Social diversity for reducing the impact of information cascades on social learning

Fernando Rosas[1,2], Kwang-Cheng Chen[3] and Deniz Gündüz[2]

[1] Centre of Complexity Science and Department of Mathematics, Imperial College London, UK

[2] Department of Electrical and Electronic Engineering, Imperial College London, UK

[3] Department of Electrical Engineering, University of South Florida, USA

Abstract

Collective behavior in online social media and networks is known to be capable of generating non-intuitive dynamics associated with crowd wisdom and herd behaviour. Even though these topics have been well-studied in social science, the explosive growth of Internet computing and e-commerce makes urgent to understand their effects within the digital society. In this work we explore how the stochasticity introduced by social diversity can help agents involved in a inference process to improve their collective performance. Our results show how social diversity can reduce the undesirable effects of information cascades, in which rational agents choose to ignore personal knowledge in order to follow a predominant social behaviour. Situations where social diversity is never desirable are also distinguished, and consequences of these findings for engineering and social scenarios are discussed.

1 Introduction

The high interconnectedness enabled by communication technologies and online media is progressively increasing the complexity of our aggregated social behaviour [1]. In fact, these complex dynamics were dramatically illustrated by the failure of our prediction tools in the forecast of recent political events, including the Brexit referendum and the latest US presidential election. A key open challenge is to clarify how the large amount of information that is constantly exchanged among individuals affects their decisions.

Fascinating dynamics take place when social agents engage in sequential decision-making. For example, most people nowadays use the Internet to check other people's recommendations prior to make decisions, which enable more informed decisions thanks to the inclusion of evidence from previous experiences. Subsequent decisions are, however, heavily influenced by earlier agents, allowing misinformation or fake news to be reinforced and spread across the social network. These non-trivial *social learning* dynamics are known to play a critical role in a number of key social phenomena, e.g., in the adoption or rejection of new technology, and in the formation of political opinions [2, 3]. Moreover, social learning also plays a key role in the context of e-commerce and digital society, e.g., in recommendation systems of online stores where users access opinions of previous customers while choosing their products [4, 5]. This is also the case in the emergence of viral media contents in various Internet portals, which are based on sequential actions of like or dislike.

A deep understanding of social learning dynamics is crucial for enabling robust platform design against fake news and data falsification, which is an urgent need in our modern networked society. As a matter of fact, digital misinformation was listed by the World Economic Forum (WEF) as one of the main threats to our modern society [6].

Social learning have been thoughtfully studied since the 90's by researchers from economics and social sciences [7–9] (for modern reviews see [2, 3]). These studies have shown that social learning is driven by two competing mechanisms. In one hand, the well-known *crowd wisdom* improve the decision-making capabilities of agents within large networks, as more information becomes available to latter agents in the decision sequence. The accumulation of social experience can, on the other hand, overload agents and generate *information cascades*, which pushes them to ignore their private knowledge and to adopt a predominant social behaviour. Interestingly, it has been shown that the combination of these two mechanisms can serve to provide network resilience against data falsification attacks [10, 11], pointing out promising possibilities for the design of resilient social learning platforms.

Motivated by the benefits that diversity can provide in biological and social systems [12, 13], in this work we study how social diversity affects the learning rate in a social learning scenario. For this, we consider a network of rational agents that have diverse preferences and prior information, having some similarities to the works reported in [14, 15]. Using a communication theoretical interpretation of this scenario, we show that social diversity is equivalent to additive noise in a communication channel —which one would expect to be detrimental for the learning process. Surprisingly, our findings show that social diversity can help to avoid information cascades, introducing important improvements in the asymptotic learning performance.

The rest of this article is structured as follows. Sections 2 introduces the considered social learning scenario, and develops our definition of social diversity. Section 3 defines information cascades, and characterize theoretically their behaviour with respect to social diversity. Section 4 presents numerical evaluations that verify the theoretical results, and finally Section 5 summarizes our main conclusions.

Notation: uppercase letters X are used to denote random variables and lowercase x realizations of them, while boldface letters $\boldsymbol{X}$ and $\boldsymbol{x}$ represent vectors. Also, $\mathbb{P}_w \{X = x | Y = y\} := \mathbb{P}\{X = x | Y = y, W = w\}$ is used as a shorthand notation.

2 Social learning model

2.1 Preliminaries and basic assumptions

Let us consider a social network composed by N agents, who are engaged in a decision-making process. In this process each agent need to make a decision between two options[1], which could correspond to a choice between two restaurants, two brands, or two political parties. It is assumed that decisions occur sequentially, and are labeled according to the order in which they take place.

The decision of the n-th agent, denoted as $X_n \in \{0, 1\}$, is based on two sources of information (see Figure 1): a *private signal* $S_n \in \mathcal{S}$, which is a continuous or discrete random variable that represents personal information that the n-th agent possesses, and *social information* given by the decisions of the previous agents, denoted by $\boldsymbol{X^{n-1}} := (X_1, \ldots, X_{n-1}) \in \{0, 1\}^{n-1}$.

All the agents are assumed to have equivalent observation capabilities, and therefore the private signals S_n are identically distributed. These signals are affected by environment conditions, which for simplicity are represented by a binary variable W. For the sake of tractability, we follow the existent literature in assuming that the private signals S_n are conditionally independent given W, leaving other cases for future work. The corresponding conditional probability distributions of S_n given the event $\{W = w\}$ are denoted by μ_w. We further assume that no realization of S_n is capable of completely determining W, which is equivalent to the measure theoretic notion of absolute continuity between μ_0 and μ_1 [16]. As a consequence of this assumption, the log-likelihood ratio

---

[1]Although generalizations for more than two options are possible, we focus in the case of binary decisions for simplifying the presentation.
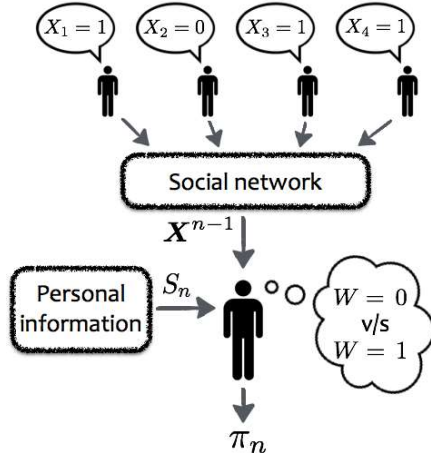
Figure 1: A social learning scenario, where an agent needs to make a decision (π_n) based on personal information coming from a private signal (S_n) and social information ($\boldsymbol{X}^{n-1}$) coming from a social network.

of these two distributions μ_1 and μ_0 is well-defined and given by the logarithm of the corresponding Radon-Nikodym derivative $\Lambda_S(s) = \log \frac{d\mu_1}{d\mu_0}(s)^2$.

A *strategy* is a rule for generating a decision X_n based on $S_n = s$ and $\boldsymbol{X}^{n-1}$, i.e. a collection of deterministic or random functions $\pi_n : \mathcal{S} \times \{0,1\}^{n-1} \to \{0,1\}$ such that $X_n = \pi_n(S_n, \boldsymbol{X}^{n-1})$.

2.2 Bayesian strategy, agents' preferences and prior information

Let us assume that the preferences of the n-th agent are encoded in an utility function $u_n(x,w)$, which determines the payoff that the agent receives when making the decision $X_n = x$ under the condition $\{W = w\}$. We consider rational agents that follow a *Bayesian strategy*, which seeks to maximize their average payoff given by $\mathbb{E}\left\{u(\pi_n(S_n, \boldsymbol{X}^{n-1}), W)\right\}$, with $\mathbb{E}\left\{\cdot\right\}$ being the expected value operator. It has been shown that the Bayesian strategy for the n-th agent can be expressed succinctly as [4]

$$\frac{\mathbb{P}\left\{W = 1 | S_n, \boldsymbol{X}^{n-1}\right\}}{\mathbb{P}\left\{W = 0 | S_n, \boldsymbol{X}^{n-1}\right\}} \underset{X_n=1}{\overset{X_n=0}{\gtrless}} e^{\nu_n} \ , \tag{1}$$

where $\nu_n = \log \frac{u_n(0,0)-u_n(0,1)}{u_n(1,1)-u_n(1,0)}$ reflects the effect of the cost function. For considering agents with diverse preferences, we assume that ν_i are independent and identically distributed (i.i.d.) random variables.

Let us further consider the case where the agents have no absolute knowledge about the prior distribution of W. Note that because W is binary, its distribution is completely determined by the value of $\mathbb{P}\left\{W = 1\right\}$. Following the framework of Bayesian inference [17], let us consider $\theta_n \in [0,1]$ to be a collection of i.i.d. random variables following a distribution $f_\theta(\theta)$ that reflects the state of knowledge of the agents about $\mathbb{P}\left\{W = 1\right\}$. In particular, if the agent has complete knowledge then $f_\theta(\theta)$ is a delta centered in the true value of $\mathbb{P}\left\{W = 1\right\}$ and hence $\theta_n = \mathbb{P}\left\{W = 1\right\}$ for all n, while if agents has no information then $f_\theta(\theta)$ corresponds to an uniform distribution over $[0,1]$.

Noting that $\boldsymbol{X}^{n-1}$ depends only on $(S_1, \ldots, S_{n-1})$, and therefore is conditionally independent of S_n, a direct application of the Bayes rule on $\mathbb{P}\left\{W = 1 | S_n, \boldsymbol{X}^{n-1}\right\}$ and $\mathbb{P}\left\{W = 0 | S_n, \boldsymbol{X}^{n-1}\right\}$

---

[2]When S_n takes a discrete number of values then $\frac{d\mu_1}{d\mu_0}(s) = \frac{\mathbb{P}\{S_n=s|W=1\}}{\mathbb{P}\{S_n=s|W=0\}}$, while if S_n is a continuous random variable with conditional p.d.f. $p(s|w)$ then $\frac{d\mu_1}{d\mu_0}(s) = \frac{p(s|w=1)}{p(s|w=0)}$.

3

shows that (1) can be re-written as

$$\Lambda_S(S_n) + \Lambda_{\boldsymbol{X}^{n-1}}(\boldsymbol{X}^{n-1}) \overset{X_n=0}{\underset{X_n=1}{\lessgtr}} \nu_n + \log \frac{\theta_n}{1-\theta_n} \quad , \tag{2}$$

where $\Lambda_S(S_n)$ and $\Lambda_{\boldsymbol{X}^{n-1}}(\boldsymbol{X}^{n-1})$ are the log-likelihood ratios of S_n and $\boldsymbol{X}^{n-1}$, respectively. Note that an efficient method for computing $\tau_n(\boldsymbol{X}^{n-1})$ has been reported in [10].

2.3 Communication theoretic interpretation

By using an adequate decision labeling, one can consider the event $\{X_n = W\}$ to be more desirable than $\{X_n \neq W\}$, or equivalently, that $u_n(1,1) \geq u_n(1,0)$ and $u_n(0,0) \geq u_n(0,1)$. The Bayesian strategy is, hence, to choose X_n as similar to W as possible using the information provided by S_n and $\boldsymbol{X}^{n-1}$. Therefore, the decisions $\pi_n(S_n, \boldsymbol{X}^{n-1}) = X_n$ can be considered to be noisy estimations of W.

To further explore this perspective, let us re-formulate (2) as

$$\Lambda_S(S_n) + \xi_n \overset{X_n=0}{\underset{X_n=1}{\lessgtr}} \tau_n(\boldsymbol{X}^{n-1}) \quad , \tag{3}$$

where $\xi_n := \log(1-\theta_n)/\theta_n - \nu_n$ and $\tau_n(\boldsymbol{X}^{n-1}) := -\Lambda_{\boldsymbol{X}^{n-1}}(\boldsymbol{X}^{n-1})$. The above can be understood as a classic signal decoder within communication theory [4, Section IV], where $\Lambda_S(S_n)$ is the decision signal and ξ_n is additive noise. Moreover, $\tau_n(\boldsymbol{X}^{n-1})$ is a decision threshold that establishes the decoding rule based on a Vonoroi tessellation that divides $\mathbb{R}$ in two semi-open intervals given by $(-\infty, \tau_n(\boldsymbol{X}^{n-1}))$ and $(\tau_n(\boldsymbol{X}^{n-1}), \infty)$.

3 Avoiding information cascades via noise

3.1 Local and global information cascades

In general, the decision $\pi_n(S_n, \boldsymbol{X}^{n-1})$ is made based in complementary evidence provided by both $\boldsymbol{X}^{n-1}$ and S_n. The n-th agent is said to fall into a *local information cascade* when the information conveyed by S_n is not included in the decision-making process due to a dominant influence of $\boldsymbol{X}^{n-1}$. The term "local" is used to emphasize that this event is related to the data fusion taking place at an individual agent. The notion of local information cascade is formalized in the following definition, which is based on the notion of conditional mutual information [18], denoted as $I(\cdot; \cdot|\cdot)$.

Definition 1. *The social information $\boldsymbol{x}_c^{n-1} \in \{0,1\}^{n-1}$ generates a* local information cascade *for the n-th agent if $I(\pi_n; S_n|\boldsymbol{X}^{n-1} = \boldsymbol{x}_c^{n-1}) = 0$.*

The above condition summarizes two possibilities: either $\pi_n(s, \boldsymbol{x}_c^{n-1})$ is constant for all values of $s \in \mathcal{S}$, or there is still variability but this variability is conditionally independent of S_n (e.g. in the case of stochastic strategies —not considered in this work). In both cases, the above definition highlights the fact that the decision π_n contains no unique information[3] coming from S_n when a local cascade takes place .

Next we define *global information cascades*, which are avalanches of local information cascades that affect all the agents after their ignition.

Definition 2. *The social information vector $\boldsymbol{x}_c^{n-1} \in \{0,1\}^{n-1}$ triggers a* global information cascade *if $I(\pi_m : S_m|\boldsymbol{X}^{n-1} = \boldsymbol{x}_c^{n-1}) = 0$ holds for all $m \geq n$.*

The relationship between local and global information cascades is explored in the next section (c.f. Proposition 1).

---

[3]For a rigorous definition of unique information in Markov chains c.f. [19].

3.2 The effect of social diversity over information cascades

Let us first introduce $F_w(z) = \mathbb{P}_w \{\Lambda_S(S_n) + \xi_n \leq z\}$ as a shorthand notation for the cumulative distribution function of $\Lambda_S(S_n) + \xi_n$ conditioned on the event $\{W = w\}$. Note that, thank to the fact that $\Lambda_S(S_1)$ and ξ_1 are independent random variables, one can compute $F_w(\cdot)$ as the convolution of their density functions.

Lemma 1. *The conditional statistics of π_n given $\boldsymbol{X}^{n-1}$ are defined by*

$$\mathbb{P}_w \{\pi_n = 0 | \boldsymbol{X}^{n-1} = \boldsymbol{x}^{n-1}\} = F_w(\tau_n(\boldsymbol{x}^{n-1})). \tag{4}$$

Proof. A direct calculation shows that

$$\mathbb{P}_w \{\pi_1(S_1) = 0\} = \mathbb{P}_w \{\Lambda_S(S_1) + \xi_1 < 0\} = F_w(0).$$

Following a similar rationale, one can find that

$$\begin{aligned}
\mathbb{P}_w \{\pi_n = 0 | \boldsymbol{X}^{n-1} = \boldsymbol{x}^{n-1}\} &= \int_{\mathcal{S}} \mathbb{P}_w \{\pi_n(s, \boldsymbol{x}^{n-1}) = 0 | S_n = s\} \, \mu_w(s) \mathrm{d}s \\
&= \int_{\mathcal{S}} \mathbb{1} \{\pi_n(s, \boldsymbol{x}^{n-1}) = 0\} \, \mu_w(s) \mathrm{d}s \\
&= \mathbb{P}_w \{\Lambda_S(s) + \xi_n < \tau_n(\boldsymbol{x}^{n-1})\} \\
&= F_w(\tau_n(\boldsymbol{x}^{n-1})) \ .
\end{aligned}$$

Above, the first equality is a consequence of the fact that S_n is conditionally independent of $\boldsymbol{X}^{n-1}$ given $W = w$, while the second equality is a consequence that π_n is a deterministic function of $\boldsymbol{X}^{n-1}$ and S_n, and hence becomes conditionally independent of W. $\square$

Next, using Lemma 1, one can show that τ_n is an effective summary of the information provided by $\boldsymbol{X}^{n-1}$ that is relevant for generating the decision π_n.

Lemma 2. *The variables $\boldsymbol{X}^{n-1} \rightarrow \tau_n \rightarrow \pi_n$ form a Markov Chain, i.e. τ_n is a sufficient statistic of $\boldsymbol{X}^{n-1}$ for predicting the decision π_n.*

Proof. Using (4), one can find that

$$\mathbb{P}_w \{\pi_n = 0 | \tau_n, \boldsymbol{X}^{n-1}\} = F_w(\tau_n) = \mathbb{P}_w \{\pi_n = 0 | \tau_n\} \ , \tag{5}$$

and therefore the conditional independency of π_n and $\boldsymbol{X}^{n-1}$ given τ_n is clear. $\square$

We now present a proposition that clarifies the relationship between local and global information cascades. This result extends [4, Theorem 1] to the current scenario.

Proposition 1. *Each local information cascade triggers a global information cascade over the social network.*

Proof. Letus first note that

$$\begin{aligned}
\tau_{n+1}(\boldsymbol{X}^n) - \tau_n(\boldsymbol{X}^{n-1}) &= \Lambda_{\boldsymbol{X}^{n-1}}(\boldsymbol{X}^{n-1}) - \Lambda_{\boldsymbol{X}^n}(\boldsymbol{X}^n) \\
&= -\Lambda_{X_n | \boldsymbol{X}^{n-1}}(X_n | \boldsymbol{X}^{n-1}) \ , \tag{6}
\end{aligned}$$

where the conditional log-likelihood is given by

$$\Lambda_{X_n | \boldsymbol{X}^{n-1}}(X_n | \boldsymbol{X}^{n-1}) = \log \frac{\mathbb{P}_1 \{X_n | \boldsymbol{X}^{n-1}\}}{\mathbb{P}_0 \{X_n | \boldsymbol{X}^{n-1}\}}.$$

5

Let us consider $\boldsymbol{x}_{\mathrm{c}}^{n-1} \in \{0,1\}^{n-1}$ such that it produce a local cascade in the n-th node. As Bayesian strategies are deterministic, local information cascades corresponds to the events where π_n is fully determined by $\boldsymbol{X}^{n-1}$, i.e. when the probability of the event $\{\pi_n = 0 | \boldsymbol{X}^{n-1} = \boldsymbol{x}_{\mathrm{c}}^{n-1}\} = \{X_n = 0 | \boldsymbol{X}^{n-1} = \boldsymbol{x}_{\mathrm{c}}^{n-1}\}$ is either 0 or 1. This, in turn, implies that $\Lambda_{X_n | \boldsymbol{X}^{n-1}}(X_n | \boldsymbol{x}_{\mathrm{c}}^{n-1}) = 0$ almost surely, and therefore, conditioned on the event $\{\boldsymbol{X}^{n-1} = \boldsymbol{x}_{\mathrm{c}}^{n-1}\}$ one has that

$$\tau_m(\boldsymbol{X}^m) = \tau_n(\boldsymbol{x}_{\mathrm{c}}^{n-1}) \qquad \text{for all } m \geq n. \tag{7}$$

Finally, by using (4), one can show that

$$\mathbb{P}_w\left\{\pi_m = 0 | \boldsymbol{X}^m, \boldsymbol{X}^{n-1} = \boldsymbol{x}_{\mathrm{c}}^{n-1}\right\} = F_w(\tau_m(\boldsymbol{X}^m)) = F_w(\boldsymbol{x}_{\mathrm{c}}^{n-1}) \tag{8}$$

Therefore, $\mathbb{P}_w\{\pi_m = 0 | \boldsymbol{X}^{n-1} = \boldsymbol{x}_{\mathrm{c}}^{n-1}, X_n, \ldots, X_m\}$ is also either zero or one, showing that the m-th agent also is affected by a local information cascade. $\qquad \square$

Let us now introduce $U_s = \operatorname{ess\,sup} \Lambda_S(S_n)$, $U_{\xi_n} = \operatorname{ess\,sup} \xi_n$, $L_s = \operatorname{ess\,inf} \Lambda_S(S_n)$ and $L_{\xi_n} = \operatorname{ess\,inf} \xi_n$ as shorthand notations for the essential supremum and infimum of $\Lambda_S(S_n)$ and ξ_n [4]. In particular, U_s and L_s correspond to the signals within $\mathcal{S}$ that most strongly support the hypothesis $\{W = 1\}$ and $\{W = 0\}$, respectively. If one of these quantities diverge, this implies that there are signals $s \in \mathcal{S}$ that provide overwhelming evidence in favour of one of the competing hypotheses. On the other hand, if U_s and L_s are both finite then the agents are said to have *bounded beliefs* [3]. Similarly, when both U_ξ and L_ξ are finite we say agents have *bounded diversity*, which implies that the diversity among priors and cost functions is not too high. Using this notions, we present the main result of this work.

Theorem 1. *Local information cascades cannot take place when agents have either unbounded beliefs or unbounded diversity.*

Proof. Note first that, due to the independency between $\Lambda_S(S_n)$ and ξ_n, one has that

$$U_{\text{total}} := \operatorname{ess\,sup} \{\Lambda_S(S_n) + \xi_n\} = U_s + U_\xi, \tag{9}$$

$$L_{\text{total}} := \operatorname{ess\,inf} \{\Lambda_S(S_n) + \xi_n\} = L_s + L_\xi. \tag{10}$$

From this, U_{total} and L_{total} are unbounded if and only the agents have unbounded beliefs or unbounded diversity.

From Lemma 1, it is clear that π_n is fully determined by $\boldsymbol{x}^{n-1} \in \{0,1\}^{n-1}$ if and only if $\tau_n(\boldsymbol{x}^{n-1})$ is such that $F_w(\tau_n(\boldsymbol{x}^{n-1}))$ is zero or 1 for $w \in \{0,1\}$. Because of the definition of F_w, this happens whenever $\tau_n(\boldsymbol{X}^{n-1}) \notin [L_{\text{total}}, U_{\text{total}}]$, proving the Proposition. $\qquad \square$

Information cascades are known to degrade the learning process, preventing the error rate of the learning process $\mathbb{P}\{\pi_n \neq W\}$ from converging to zero when the social network grows [4]. Therefore, Theorem 1 reveals a non-intuitive value of social diversity, as it can safeguard social learning from information cascades. In this way, social diversity can guarantee perfect social learning to happen asymptotically, even when agents have bounded beliefs and are hence prone to herd behaviour [4]. However, this benefit usually comes at the price of a slower convergence, which can be detrimental for the first agents of the decision sequence. This trade-off is explored in the next section.

---

[4]The essential supremum is the smallest upper bound of a random variable that holds almost surely, being the natural measure-theoretic extension of the notion of supremum [20].

4 Proof of concept

For illustrating the findings presented in Section 3, this section presents results of simulations of a social network following the model presented in Section 2. We considered two scenarios: one where S_n are binary variables that follow a binnary symmetric channel with $\mathbb{P}\{S_n \neq w | W = w\} = 1/4$, and other where S_n given $\{W = w\}$ are Gaussian variables $N(\mu_w, \sigma^2)$ with $\mu_w = (-1)^{1-w}$ and $\sigma^2 = 4$. These two signal models were choosen because it is known that agent following binary signals are strongly affected by information cascades, while agents following Gaussian signals are not affected by them (for further details about these scenarios c.f. [4, Section VI]). For simpicity, the social diversity has been modeled considering ξ_n to be i.i.d. following a Gaussian distribution $N(0, \sigma_\xi^2)$, and hence σ_ξ^2 quantifies the "diverstiy strength" of the social network. Each scenario was simulated 10^5 realizations, and the statistics of the learning error rate, defined as $\mathbb{P}\{\pi_n \neq W\}$ were computed afterwards.

In agreement with Theorem 1, results confirm that social learning processes can be benefited by social diversity. Figure 2a shows how the results of a collective inference carried out by agents driven by binary private signals achieve better performance asymptotically. However, for some values of social diversity the learning rate can be rather slow, making social learning not useful for small social networks. In all the studied cases it was seen that social diversity degrades the performance of the first agents in the decision sequence; however an adequate level of diversity can introduce a fast learning rate. In contrast, as illustrated in Figure 2a for agents following Gaussian signals, social diversity was found to be always detrimental in cases where agents have unbounded beliefs. This confirms the fact that the benefits of social diversity is to avoid information cascades, which are the main cause of poor performance of social learning in large networks [4].
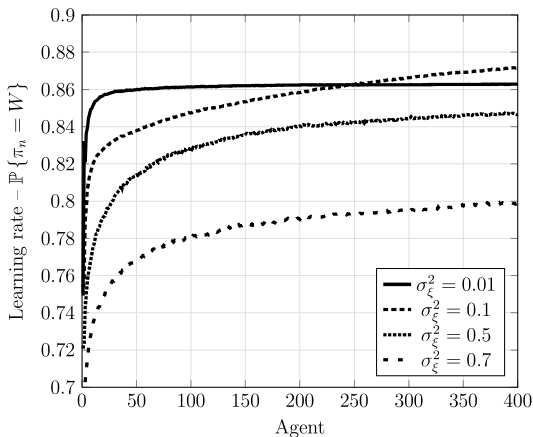
(a) Binary private signals
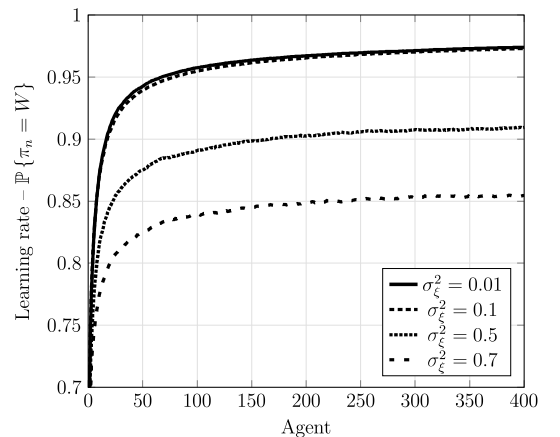
(b) Gaussian private signals

Figure 2: Social learning rate for agents following binary or Gaussian private signals, under various levels of social diversity (σ_ξ^2). Social networks that follow binary signals are vulnerable to information cascades, and hence a non-zero social diversity improve their asymptotic learning rate. In contrast, social networks that follow Gaussian signals are inmune to information cascades, and hence social diversity have a purely detrimental effect.

The different effect that social diversity has over agents located at different positions in the inference process is further illustrated by Figure 3. We found that, for each agent, there exists an optimal level of social diversity that reduces the effect of information cascades without introducing too much noise. Agents located in the first places of the decision sequence are always affected negatively by social diversity, and hence for them is optimal to have $\sigma_\xi^2 = 0$.
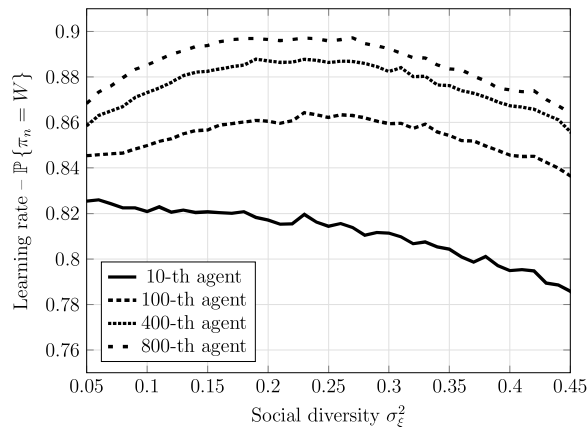
Figure 3: An optimal level of social diversity exist that can improve the social learning performance of agents located late in the inference process. However desirable, this better performance of late agents comes at the expense of a detrimental effect to the first agents.

5 Conclusion

This paper aims to undestand how social learning is affected when it is pursued by a diverse population. Our scenario considered rational agents with heterogeneous preferences, as encoded by their utility functions, and diverse prior information about the target variable. A communication theoretic analysis showed that this kind of social diversity is equivalent to additive noise in a communication channel. However, it was found that an unbounded social diversity prevent information cascades and, hence, introduces important improvements into the asymptotic social learning rate that can be achieved by a population. Social learning is, therefore, one of those rare cases where noise can improve the overall performance.

To understand how can noise be beneficial, let us point out that rational social agents maximize their individual performance while ignoring the consequences of their actions on the aggregated behaviour. This selfish quality of the agent's behavior makes their actions locally optimal while being globaly suboptimal. In this context, the heterogeneity introduced by social diversity makes the decisions of each agent less informative to others. This generates a reduced social pressure that, in turn, prevents information cascades and herd behaviour, introducing great improvements in the asymptotic social learning performance.

It is to be noted that the benefits of social diversity are only experienced by agents that are prone to information cascades. Therefore, social diversity is not beneficial, e.g., for agents with unbounded beliefs. However, in most applications agent's beliefs are bounded, either because their signals information content is limited or because the signals themselves are bounded. The latter is the case in most engineering applications, e.g. in the scenario studied in [10].

Finally, it is important to remark that social diversity provides benefits to the latter agents in the decision sequence, while degrading the performance of the first agents. Therefore, social diversity might in general be detrimental for the performance of social learning in small networks.

Acknowledgements

References

[1] Y. Bar-Yam, "Complexity rising: From human beings to human civilization, a complexity profile," *Encyclopedia of Life Support Systems (EOLSS), UNESCO, EOLSS Publishers, Oxford, UK*, 2002.

[2] D. Easley and J. Kleinberg, "Networks, crowds, and markets," *Cambridge University Press*, vol. 1, no. 2.1, pp. 2–1, 2010.

[3] D. Acemoglu, M. A. Dahleh, I. Lobel, and A. Ozdaglar, "Bayesian learning in social networks," *The Review of Economic Studies*, vol. 78, no. 4, pp. 1201–1236, 2011.

[4] F. Rosas, J.-H. Hsiao, and K.-C. Chen, "A technological perspective on information cascades via social learning," *IEEE Access*, vol. 5, pp. 22 605–22 633, 2017.

[5] J. Hsiao and K. C. Chen, "Steering information cascades in a social system by selective rewiring and incentive seeding," in *to be included in 2016 IEEE International Conference on Communications (ICC)*, 2016.

[6] L. Howell, "Digital wildfires in a hyperconnected world," *WEF Report*, 2013.

[7] A. V. Banerjee, "A simple model of herd behavior," *The Quarterly Journal of Economics*, pp. 797–817, 1992.

[8] S. Bikhchandani, D. Hirshleifer, and I. Welch, "A theory of fads, fashion, custom, and cultural change as informational cascades," *Journal of political Economy*, pp. 992–1026, 1992.

[9] ——, "Learning from the behavior of others: Conformity, fads, and informational cascades," *The Journal of Economic Perspectives*, vol. 12, no. 3, pp. 151–170, 1998.

[10] F. Rosas and K.-C. Chen, "Social learning against data falsification in sensor networks," in *International Workshop on Complex Networks and their Applications*. Springer, 2017, pp. 704–716.

[11] F. Rosas, K.-C. Chen, and D. Gunduz, "Social learning for resilient data fusion against data falsification attacks," *arXiv preprint arXiv:1804.00356*, 2018.

[12] J. Mathiesen, N. Mitarai, K. Sneppen, and A. Trusina, "Ecosystems with mutually exclusive interactions self-organize to a state of high diversity," *Physical review letters*, vol. 107, no. 18, p. 188101, 2011.

[13] F. C. Santos, M. D. Santos, and J. M. Pacheco, "Social diversity promotes the emergence of cooperation in public goods games," *Nature*, vol. 454, no. 7201, p. 213, 2008.

[14] L. Smith and P. Sørensen, "Pathological outcomes of observational learning," *Econometrica*, vol. 68, no. 2, pp. 371–398, 2000.

[15] V. Bala and S. Goyal, "Conformism and diversity under social learning," *Economic theory*, vol. 17, no. 1, pp. 101–120, 2001.

[16] M. Loeve, *Probability Theory I*. Springer, 1978.

[17] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian data analysis*. CRC press Boca Raton, FL, 2014, vol. 2.

[18] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[19] F. Rosas, V. Ntranos, C. J. Ellison, S. Pollin, and M. Verhelst, "Understanding interdependency through complex information sharing," *Entropy*, vol. 18, no. 2, p. 38, 2016.

[20] J. Dieudonne, *Treatise on Analysis*. Associated Press, New York, 1976, vol. II.