wrong herd is eventually overthrown by some player with sufficiently precise information. Correct learning must happen. Çelen and Kariv (2004) study the observing structure that each player just observe the action of his/her immediate predecessor. They find that people's choices and beliefs cycle. In other words, learning needs not to happen. Mossel et al. (2015) place strategic players on a directed social network and let them take actions repeatedly. They assume a player  $A$  could only observe the action of player  $B$  iff a directed edge connected them. They find that long run efficiency occurs as long as the number of neighborhoods of every player is bounded above and each player can observe his/her observers by a chain with bounded length.

We think one simple but important variation is missed in the literature-the payoff structure may be subject to exogenous shocks in the learning process. For example, the taste of the main course could be different in different seasons for the reason that some key ingredients tastes worse in winter. The benefits of taking the same course may change from year to year as the instructor changes.

Based on this consideration, we introduce exogenous public payoff shocks into observational learning models. There are certainly many different ways to model payoff shocks. In this article we consider the most simple one. We assume there are two infinite sequence of periods. In one sequence, the payoff of one action is changed by a payoff shock with value  $c$ . Even such a simple structure of payoff shocks makes a difference: it makes confounded learning fragile even if players have sufficiently divergent preferences.

To illustrate the idea, let us briefly review the concepts of learning and confounded learning. The first thing we would like to mention is that the usage of the word “long run learning” in observational learning literature is different from the usage in daily life and possibly other literatures. A “long run learning belief” is often just a stationary belief over all potential states and needs not to assign all the weight to the true state. Reaching a long run learning belief doesn’t mean efficiency-people need not to assign all the weight to the true state. The well-known “wrong herd” in Banerjee (1992) is associated with a long run learning belief which assigns sufficient weight to the wrong state. One of the many contributions of Smith and Sørensen (2000) is that inefficiency at a long run learning belief is not necessarily associated with wrong herd-the learning could be confounded. The key feature of confounded learning is its inconclusiveness-the previous actions don’t provide overwhelming evidences about which state is true, and each person still has to use their own private information to decide. A sharp reader may question such a belief cannot be stationary: since each person still use private information to decide, his/her action contains certain amount of information; upon observing such an informative action, the belief must be updated and cannot stay the same! The smart point that Smith and Sørensen (2000) point out is that an observable action needs not to be informative even if the action-taker uses private information to decide. Actions are uninformative as long as the probability of the observed action is the same across different states. Mathematically, it means

$$\Pr(b|\lambda^*, A) = \Pr(b|\lambda^*, B), \tag{1}$$

the probability of observing action  $b$  under confounded belief  $\lambda^*$  is the same across true state  $A$  and alternative state  $B$ . Of course equation 1 is just necessary but not sufficient for a belief to be confounded. A confounded belief must also be moderate in the sense that this belief doesn’t dominate one’s private information. Smith and Sørensen (2000) shows that confounded learning robustly arise in the sense that equation 1 robustly has moderate solutions.

The mathematics behind confounded learning changes after we introduce the simple public payoff shock. Now for a belief  $\lambda^*$  to be confounded, any observable action must be uninformative under belief  $\lambda^*$  with and without shock  $c$ . Mathematically, that means the following equation system

$$\begin{aligned} \Pr(a|\lambda^*, A, 0) &= \Pr(a|\lambda^*, B, 0), \\ \Pr(a|\lambda^*, A, c) &= \Pr(a|\lambda^*, B, c). \end{aligned} \tag{2}$$

has moderate solutions. Despite that each equation in the system robustly has moderate solutions, we prove that the two equations almost never share a moderate solution. In other words, confounded learning almost never arise in models with payoff shock.

The existence of confounded learning undermines long run efficiency. In fact, in a model without payoff shock, its confounded learning beliefs may be uncountable and even homeomorphic to a cantor set. (See Wang (2021)). However, no such belief could stay stationary as long as an arbitrary small shock  $c$  arrives regularly. In this sense, we say confounded learning is fragile under payoff shock.

This article primarily contributes to the observational learning literature by proving that confounded learning is fragile to almost all small payoff shock in a widely considered complete information, divergent preferences setting. In reaching that conclusion, we develop a math framework which may be useful in proving other generic results in observational learning models. Some of our math results, like private signals with real-analytic density functions are dense in certain sense, may have independent usage for other researchers.

Our work is sharply different from a recently a growing literature on observational learning with payoff externalities, which focus on long run efficiency in an environment that players' payoffs depend on not only the underlying state but also actions of other players. See Ali and Kartik (2012), Eyster et al. (2014), Arieli (2017), Monzón (2019) and Song and Zhang (2020). Our work is different both from the set up perspective and the result perspective. The payoff changes in this literature is endogenous. In contrast, the payoff shock in our model is exogenous. What is more important is that this literature does not discuss confounded learning. They aims at study whether long run correct learning arises under different conditions. In contrast, our result is a generic result that confounded learning is fragile under exogenous public payoff shock.

This article is organized as following: in section 2 we set up the model; in section 3 we solve it; in section 4 we describe different types of long run learning in our model; in section 5 we rigorously state the results; in section 6 we prove.

## 2 Model with Public Payoff Shock

Our model is based on the classic model in Smith and Sørensen (2000) where confounded learning arises due to players having divergent preferences. We twist the model by assuming that in each period a payoff shock of value  $c$  arrives with a commonly known positive probability. The shock changes the payoff of action  $b$  by value  $c$ . Below is a detailed description.

M	$a$	$b$
$A$	$u$	$0$
$B$	$0$	$1$

MM	$a$	$b$
$A$	$0$	$v$
$B$	$1$	$0$

Table 1: Payoff without payoff shock

M	$a$	$b$
$A$	$u$	$c$
$B$	$0$	$1 + c$

MM	$a$	$b$
$A$	$0$	$v + c$
$B$	$1$	$c$

Table 2: Payoff without payoff shock

The model is of discrete-time. In period 0, nature chooses one realization out of two potential states  $A$  and  $B$  according to a flat prior. In each period  $t \geq 1$ , one player arrives and needs to make a choice between actions  $a$  and  $b$ . There are two types of players-“match” type and “mismatch” type. The match type receives positive payoff iff his/her actions matches the realized underlying state. The mismatch type receives positive payoff if the chosen action is different from nature’s chosen state. In each period, a payoff shock arrives with a commonly known probability. The arrival of payoff shock is independent from other primitives of the model, and the specific arrival probability doesn’t matter for our results. Table 1 gives the payoff matrices of match type (M) and mismatch type (MM) without payoff shock. Table 2 gives the payoff matrices with shock. Here  $u, v > 0$  and  $u \neq v$ . The payoff tables are commonly known by all players.

Before deciding, player at period  $t$  observes the ordered choices made by his predecessors and whether each predecessor meets a payoff shock. He/she doesn’t observe the type of each predecessor. But players commonly know that each player is of match type with probability  $p$ . Player  $t$  observes his/her own type and payoff table before deciding.

Besides, player  $t$  also observes a noisy private signal about nature’s choice. This noisy signal is a state-contingent random variable  $\mathcal{S}$ . Its distribution is  $F^A(s)$  if nature chooses state  $A$ , and is  $F^B(s)$  if nature chooses state  $B$ . Following the literature, private signal is modeled as a direct signal, that is,  $\mathcal{S} = s$  means that “the probability that nature chooses state  $A$  is  $s$ ”. We also assume that there is no fully revealing signal. For technical reason, we make the following assumption

**Assumption 1 (Continuous and Uniformly Bounded Assumption)** *We assume the distributions of private signal  $\mathcal{S}$  under both states ( $F^A(s), F^B(s)$ ) have continuous and uniformly bounded density functions ( $f^A(s), f^B(s)$ ).*

Besides, all players commonly know the state-contingent distributions of signal  $S$  and that  $\{S_t\}$  is i.i.d. across periods.

To summarize, the information set of player  $t$  before deciding is

$$\{S_t, T_t, h_t\}. \quad (3)$$

Here  $S_t$  is his/her private signal's realization,  $T_t$  is his/her type realization. These are private information known only to player  $t$ . That  $h_t$  is the public history up to period  $t$ . It is an ordered sequence

$$h_t = \{(\alpha_1, \chi_1), \dots, (\alpha_{t-1}, \chi_{t-1}), \chi_t\}. \quad (4)$$

Here  $\alpha_k \in \{a, b\}$  ( $1 \leq k \leq t-1$ ) is the action chosen at period  $k$ ,  $\chi_k$  denotes whether a payoff shock  $c$  occurs at period  $k$ . Public history  $h_t$  is observed not only by player  $t$ , but also all players arrives later.

We shall refer the above model as a *shock model*. We often need to compare it with the corresponding *no shock model*. By *corresponding*, we mean the no shock model has the same primitives as the shock model except that payoff is described by table 1 in every period. Roughly speaking, our result says all the confounded learning beliefs of the no shock model turns non-stationary in the corresponding shock model for almost all small shocks.

### 3 Solution of the Shock Model

The shock model can be solved through a standard martingale argument. The main conclusion is that long run learning still happens almost surely, as in any no shock model. Readers who are not familiar with the standard martingale argument in solving an observational learning model can read Smith and Sørensen (2000) and Wang (2019).

Without loss of generality, let us assume that nature chooses state  $A$  in period 0. The public belief  $\lambda_t$  at period  $t$  is the posterior likelihood ratio of state being  $B$  over being  $A$  conditional on public history  $h_t$ :

$$\lambda_t = \frac{\Pr(B|h_t)}{\Pr(A|h_t)}. \quad (5)$$

Player  $t$ 's posterior belief of state being  $B$  over being  $A$  is

$$\frac{\Pr(B|h_t, S_t)}{\Pr(A|h_t, S_t)} = \lambda_t \frac{1 - S_t}{S_t}. \quad (6)$$

So we can compute the decision rule of player  $t$ . Player  $t$  follows a cutoff strategy with the cutoff described as following:

- If shock arrives at period  $t$ , a match type chooses action  $b$  iff

$$S_t < \frac{\lambda_t(1 + c)}{\lambda_t(1 + c) + (u - c)} \equiv m(\lambda_t, c). \quad (7)$$

If no shock arrives at period  $t$ , a match type chooses action  $b$  iff

$$S_t < \frac{\lambda_t}{\lambda_t + u} = m(\lambda_t, 0). \quad (8)$$

- If shock arrives at period  $t$ , a mismatch type chooses action  $b$  iff

$$S_t > \frac{\lambda_t(1 - c)}{\lambda_t(1 - c) + (v + c)} \equiv mm(\lambda_t, c). \quad (9)$$

If no shock arrives at period  $t$ , a mismatch type chooses action  $b$  iff

$$S_t > \frac{\lambda_t}{\lambda_t + v} \equiv mm(\lambda_t, 0). \quad (10)$$

After player  $t$  makes his/her decision, the public belief evolves as following

$$\lambda_{t+1} = \frac{\Pr(B|h_{t+1})}{\Pr(A|h_{t+1})} = \frac{\Pr(B|h_t) \Pr(\alpha_t|B, h_t)}{\Pr(A|h_t) \Pr(\alpha_t|A, h_t)} = \lambda_t \frac{\Pr(\alpha_t|B, h_t)}{\Pr(\alpha_t|A, h_t)}. \quad (11)$$

Then we could verify that  $\{\lambda_t\}$  forms a martingale:

$$E[\lambda_{t+1}|A, h_t] = \sum_{\alpha_t \in \{a, b\}} \left( \lambda_t \frac{\Pr(\alpha_t|B, h_t)}{\Pr(\alpha_t|A, h_t)} \right) \Pr(\alpha_t|A, h_t) = \lambda_t. \quad (12)$$

Therefore, martingale  $\{\lambda_t\}$  converges almost surely to a limit random variable  $\lambda_\infty$ . This means that if we randomly draw an infinite history  $h$  out, then

$$\lim_{t \rightarrow +\infty} \lambda_t(h) = \lambda_\infty(h) \quad (13)$$

for almost all  $h$ . This means that the public belief updated along history  $h$  eventually settle down to the value  $\lambda_\infty(h)$  for almost all  $h$ . Any value that r.v.  $\lambda_\infty$  takes with positive probability is referred as a long run learning belief.

Therefore, after introducing a public observable payoff shock, the classic result of observational learning still holds: the public belief eventually settles down almost surely.

## 4 Different Types of Long Run Learning

In this section we classify all the long run learning beliefs into three classes: (1) confounded learning; (2) herding learning; and (3) mixed learning.

We start with confounded learning. We recall its key property is that both types still use private information to decide at confounded learning beliefs(active). Therefore, at a confounded learning belief  $\lambda^*$ , the cutoffs of both types' strategy must be within the domain of private signals. That is,

$$m(\lambda^*, 0), m(\lambda^*, c), mm(\lambda^*, 0), mm(\lambda^*, c) \in (\underline{s}, \bar{s}). \quad (14)$$

Condition 14 can be rewritten as following. Let

$$\mathcal{E}_0 = \{\lambda | m(\lambda, 0), m(\lambda, 0) \in (\underline{s}, \bar{s})\}; \mathcal{E}_c = \{\lambda | m(\lambda, c), m(\lambda, c) \in (\underline{s}, \bar{s})\}. \quad (15)$$

Here  $\mathcal{E}_0$  is the set of all the public beliefs such that both types are active if no shock arrives. We call  $\mathcal{E}_0$  to be the confounding region without shock, and similarly call  $\mathcal{E}_c$  to be the confounding region with shock  $c$ . Then

**Definition 2** *A long run learning belief  $\lambda^*$  is confounded iff  $\lambda^* \in \mathcal{E}_0 \cap \mathcal{E}_c$ .*

The necessary condition for a belief  $\lambda^*$  to be a long run learning belief is that  $\lambda^*$  must be stationary with and without shock. Thus we could obtain the following necessary condition for a belief  $\lambda^*$  to be confounded:

**Condition 3 (Necessary Condition for Confounded Learning Belief)** *That  $\lambda^*$  is confounded learning belief in a model with shock  $c$  only if  $\lambda^* \in \mathcal{E}_0 \cap \mathcal{E}_c$  and solves*

$$\Pr(b|B, \lambda^*, 0) = \Pr(b|A, \lambda^*, 0); \Pr(b|B, \lambda^*, c) = \Pr(b|A, \lambda^*, c). \quad (16)$$

Now we turn to describe herding beliefs. Such a belief must make both types of players ignore their private information and act according to the public belief (inactive) in all periods. Therefore

**Definition 4** *A long run learning belief  $\lambda^*$  is herding iff  $\lambda^* \in H_0 \cap H_c$ .*

Here

$$\begin{aligned} H_0 &= \{\lambda \in (0, +\infty) | m(\lambda, 0), mm(\lambda, 0) \notin (\underline{s}, \bar{s})\}; \\ H_c &= \{\lambda \in (0, +\infty) | m(\lambda, c), mm(\lambda, c) \notin (\underline{s}, \bar{s})\} \end{aligned} \quad (17)$$

are the herding region without (with shock  $c$ ) respectively.

A sharp reader may have guessed the meaning of a mixed learning.

**Definition 5** *A long run learning belief  $\lambda^*$  is mixed iff  $\lambda^* \in \mathcal{E}_0 \cap H_c$  (first type) or  $\lambda^* \in \mathcal{E}_c \cap H_0$  (second type).*

These are the beliefs that are confounded without shock and are herding under shock, or vice versa. We pay special attention to the first type mixed learning. In fact, if a confounded learning belief of a no shock model is stationary in the shock model, it must either be confounded or be first type mixed. To see this, we classify all the interior public beliefs under shock  $c$  as the herding region and the non-herding region  $NH_c = (0, +\infty) - H_c$ . For any belief in  $NH_c - \mathcal{E}_c$ , exactly one type of player is active. Since the actions of this type are not confounded by the other type, they are informative. Hence any belief in  $NH_c - \mathcal{E}_c$  cannot be stationary.

## 5 Fragility of Confounded Learning

In this section we state our main results.

**Theorem 6** *For almost all private signals and almost all payoff shocks, the shock model admits no confounded learning, despite players have sufficiently divergent preferences.*

*Furthermore, the shock model admits no first type mixed learning, for all private signals and all small payoff shocks.*

*To summarize, in almost all cases, none of the confounded learning beliefs in a no shock model stays stationary in the corresponding shock model.*

We spend the rest of this section to set-up the necessary math so that we can rigorously translate the above theorem. In particular, we want to explain the meaning of *almost all private signals* and *almost all payoff shocks*.

Let us recall that an observational learning model with payoff shock is described by the following parameters: (1)  $p$ -the proportion of match type players; (2)  $u, v$ -the payoff parameters of both types; (3)  $(\underline{s}, \bar{s})$ -the strength of private signal; (4)  $(f^A(s), f^B(s))$ -the private signal. We first nail down those parameters that makes our main theorem meaningful. Obviously, if the confounding region without shock  $\mathcal{E}_0$  is empty, then naturally there is no confounded learning or mixed learning of first type. So we only consider parameter tuples  $(p, u, v, \underline{s}, \bar{s})$  with  $\mathcal{E}_0 \neq \emptyset$ .

After arbitrarily pick a meaningful parameter tuple, we consider all private signals with support  $(\underline{s}, \bar{s})$  and satisfying assumption 1. We claim that this set of private signals can be identified with the following subset of  $L_\infty(\underline{s}, \bar{s}) \cap C^0(\underline{s}, \bar{s})$

$$F_{(\underline{s}, \bar{s})}^\infty = \left\{ f(s) \left| \int_{\underline{s}}^{\bar{s}} f(s) \frac{1-2s}{s} ds = 0; \int_{\underline{s}}^{\bar{s}} f(s) ds = 1; f(s) \geq 0; f(s) \frac{1-s}{s} \in L_\infty(\underline{s}, \bar{s}) \right. \right\}. \quad (18)$$

To see this, we recall that private signal is modeled as direct signal-a realization of value  $s$  means that “probability of state being  $A$  is  $s$ ”. Then we must have

$$\frac{f^B(s)}{f^A(s)} = \frac{1-s}{s} \quad (19)$$

on the common support of  $(f^A(s), f^B(s))$ . Equation 19 is necessary because it impose the requirement that the likelihood of direct signals  $\frac{1-s}{s}$  must agree with the likelihood of density functions  $\frac{f^B(s)}{f^A(s)}$ . Obviously, a pair of density functions  $(f^A(s), f^B(s))$  satisfying equation 19 gives out a description of direct signals. Using this relation we could identify a pair of density functions  $(f^A(s), f^B(s))$  with  $f^A(s)$ . The  $F_{(\underline{s}, \bar{s})}^\infty$  is the set of density functions  $f(s)$  that could serve as  $f^A(s)$  in a pair. In other words, any  $f(s) \in F_{(\underline{s}, \bar{s})}^\infty$  is a density function satisfying that  $f(s) \frac{1-s}{s}$  is also a continuous uniformly bounded density function. This signal set  $F_{(\underline{s}, \bar{s})}^\infty$  is embedded into  $L_\infty(\underline{s}, \bar{s})$  and is endowed with the subspace topology induced by  $L_\infty$ -norm. Now we can nail down the meaning of *almost all signals*:

**Definition 7**  *Holding a meaning parameter tuple  $(p, u, v, \underline{s}, \bar{s})$ , we say a statement  $P$  holds for almost all signals if it holds for all the signals  $f(s) \in F_{(\underline{s}, \bar{s})}^\infty$  except for a first category set. Here the first category is defined with respect to the topology induced by  $L_\infty$ -norm.*

Now we turn to nail down the meaning of “almost all shocks”. Given a parameter tuple  $(p, u, v, \underline{s}, \bar{s})$  with non-empty  $\mathcal{E}_0$ , it is trivial to say the model with shock  $c$  admits no confounded learning if the confounding region under shock  $c$  becomes empty. Therefore, we only consider shock in the set

$$C = \{c \in (-\infty, +\infty) | \mathcal{E}_c \neq \emptyset\} - \{0\}. \quad (20)$$

We observe that this shock set  $C$  must be bounded. Actually,  $C$  must be a subset of  $(-1, u) \cap (-v, 1)$ , which is the set of all shocks that guarantees both types are active. On the other hand, we also observe that  $C$  must be non-empty. This is because we have guaranteed that  $\mathcal{E}_0$  is non-empty by choosing the meaningful parameter tuple  $(p, u, v, \underline{s}, \bar{s})$ . A small enough shock just slightly shift the confounding region and cannot eliminate it. So  $C$  always contain an interval around 0. We also need a technical assumption:

**Assumption 8** *We arbitrarily pick a small enough  $\varepsilon_0 > 0$  and consider*

$$C_{\varepsilon_0} = C \cap (-1 + \varepsilon_0, u - \varepsilon_0) \cap (-v + \varepsilon_0, 1 - \varepsilon_0) \quad (21)$$

*to be the set of meaningful shocks.*

The purpose of this assumption is to keep cutoffs  $m(\lambda, c), mm(\lambda, c)$  bounded away from 0 and  $+\infty$  for all  $c \in C_{\varepsilon_0}$ . Given that we can take  $\varepsilon_0$  arbitrarily small, we consider this as a mild assumption.

**Definition 9** *We say a statement holds for almost all shocks if it holds a.e. on  $C_{\varepsilon_0}$ .*

Now we can translate the first part of our main theorem into rigorous mathematical language:

**Theorem 10** *Given any parameter tuple  $(p, u, v, \underline{s}, \bar{s})$  with non-empty  $\mathcal{E}_0$ , for almost all private signals, the following equation system*

$$\Pr(b|B, \lambda^*, 0) = \Pr(b|B, \lambda^*, 0), \Pr(b|B, \lambda^*, c) = \Pr(b|B, \lambda^*, c); \quad (22)$$

*has no solution on  $\mathcal{E}_0 \cap \mathcal{E}_c$  for almost all  $c$ .*

The proof of this theorem is complicated and is delayed to the next section.

The second part of theorem 6 can be easily proved. If private signal is unbounded, then the herding region under shock  $c$  is empty and the result is immediate. Otherwise, a small

shock  $c$  just slightly shift the non-herding region from  $NH_0$  to  $NH_c$ . Because  $NH_0$  strictly covers  $\mathcal{E}_0$ <sup>1</sup>, a slight shift doesn't change it. So  $\mathcal{E}_0 \cap H_c$  must be empty for small  $c$ . In fact, given all other primitives, we call a shock  $c$  *small* iff it satisfies that  $\mathcal{E}_0 \cap H_c = \emptyset$ .

We close this section with the following corollary:

**Corollary 11** *If private signal is of unbounded strength, in the long run all the weight is assigned to the true state for almost all signals and almost all shocks.*

To see it, we observe that  $\mathcal{E}_c$  is actually  $(0, +\infty)$  for all  $c$  in  $C_{\varepsilon_0} \cup \{0\}$ . Then theorem 10 implies that the only stationary beliefs are  $\lambda^* = 0$  and  $\lambda^* = +\infty$ . Because the standard martingale argument rules out the possibility that all the weight is assigned to the incorrect state<sup>2</sup>, the result follows.

## 6 Proof of Main Result

In this section we prove our theorem 10. Before proof, we first rewrite the equation system 22 so that the notation reflect the underlying private signal  $f(s)$ : first, we have

$$\begin{aligned} & \Pr(b|B, \lambda^*, c) - \Pr(b|A, \lambda^*, c) \\ &= p[F^B(m(\lambda^*, c)) - F^A(m(\lambda^*, c))] + (1-p)[F^B(mm(\lambda^*, c)) - F^A(mm(\lambda^*, c))]. \end{aligned}$$

Then use that private signal pair  $(f^A(s), f^B(s))$  is given by  $(f(s), f(s)\frac{1-s}{s})$  and rewrite probability as integral of density function, we have

$$\begin{aligned} & p[F^B(m(\lambda^*, c)) - F^A(m(\lambda^*, c))] + (1-p)[F^B(mm(\lambda^*, c)) - F^A(mm(\lambda^*, c))] \\ &= p \int_{\underline{s}}^{m(\lambda^*, c)} f(s) \frac{1-2s}{s} ds + (1-p) \int_{mm(\lambda^*, c)}^{\bar{s}} f(s) \frac{1-2s}{s} ds \end{aligned} \quad (23)$$

provided that  $m(\lambda^*, c), mm(\lambda^*, c) \in (\underline{s}, \bar{s})$ . We shall denote the last line of equation 23 as  $G_f(\lambda^*, c)$ . Then the equations system 22 can be rewritten as

$$G_f(\lambda^*, 0) = G_f(\lambda^*, c) = 0. \quad (24)$$

<sup>1</sup>The non-herding region is the union of active regions of both type while the confounding region is the intersection of these active regions.

<sup>2</sup>Following Fatou's lemma,  $\lim_{t \rightarrow +\infty} E[\lambda_t] \geq E[\lim_{t \rightarrow \infty} \lambda_t]$ . If  $\lambda^*$  could be  $+\infty$  with positive probability, then the right-hand side is  $+\infty$ . This is impossible since the left-hand side is  $\lambda_0$ , which is finite.

## 6.1 Strongly Bounded Signals

We start with the case where  $0 < \underline{s} < \bar{s} < 1$ . We name all these signals as “strongly bounded”. For each private signal  $f(s)$ , let

$$S_f = \{c \in C_{\varepsilon_0} | G_f(\lambda^*, 0) = G_f(\lambda^*, c) = 0 \text{ for some } \lambda^* \in \mathcal{E}_0 \cap \mathcal{E}_c\}. \quad (25)$$

In other words,  $S_f$  is the set of shocks such that confounded learning belief  $\lambda^*$  arises in a shock model. We call  $S_f$  as the stationary set for signal  $f(s)$ .

We make the following definition

**Definition 12**  $F_n \subset F_{(\underline{s}, \bar{s})}^\infty$  is the set of private signals whose stationary set  $S_f$  has measure no less than  $\frac{1}{n}$ .

Our purpose is to prove that  $F_n$  is nowhere dense in  $F_{(\underline{s}, \bar{s})}^\infty$ . Then, except a first category set  $\bigcup_n F_n$ , the stationary set of each private signal  $f(s)$  must have measure zero. This completes the proof of theorem 10 for strongly bounded signals.

To prove the nowhere denseness of  $F_n$ , we turn to show that  $F_n$ 's closure in  $F_{(\underline{s}, \bar{s})}^\infty$ , which we denoted as  $\bar{F}_n$  has no interior point. To do so, we must first characterize the closure  $\bar{F}_n$ . The following proposition says that  $\bar{F}_n$  is almost  $F_n$ .

**Proposition 13** If  $f(s) \in \bar{F}_n$ , then there exists  $S \subset C_{\varepsilon_0}$  with  $m(S) \geq \frac{1}{n}$  such that  $\forall c \in S$ , equation system

$$G_f(\lambda^*, 0) = G_f(\lambda^*, c) = 0 \quad (26)$$

has solutions on  $\bar{\mathcal{E}}_0 \cap \bar{\mathcal{E}}_c$ .

Comparing definition 12 and proposition 13, we see the only difference is that: if  $f(s) \in \bar{F}_n$ , then the equation system could have solution on the closure of  $\mathcal{E}_0 \cap \mathcal{E}_c$ .

**Proof.** For any  $f(s) \in \bar{F}_n$ , there exists  $\{f_n(s)\} \in F_n$  such that  $\|f_n - f\|_{L_\infty} \rightarrow 0$ . From definition 12, each  $f_n(s)$  has an associated stationary set  $S_n$  with  $m(S_n) \geq \frac{1}{n}$ . We let

$$S_{i.o.} = \bigcap_k \bigcup_{k \geq n} S_n \quad (27)$$

to be shocks that shows up infinite times in  $\{S_n\}$ . We could prove that  $m(S_{i.o.}) \geq \frac{1}{n}$ .

Arbitrarily choose a  $c_0 \in S_{i.o.}$ , there exists subsequence  $n_k$  such that  $c_0 \in S_{n_k}$ . Since  $c_0$  is a stationary shock for  $f_{n_k}(s)$ , for each  $n_k$ , there exists  $\lambda_{n_k}^{c_0}$  being confounded belief for  $f_{n_k}(s)$ .

That is,

$$G_{f_{n_k}}(\lambda_{n_k}^{c_0}, c_0) = G_{f_{n_k}}(\lambda_{n_k}^{c_0}, 0) = 0 \text{ and } \lambda_{n_k}^{c_0} \in \mathcal{E}_0 \cap \mathcal{E}_{c_0} \quad (28)$$

Then  $\{\lambda_{n_k}^{c_0}\}$  has at least one cluster point  $\Lambda_{c_0}$  on  $\bar{\mathcal{E}}_0 \cap \bar{\mathcal{E}}_{c_0}$ . Let  $n_{k_q}$  be a sub-subsequence of  $n_k$  such that  $\lambda_{n_{k_q}}^{c_0}$  converges to  $\Lambda_{c_0}$ . We claim that

$$\lim_{n_{k_q} \rightarrow +\infty} |G_{f_{n_{k_q}}}(\lambda_{n_{k_q}}^{c_0}, c_0) - G_f(\Lambda_{c_0}, c_0)| = \lim_{n_{k_q} \rightarrow +\infty} |G_{f_{n_{k_q}}}(\lambda_{n_{k_q}}^{c_0}, 0) - G_f(\Lambda_{c_0}, 0)| = 0. \quad (29)$$

This claim directly implies that

$$G_f(\Lambda_{c_0}, c_0) = G_f(\Lambda_{c_0}, 0) = 0 \text{ with } \Lambda_{c_0} \in \bar{\mathcal{E}}_0 \cap \bar{\mathcal{E}}_{c_0}, \quad (30)$$

which completes the proof of proposition 13.

To prove claim 29, we do the following transformations. The first transformation is

$$\begin{aligned} & |G_{f_{n_{k_q}}}(\lambda_{n_{k_q}}^{c_0}, \mathbf{c}) - G_f(\Lambda_{c_0}, \mathbf{c})| \\ = & |G_{f_{n_{k_q}}}(\lambda_{n_{k_q}}^{c_0}, \mathbf{c}) - G_f(\lambda_{n_{k_q}}^{c_0}, \mathbf{c}) + G_f(\lambda_{n_{k_q}}^{c_0}, \mathbf{c}) - G_f(\Lambda_{c_0}, \mathbf{c})| \\ = & \left| -p \int_{m(\lambda_{n_{k_q}}^{c_0}, \mathbf{c})}^{\bar{s}} \left[ f_{n_{k_q}}(s) - f(s) \right] \frac{1-2s}{s} ds + (1-p) \int_{mm(\lambda_{n_{k_q}}^{c_0}, \mathbf{c})}^{\bar{s}} \left[ f_{n_{k_q}}(s) - f(s) \right] \frac{1-2s}{s} ds \right. \\ & \left. + p \int_{m(\Lambda_{c_0}, \mathbf{c})}^{m(\lambda_{n_{k_q}}^{c_0}, \mathbf{c})} f(s) \frac{1-2s}{s} ds + (1-p) \int_{mm(\lambda_{n_{k_q}}^{c_0}, \mathbf{c})}^{mm(\Lambda_{c_0}, \mathbf{c})} f(s) \frac{1-2s}{s} ds \right|. \end{aligned} \quad (31)$$

Here we use symbol  $\mathbf{c}$  to represent either 0 or  $c_0$ . Besides, to obtain the first term in the last line of equation 31, we use the fact that  $p \int_{m(\lambda_{n_{k_q}}^{c_0}, \mathbf{c})}^{\bar{s}} [f(s) - f_{n_{k_q}}(s)] \frac{1-2s}{s} ds = -p \int_{\underline{s}}^{m(\lambda_{n_{k_q}}^{c_0}, \mathbf{c})} [f(s) - f_{n_{k_q}}(s)] \frac{1-2s}{s} ds$ , which follows from  $\int_{\underline{s}}^{\bar{s}} f(s) \frac{1-2s}{s} ds = \int_{\underline{s}}^{\bar{s}} f_n(s) \frac{1-2s}{s} ds = 0$ .

Then we make the following observation:

$$m(\lambda_{n_{k_q}}^{c_0}, \mathbf{c}), mm(\lambda_{n_{k_q}}^{c_0}, \mathbf{c}), m(\Lambda_{c_0}, \mathbf{c}), mm(\Lambda_{c_0}, \mathbf{c}) \in [b, \bar{b}] \subset (0, 1) \quad (32)$$

Actually, the extreme values of  $m(\lambda, c)$  and  $mm(\lambda, c)$  on compact region  $\bar{\mathcal{E}}_0 \times \bar{C}_{\varepsilon_0}$  can be computed through the partial derivatives of  $m(\lambda, c)$  and  $mm(\lambda, c)$ . We remark observation 32 relies on two things: (1)  $\lambda_{n_{k_q}}^{c_0}, \Lambda_{c_0} \in \bar{\mathcal{E}}_0$  and  $0 < \inf \mathcal{E}_0 < \sup \mathcal{E}_0 < +\infty$ . Here we use assumption that  $0 < \underline{s} < \bar{s} < 1$  to obtain that  $\mathcal{E}_0$  is bounded away from 0,  $+\infty$ ; (2)  $C_{\varepsilon_0}$  is bounded away from  $-1, 1, -v, u$  following assumption 8.

Following this observation, we can apply triangle inequality to equation 31 and obtain that

$$\begin{aligned}
& |G_{f_{n_{k_q}}}(\lambda_{n_{k_q}}^{c_0}, \mathbf{c}) - G_f(\Lambda_{c_0}, \mathbf{c})| \\
\leq & \left[ \max_{s \in [b, 1]} \left| \frac{1 - 2s}{s} \right| \right] \left[ \|f_{n_{k_q}}(s) - f(s)\|_{L_\infty} + p \|f(s)\| |m(\Lambda_{c_0}, \mathbf{c}) - m(\lambda_{n_{k_q}}^{c_0}, \mathbf{c})| \right. \\
& \left. + (1 - p) \|f(s)\| |mm(\Lambda_{c_0}, \mathbf{c}) - mm(\lambda_{n_{k_q}}^{c_0}, \mathbf{c})| \right]. \tag{33}
\end{aligned}$$

Since  $\lambda_{n_{k_q}}^{c_0} \rightarrow \Lambda_{c_0}$ , by the continuity of  $m(\lambda, c)$ ,  $mm(\lambda, c)$ , the last two terms in equation 33 vanishes. Recall that  $\|f_{n_{k_q}}(s) - f(s)\|_{L_\infty} \rightarrow 0$ , the first term vanishes as well. So we prove claim 29. ■

Now, to prove that  $\overline{F}_n$  has no interior point, we turn to prove the following two propositions.

**Proposition 14** *For any  $f(s) \in F_{(\underline{s}, \overline{s})}^\infty$ , there exists a sequence of real-analytic private signals  $\{f_n^\omega(s)\} \subset F_{(\underline{s}, \overline{s})}^\infty \cap C^\omega(\underline{s}, \overline{s})$  converging to  $f(s)$  in  $L_\infty$ -norm. In other words, the signals with real-analytic density functions are dense in  $F_{(\underline{s}, \overline{s})}^\infty$ .*

**Proposition 15** *Any real-analytic signal  $f^\omega(s)$  is not in  $\overline{F}_n$ .*

Following the above two propositions, any  $f(s) \in \overline{F}_n$  can be approximated by a sequence of real-analytic signals which are not in  $\overline{F}_n$ . Therefore, no point in  $\overline{F}_n$  could be interior. Now we prove the above two propositions:

**Proof of Proposition 14.** Following the Whitney smooth approximation theorem (theorem 6.1.5 in Lee (2000)) and the Whitney analytic approximation theorem (theorem 1.6.5 in Narasimhan (1968)), for any  $f(s) \in F_{(\underline{s}, \overline{s})}^\infty$ , there exists a sequence of  $\overline{g}_n(s) \in C^\omega(\underline{s}, \overline{s})$  such that

$$\left\| \overline{g}_n(s) - \frac{f(s)}{s} \right\|_{L_\infty} < \frac{1}{n}. \tag{34}$$

Here  $\frac{f(s)}{s} \in L_\infty(\underline{s}, \overline{s})$  since we assume that  $f(s) \frac{1-s}{s}, f(s) \in L_\infty(\underline{s}, \overline{s})$  for any  $f(s) \in F_{(\underline{s}, \overline{s})}^\infty$ . The problem is that  $\overline{g}_n(s)$  are not necessarily private signals. They may fail the four requirements listed in the description of  $F_{(\underline{s}, \overline{s})}^\infty$  as in 18. We shall construct a sequence of analytic signals that uniformly converge to  $f(s)$  based on the existence of  $\{\overline{g}_n(s)\}_{n \in \mathbb{N}}$ .

First of all, we verify that

$$\epsilon_n \equiv \int_{\underline{s}}^{\bar{s}} \bar{g}_n(s)(1-2s)ds \rightarrow \int_{\underline{s}}^{\bar{s}} \frac{f(s)}{s}(1-2s)ds = 0. \quad (35)$$

Let  $r(s)$  be a polynomial that satisfies  $\int_{\underline{s}}^{\bar{s}} r(s)(1-2s)ds = 1$ . Let  $\hat{g}_n(s) = \bar{g}_n(s) - \epsilon_n r(s)$ , then  $\int_{\underline{s}}^{\bar{s}} \hat{g}_n(s)(1-2s)ds = 0$ . Let

$$\delta_n = \min\left\{\min_{s \in [\underline{s}, \bar{s}]} \hat{g}_n(s), 0\right\}. \quad (36)$$

We can verify that  $-\frac{1}{n} - |\epsilon_n| \|r(s)\|_{L_\infty[\underline{s}, \bar{s}]} \leq \delta_n \leq 0$  and hence  $\delta_n \rightarrow 0$ . Now we lift  $\hat{g}_n(s)$  up to eliminate possible negative values. We construct a lifting function  $h(s)$  as following: given  $(\underline{s}, \bar{s}) \subset (0, 1)$ , for any  $a < 0, b > 1$ , there exists an  $\alpha \in \mathbb{R}$  such that

$$\int_{\underline{s}}^{\bar{s}} \frac{1}{2}(x-a)(b-x)e^{\alpha(x-\frac{1}{2})}(1-2x)dx = 0. \quad (37)$$

We let  $h(x) = \frac{1}{2}(x-a)(b-x)e^{\alpha(x-\frac{1}{2})}$ . Intuitively, a large positive  $\alpha$  makes  $h(s)(1-2s)$  large on  $(\frac{1}{2}, \bar{s})$ , where  $h(s)(1-2s)$  is negative and small on  $(\underline{s}, \frac{1}{2})$ , where  $h(s)(1-2s)$  is positive; and a large negative  $\alpha$  has the opposite effect on  $h(s)(1-2s)$ . By continuity of  $\alpha$ , some intermediate value makes equation 37 holds.

Let  $\delta = \min_{s \in [0, 1]} h(s)$ . We could verify that  $\delta > 0$ . We let

$$\tilde{g}_n(s) = \hat{g}_n(s) - \frac{\delta_n}{\delta} h(s). \quad (38)$$

We could verify that  $\tilde{g}_n(s) \geq 0$ . Furthermore, since equation 37 holds, we have  $\int_{\underline{s}}^{\bar{s}} \hat{g}_n(s)(1-2s)ds = \int_{\underline{s}}^{\bar{s}} \tilde{g}_n(s)(1-2s)ds = 0$ .

Finally, we let

$$g_n(s) = \frac{\tilde{g}_n(s)s}{\int_{\underline{s}}^{\bar{s}} \tilde{g}_n(s)s ds} = \frac{[\bar{g}_n(s) - \epsilon_n r(s) - \frac{\delta_n}{\delta} h(s)]s}{\int_{\underline{s}}^{\bar{s}} [\bar{g}_n(s) - \epsilon_n r(s) - \frac{\delta_n}{\delta} h(s)]s ds}. \quad (39)$$

We can verify that  $\int_{\underline{s}}^{\bar{s}} g_n(s)ds = 1, \int_{\underline{s}}^{\bar{s}} g_n(s)\frac{1-2s}{s}ds = 0, g_n(s) \geq 0, g_n(s) \in C^\omega(\underline{s}, \bar{s})$  and  $g_n(s)\frac{1-s}{s} \in L_\infty(\underline{s}, \bar{s})$ . Therefore,  $\{g_n(s)\}$  is a sequence of analytic signals. It is direct to verify that  $\|g_n(s) - f(s)\|_{L_\infty} \rightarrow 0$ . ■

**Proof of Proposition 15.** For any real analytic signal  $f(s)$ , we could verify that

$$G_1(x) = p \int_{\underline{s}}^x f(s) \frac{1-2s}{s} ds \quad ; \quad G_2(x) = (1-p) \int_x^{\bar{s}} f(s) \frac{1-2s}{s} ds \quad (40)$$

are analytic on  $(\underline{s}, \bar{s})$ . For any  $\lambda \in (0, +\infty)$ , let

$$M_\lambda = \{c \in C_{\varepsilon_0} | m(\lambda, c), mm(\lambda, c) \in [\underline{s}, \bar{s}]\}. \quad (41)$$

we could verify that  $m(\lambda, c), mm(\lambda, c) \in (\underline{s}, \bar{s})$  for all  $c \in \overset{\circ}{M}_\lambda$ . Since composition of analytic functions is analytic, for any  $\lambda \in (0, +\infty)$ , we have

$$G_f(\lambda, c) = G_1(m(\lambda, c)) + G_2(mm(\lambda, c)) \quad (42)$$

is analytic in  $c$  on  $\overset{\circ}{M}_\lambda$ , provided that  $\overset{\circ}{M}_\lambda \neq \emptyset$ . Similarly, for any  $c \in C_{\varepsilon_0}$ ,  $G_f(\lambda, c)$  is analytic in  $\lambda$  on  $\mathcal{E}_c$ .

It is well-known that a real-analytic function could have at most finitely many zeros on a compact interval. So  $G_f(\lambda, 0) = 0$  could have at most finitely solutions on  $\mathcal{E}_0$ . Given that  $\bar{\mathcal{E}}_0 - \mathcal{E}_0$  contains two points, we know

$$\Lambda_f = \{\lambda \in \bar{\mathcal{E}}_0 | G_f(\lambda, 0) = 0\} \quad (43)$$

is a finite set.

Recall that  $f(s) \in \bar{F}_n$  implies that the set

$$S = \{c \in C_{\varepsilon_0} | G_f(\lambda, 0) = G_f(\lambda, c) = 0, \text{ for some } \lambda \in \bar{\mathcal{E}}_0 \cap \bar{\mathcal{E}}_c\} \quad (44)$$

must satisfy  $m(s) \geq \frac{1}{n}$ . Obviously, if  $\lambda$  is a confounded belief in a shock model with shock  $c \in S$ , then  $\lambda \in \Lambda_f$ . Then we could divide  $S$  into finitely many subsets as following: let

$$S_\lambda = \{c \in S | G_f(\lambda, 0) = G_f(\lambda, c) = 0 \text{ for one } \lambda \in \Lambda_f\}. \quad (45)$$

Then we could verify that  $S = \bigcup_{\lambda \in \Lambda_f} S_\lambda$ . As  $m(S) \geq \frac{1}{n}$  and  $\Lambda_f$  is finite, there exists at least one  $\lambda_0 \in \Lambda_f$  such that  $m(S_{\lambda_0}) > 0$ .

Therefore, if an real-analytic signal  $f(s) \in \bar{F}_n$ , then there exists a  $\lambda_0 \in \bar{\mathcal{E}}_0$  and a  $S_{\lambda_0} \subset C_{\varepsilon_0}$

such that

$$G_f(\lambda_0, 0) = G_f(\lambda_0, c) = 0, \forall c \in S_{\lambda_0}; \text{ and } \lambda_0 \in \bigcap_{c \in S_{\lambda_0}} \bar{\mathcal{E}}_c, m(S_{\lambda_0}) > 0. \quad (46)$$

We are going to show this is impossible.

We observe that  $\lambda_0 \in \bar{\mathcal{E}}_c$  implies that  $c \in M_{\lambda_0}$ . Then  $\lambda_0 \in \bigcap_{c \in S_{\lambda_0}} \bar{\mathcal{E}}_c$  implies that  $S_{\lambda_0} \subset M_{\lambda_0}$ . We can compute  $M_\lambda$  and see that it is the intersection of a (possibly degenerate) closed interval with  $C_{\varepsilon_0}$ . That  $m(S_{\lambda_0}) > 0$  guarantees that  $M_{\lambda_0}$  must have a non-empty interior. Then equation 46 implies that

$$G_f(\lambda_0, c) = 0; \forall c \in S_{\lambda_0} \cap \overset{\circ}{M}_{\lambda_0} \text{ and } m(S_{\lambda_0} \cap \overset{\circ}{M}_{\lambda_0}) > 0. \quad (47)$$

This contradict that  $G_f(\lambda, c)$  is analytic in  $c$  on  $\overset{\circ}{M}_\lambda$ . ■

## 6.2 Partially Bounded and Unbounded Signals

In this subsection we allow  $\underline{s} = 0$  or (and)  $\bar{s} = 1$ . The case that  $0 = \underline{s} < \bar{s} = 1$  is known as private signal is of unbounded strength. If  $0 = \underline{s} < \bar{s} < 1$ , then players could obtain arbitrarily precise signal about state being  $A$  but could only obtain bounded signal about state being  $B$ . We name such signals as “partially bounded”. Similarly,  $0 < \underline{s} < \bar{s} = 1$  is partially bounded as well.

We would like to extend the previous method. One difficulty arise if we try to describe  $\bar{F}_n$  by running the proof of proposition 13. If we ask ourselves about the limit  $f(s)$  of convergent sequence  $\{f_n(s)\} \subset F_n$  with confounded belief  $\lambda_n$  converging to 0, we shall find proposition 13’s proof tells us that  $f(s)$  should satisfy

$$G_f(0, 0) = G_f(0, c) = 0, \forall c \in S \subset C_{\varepsilon_0} \text{ with } m(S) \geq \frac{1}{n}. \quad (48)$$

This condition is satisfied by all signals and has no bite.

To resolve this difficulty, we turn to prove the the signals with sufficiently small or sufficiently large confounded beliefs are small in the first category sense. In fact, we have the following two propositions.

**Proposition 16** *There exists  $\underline{\lambda} > 0$  and  $\bar{\lambda} < +\infty$  such that*

$$\begin{aligned} G_f(\lambda, 0) = G_f(\lambda, c) = 0 \text{ for some } c \in C_{\varepsilon_0} \text{ and some } \lambda \in \mathcal{E}_0 \cap \mathcal{E}_c \cap ((0, \underline{\lambda}) \cup (\bar{\lambda}, +\infty)) \\ \Rightarrow \exists [z_1, z_2] \subset (\underline{s}, \bar{s}) \text{ s.t. } f(s) = 0 \text{ on } [z_1, z_2]. \end{aligned} \quad (49)$$

In other words, if private signal  $f(s)$  admits small ( $< \underline{\lambda}$ ) or large ( $> \bar{\lambda}$ ) confounded belief, then it must be locally zero. The next proposition states that locally zero private signals are of first category.

**Proposition 17** *The set of locally zero signals*

$$F_{zero} = \{f(s) \in F_{(\underline{s}, \bar{s})}^\infty \mid \exists \underline{s} < z_1 < z_2 < \bar{s} \text{ s.t. } f(s) = 0 \text{ a.e. on } [z_1, z_2].\}$$

*is of first category in  $F_{(\underline{s}, \bar{s})}^\infty$ .*

Now we prove the above two propositions.

**Proof of Proposition 16.** For any  $\lambda \in \mathcal{E}_0$ , differentiate  $G_f(\lambda, c)$  w.r.t  $c$  on  $\mathring{M}_\lambda$ , we have

$$\frac{\partial}{\partial c} G_f(\lambda, c) = pf(m(\lambda, c)) \frac{1 - 2m(\lambda, c)}{m(\lambda, c)} \frac{\partial}{\partial c} m(\lambda, c) - (1 - p)f(mm(\lambda, c)) \frac{1 - 2mm(\lambda, c)}{mm(\lambda, c)} \frac{\partial}{\partial c} mm(\lambda, c).$$

It is easy to verify that on  $(\lambda, c) \in (0, +\infty) \times \{(-1, u) \cap (-v, 1)\}$

$$\frac{\partial}{\partial c} m(\lambda, c) = \frac{\lambda(u+1)}{[\lambda(1+c) + (u-c)]^2} > 0; \quad \frac{\partial}{\partial c} mm(\lambda, c) = \frac{-\lambda(v+1)}{[\lambda(1-c) + (v+c)]^2} < 0. \quad (50)$$

Therefore, we have

$$\frac{\partial}{\partial c} G_f(\lambda, c) \begin{cases} \geq 0 \text{ on } \mathring{M}_\lambda \Leftarrow m(\lambda, c), mm(\lambda, c) < \frac{1}{2} \text{ on } \mathring{M}_\lambda \\ \leq 0 \text{ on } \mathring{M}_\lambda \Leftarrow m(\lambda, c), mm(\lambda, c) > \frac{1}{2} \text{ on } \mathring{M}_\lambda \end{cases}$$

We could make use of partial derivatives in equation 50 to compute

$$m(\lambda, c), mm(\lambda, c) < \frac{1}{2} \Leftrightarrow \lambda < \underline{\lambda} \equiv \min \left\{ \frac{u - \sup C_{\varepsilon_0}}{1 + \sup C_{\varepsilon_0}}, \frac{v + \inf C_{\varepsilon_0}}{1 - \inf C_{\varepsilon_0}} \right\} \quad (51)$$

and

$$m(\lambda, c), mm(\lambda, c) > \frac{1}{2} \Leftrightarrow \lambda > \bar{\lambda} \equiv \max \left\{ \frac{u - \inf C_{\varepsilon_0}}{1 + \inf C_{\varepsilon_0}}, \frac{v + \sup C_{\varepsilon_0}}{1 - \sup C_{\varepsilon_0}} \right\}. \quad (52)$$

By assuming that  $C_{\varepsilon_0}$  is bounded away from  $\{-1, 1, -v, u\}$ , we have  $\underline{\lambda} > 0, \bar{\lambda} < +\infty$ . Combining all the computation above, we have

$$\begin{aligned} \lambda \in (0, \underline{\lambda}) &\Rightarrow \frac{\partial}{\partial c} G_f(\lambda, c) \geq 0 \text{ on } \mathring{M}_\lambda; \\ \lambda \in (\bar{\lambda}, +\infty) &\Rightarrow \frac{\partial}{\partial c} G_f(\lambda, c) \leq 0 \text{ on } \mathring{M}_\lambda. \end{aligned} \quad (53)$$

Assume  $f(s)$  has a small confounded belief, that is,

$$G_f(\lambda^*, 0) = G_f(\lambda^*, c_0) = 0 \text{ for some } c_0 \in C_{\varepsilon_0} \text{ and some } \lambda^* \in \mathcal{E}_0 \cap \mathcal{E}_{c_0} \cap (0, \underline{\lambda}). \quad (54)$$

We first observe that  $0, c_0 \in \mathring{M}_{\lambda^*}$  since  $\lambda^* \in \mathcal{E}_{c_0} \cap \mathcal{E}_0$ . Furthermore, because  $\mathring{M}_{\lambda^*}$  is an open interval, so  $[0, c_0] \subset \mathring{M}_{\lambda^*}$ .<sup>3</sup> Because  $\lambda^* < \underline{\lambda}$ ,  $\frac{\partial}{\partial c} G_f(\lambda^*, c) \geq 0$  on  $\mathring{M}_{\lambda^*}$ . So

$$0 = G_f(\lambda^*, c_0) - G_f(\lambda^*, 0) = \int_0^{c_0} \frac{\partial}{\partial c} G_f(\lambda^*, c) dc = 0 \quad (55)$$

implies that  $\frac{\partial}{\partial c} G_f(\lambda^*, c) = 0$  a.e. on  $[0, c_0]$ . Furthermore,

$$\frac{\partial}{\partial c} G_f(\lambda^*, c) = \overbrace{pf(m(\lambda^*, c)) \frac{1 - 2m(\lambda^*, c)}{m(\lambda^*, c)} \frac{\partial}{\partial c} m(\lambda^*, c)}^{\geq 0} - \overbrace{(1-p)f(mm(\lambda^*, c)) \frac{1 - 2mm(\lambda^*, c)}{mm(\lambda^*, c)} \frac{\partial}{\partial c} mm(\lambda^*, c)}^{\geq 0}.$$

implies that

$$f(m(\lambda^*, c)) = 0, f(mm(\lambda^*, c)) = 0, \forall c \in [0, c_0]. \quad (56)$$

It is direct to verify that  $Range(m(\lambda^*, c)), Range(mm(\lambda^*, c))$  where  $c \in [0, c_0]$  must be closed intervals within  $(\underline{s}, \bar{s})$ . So we have obtained that a small robust belief implies local zeroness.

The proof for a large robust belief is similar. ■

---

<sup>3</sup>Without loss of generality, we assume  $c_0 > 0$ .

**Proof of Proposition 17.** Let

$$F_{[z_1, z_2]} = \{f(s) \in F_{(\underline{s}, \bar{s})}^\infty \mid f(s) = 0 \text{ a.e. on } [z_1, z_2] \subset (\underline{s}, \bar{s})\}. \quad (57)$$

Then

$$F_{zero} = \bigcup_{\substack{\underline{s} < z_1 < z_2 < \bar{s} \\ z_1, z_2 \in \mathbb{Q}}} F_{[z_1, z_2]}. \quad (58)$$

So we turn to show that  $F_{[z_1, z_2]}$  is nowhere dense in  $F_{(\underline{s}, \bar{s})}^\infty$ .

To do so, we turn to show the following: for any  $f(s) \in F_{[z_1, z_2]}$  and any  $\varepsilon > 0$ , we can construct a  $g(s)$  satisfying that (1)  $\|g(s) - f(s)\|_{L_\infty} < \varepsilon$  and (2)  $\exists \delta > 0$  such that any  $h(s) \in B_\delta(g(s))$  is not an element in  $F_{[z_1, z_2]}$ .

Basically, we just need to continuously lift up  $f(s)$  on  $[z_1, z_2]$  by a number smaller than  $\varepsilon$  and simultaneously adjust  $f(s)$  somewhere else to keep it a private signal. Then it is obvious any  $h(s)$  that's close enough to  $g(s)$  under  $L_\infty$ -norm cannot be within  $F_{[z_1, z_2]}$ . ■

Knowing that signals admitting small or large confounded beliefs must be of first category, we just need to prove that signals with moderate confounded belief is of first category as well. This can be done using the same proof as proposition 13 with one slight modification. Instead of  $F_n$ , we consider  $F_n^{[\underline{\lambda}, \bar{\lambda}]} \subset F_{(\underline{s}, \bar{s})}^\infty$  to be the set of private signals such that

$$G_f(\lambda, 0) = G_f(\lambda, c) = 0 \text{ for some } \lambda \in \mathcal{E}_0 \cap \mathcal{E}_c \cap [\underline{\lambda}, \bar{\lambda}] \quad (59)$$

for each shock  $c$  in  $S \subset C_{\varepsilon_0}$  with measure  $m(S) \geq \frac{1}{n}$ . Then the following proposition describes its closure  $\bar{F}_n^{[\underline{\lambda}, \bar{\lambda}]}$ .

**Proposition 18** *If  $f(s) \in \bar{F}_n^{[\underline{\lambda}, \bar{\lambda}]}$ , then there exists  $S \subset C_{\varepsilon_0}$  with  $m(S) \geq \frac{1}{n}$  such that  $\forall c \in S$ , equation system*

$$G_f(\lambda, 0) = G_f(\lambda, c) = 0 \quad (60)$$

*has solutions on  $\bar{\mathcal{E}}_0 \cap \bar{\mathcal{E}}_c \cap [\underline{\lambda}, \bar{\lambda}]$ .*

The method in proposition 13 can be directly used to prove proposition 18. By restricting the solution to interval  $[\underline{\lambda}, \bar{\lambda}]$ , we are out of the problem described in the beginning of this subsection.

Furthermore, proposition 14 and 15 directly extend to partially bounded and unbounded

signals. Therefore, we have completed the proof of theorem 10 in these cases.

## 7 Conclusion

In this article we consider observational learning model with exogenous payoff shock. We find that even a very simple shock structure eliminates the phenomenon that long run learning could be confounded. The public belief must eventually aggregate overwhelming evidence against one person's private information.

We think observational learning with a rich payoff shock might worth more exploration. Theoretically, for example, we can ask whether information aggregation becomes faster under a random payoff shock with sufficiently rich structure. This may be helpful in resolving the slowness of sequential observational learning when the private signals have thin tails, as pointed out in Hann-Caruthers et al. (2018) and Rosenberg and Vieille (2019). In reality, it is often easier to alter players' payoffs rather than improve the precision of players private signals. For example, in an environment that the public mistakenly herd on effectiveness about vaccine, to launch a educational campaign might be less effective than simply subsidize the vaccine.

## References

- ALI, S. N. AND N. KARTIK (2012): "Herding with collective preferences," *Economic Theory*, 51, 601–626.
- ARIELI, I. (2017): "Payoff externalities and social learning," *Games and Economic Behavior*, 104, 392–410.
- BANERJEE, A. V. (1992): "A simple model of herd behavior," *The Quarterly Journal of Economics*, 107, 797.
- BIKHCHANDANI, S., D. HIRSHLEIFER, AND I. WELCH (1992): "A theory of fads, fashion, custom, and cultural Change as informational cascades," *Journal of Political Economy*, 100, 992–1026.
- ÇELEN, B. AND S. KARIV (2004): "Observational learning under imperfect information," *Game and Economics Behavior*, 47, 72–86.

- EYSTER, E., A. GALEOTTI, N. KARTIK, AND M. RABIN (2014): “Congested observational learning,” *Game and Economic Behavior*, 87, 519–538.
- HANN-CARUTHERS, W., V. V. MARTYNOV, AND O. TAMUZ (2018): “The speed of sequential asymptotic learning,” *Journal of Economic Theory*, 173, 383–409.
- LEE, J. M. (2000): *Introduction to Smooth Manifolds*, Graduate Texts in Mathematics, Springer, 1 ed.
- MONZÓN, I. (2019): “Observational learning in large anonymous games,” *Theoretical Economics*, 14, 403–435.
- MOSSEL, E., A. SLY, AND O. TAMUZ (2015): “Strategic Learning and the Topology of Social Networks,” *Econometrica*, 83, 1755–1794.
- NARASIMHAN, R. (1968): *Analysis on Real and Complex Manifolds*, vol. 35 of *North-Holland mathematical library*, Elsevier Science Publishers, 1 ed.
- ROSENBERG, D. AND N. VIEILLE (2019): “On the efficiency of social learning,” *Econometrica*, 87, 2141–2168.
- SMITH, L. AND P. SØRENSEN (2000): “Pathological Outcomes of Observational Learning,” *Econometrica*, 68, 371–398.
- SONG, Y. AND J. ZHANG (2020): “Social learning with coordination motives,” *Games and Economic Behavior*, 123, 81–30.
- WANG, X. (2019): “Confounded observational learning with common values,” *Working paper*.
- (2021): “The set of confounded learning beliefs is (almost) homeomorphic to standard cantor set in observational learning,” *Working paper*.