A Technological Perspective on Information Cascades via Social Learning

Fernando Rosas, Jen-Hao Hsiao and Kwang-Cheng Chen, *IEEE Fellow*

***Abstract*—Collective behavior in human society is attracting a lot of attention, particularly due to novel emergent phenomena associated with online social media and networks. In effect, although crowd wisdom and herding behaviour have been well-studied in social science, the rapid development of Internet computing and e-commerce brings further needs of in-depth comprehension of their consequences and impact from a technological perspective. Based on *social learning*, an analytical knowledge originated in social science, we re-examine the well-known phenomenon of *information cascade* where rational agents can ignore personal knowledge in order to follow a predominant social behaviour triggered by earlier decisions made by peers. Moreover, we look into the cascade behavior from a communication theoretic perspective, interpreting social learning as a distributed data processing scheme. This perspective enables the development of a novel framework, which allows a characterization of the conditions that trigger information cascades and trace their impact on the accuracy of the collective inference. Finally, potential applications and examples of information cascade have been presented under various cyber technological scenarios, illustrating the prolific interplay between communication technology and computational social science.**

I. Introduction

The surprising outcomes of recent political polls, such as the Brexit referendum and the latest US presidential election, is revealing the limitation of our current understanding of social behaviour in a highly-interconnected world. It has been claimed that, similar to the way in which evolution takes place among living species, human society evolves in time from simple to more complex forms of organization and behavior [1]. One of the distinctive and more challenging characteristics of complex systems, which has been widely acknowledged in social scenarios, is that the aggregation of the activities of simple components or agents can generate complex and unpredictable outcomes [2]–[4]. Therefore, just as thermodynamics and statistical mechanics went beyond classical mechanics in order to provide an adequate framework for the description of gasses and liquids, a new theory might be necessary in order to enable a deeper understanding of important phenomena that characterize modern society [1].

New information dynamics are defying our traditional tools of analysis, which were forged in times when the world was simpler and easier to predict [5], [6]. In the early days of Internet, the promises of abundance of information and the anti-authoritarian structure were thought to be seeds that would bring great benefits to society [7]. However, it has been shown that the heterogeneity and absence of recognized information sources can generate doubts about causation, which in turn can stimulate speculations and misinformation [8]. Moreover, the excess of available information and the limited processing capabilities of individuals trigger confirmation biases, which stimulate the exclusive use of information sources that support one's existing beliefs or points of view [9]. Furthermore, online recommendation algorithms constantly and invisibly filter user's queries, presenting contents that might better satisfy the user's profile and preferences. All these elements are creating so-called digital echo chambers, where disjoint groups of society are progressively reinforced in their beliefs —whatever they might be [10]. The digital misinformation that these mechanisms are generating is so severe that the World Economic Forum (WEF) listed this as one of the main threats to our modern society [11].

In order to address these issues, one big challenge is to clarify the effects and consequences of the large amount of information that is constantly generated and exchange between individuals in a digital society [12]. As a matter of fact, the massive deployment and use of Internet mobile terminals and devices is enabling massive information networks, making the global mobile data traffic of 2015 to grow more than 74% reaching 3.7 exabytes per month, being driven by 2.7 billion connected devices [13]. Moreover, social habits are evolving concurrently with the pervasive use of Internet, making social networks an essential tool for social interaction and information exchange [14]–[16]. For example, most people nowadays use the Internet to check other people's recommendations prior to making decisions for traveling, buying a product or choosing a restaurant. In these cases, subsequent decisions are influence by earlier agents, which allows possible misinformation and cascades across the network. Such complex interactions may defy intuition and are difficult to predict, and therefore an in-depth understanding of the inner mechanisms is very much desirable.

In particular, one crucial technological goal is to understand the way in which decision making is affected by the mechanisms of large distributed data processing [17]. In addition to social science and Internet computing, this knowledge is further crucial to system automation and cyber-physical systems, machine learning and artificial intelligence [18]. It is noted that most engineering approaches focuses on a combination of distributed information gathering and centralized computing.

F. Rosas and J.H. Hsiao are with the Graduate Institute of Communication Engineering, National Taiwan University, Taipei 10617, Taiwan (email {nrosas, r04942065}@ntu.edu.tw). K.C. Chen is with the Department of Electrical Engineering, University of South Florida, Tampa, FL 33620, USA (email kwangcheng@usf.edu), and was with the Graduate Institute of Communication Engineering, National Taiwan University, Taipei 10617, Taiwan. This research is partially supported by the U.S. Air Force under Grant AOARD-14-4053, and by Mediatek Inc. and the Ministry of Science and Technology, Taiwan, under Grant 104-2622-9-002-002.

However, large-scale deployment of intelligent systems/agents allows distributed processing by agents who sequentially collect and exchange data and perform local inferences, which allows the attainment of more graceful scaling properties with respect to the network size [19]–[21].

Decision making based on distributedly sensed information was intensively studied during the 1980s and early 1990s in the context of distributed radar systems, where the goal was to design schemes that detect events as accurately as possible considering various communication restrictions (c.f. [22]–[24] and references therein). Unfortunately, it has been shown that the design of optimal schemes for information transfer and distributed processing for the general case is NP-hard [25]. In fact, although in many cases these schemes can be described as a set of thresholds against which likelihood functions must be compared, the determination of the optimal thresholds is in general an intractable problem [22]. For example, it has been shown that using equal thresholds can be suboptimal even for the simple case of a network of identically distributed sensors arranged in a star topology [26], being only asymptotically optimal for large networks [22], [27]. Moreover, symmetric strategies are not useful for more complex network topologies (e.g. [28], [29]), and hence heuristic methods for finding the thresholds are necessary. Renewed interest on this problem took place after the emergence of wireless sensor networks, by considering the effect of noise, outage events and the impact of energy and bandwidth constraints (c.f. [30], [31] and references therein). Other aspects have also been analysed, including robust distributed estimation [32], [33], multi-objective optimization [34], distributed parameter optimization, tracking and many others [35]–[37]. However, most of these works focus on networks of star topology and are based on very distinctive roles for regular nodes and fusion centers, which makes them not well-suited for large-scale distributed networks with ad hoc topologies.

In parallel, remarkable efforts have been made in economics and social science to analyze sequential information processing and learning on social networks (an extensive literature review about social learning is provided in Section III). In these models, agents make decisions combining private knowledge with social information that they gather from looking at theirs peers' actions. Interestingly, it has been shown that the aggregation of rational decisions can generate irrational global behaviour, degrading the "wisdom of the crowds" into mere herd behaviour. This phenomenon, called *information cascades*, arises when the social information overloads agents, forcing them to ignore their private knowledge and to adopt the predominant social behaviour (an introduction and literature review on information cascades is provided in Section II). It is believed that information cascades play crucial roles in the formation of political opinions, the adoption or rejection of new technology and many other important social phenomena [38]. There have existed further interest in understanding the role of information cascades in the context of e-commerce and the digital society [39]. For example, information cascades can have tremendous consequences in online stores where customers can review the opinion of previous customers before deciding to buy a product, or in the emergence of viral media contents based on sequential actions of like or dislike [40]. Therefore, developing further in-depth analytical understanding on the mechanisms that trigger information cascades and their effect on social learning emerges as a fundamental issue in modern human society.

The main motivation behind this article is to explore social learning as a distributed signal processing method, building a bridge between the research done separately by economists and sociologist, and electrical engineers and computer scientists. We intend to provide a quantitative framework to analyse the impact of information cascades over the performance of social learning, while the current literature primarily pays attention to the conditions that guarantee the achievement of a perfect inference asymptotically (i.e. the inference error rate goes to zero). Furthermore, following an engineering perspective, social learning can be quite useful as a distributed data processing scheme for a range of applications if the error rate does not goes to zero but can be bounded below some critical value. Therefore, we develop novel upper bounds for the asymptotic inference performance, which can be used as a design guide for applying social learning in engineering applications. Moreover, these bounds provide fundamental insights that delineate the way in which information cascades influences the asymptotic error rate of social learning. Finally, our framework also provides analytical formulas for the exact performance of each agent, allowing an efficient exploration of the error rates in non-asymptotic regimes.

The rest of this article is structured as follows. Sections II and III present an introduction to information cascades and social learning, presenting the fundamental ideas and discussing some of the relevant literature. After this, Section IV presents our novel perspective of social learning as a data aggregation scheme. Our main results about the characterization of information cascades, related to this novel perspective, are presented in Section V, and are then illustrated for the case of social networks driven by binary signals in Section VI. Section VII uses the results of Section V to derive novel bounds for the achievable performance of social learning, providing a deeper understanding of the impact of information cascades over collective signal processing. These results are then illustrated by numerical evaluations in Section VIII. Sections IX and X discuss several important applications that information cascades have in prominent cyber technological scenarios, including cyber physical security and machine learning. Finally, Section XI summarizes our main conclusions.

II. An introduction to information cascades

This section present a general introduction to information cascades, reviewing the state of the art and providing the necessary background for the unfamiliar reader. In the sequel, first Section II-A discusses some fundamental aspects and provide historical remarks. Then, Section II-B presents some social implications, and provides a preliminary definition of what an information cascade is (which is latter revisited in Section V-A). Finally, statistical approaches to study information cascades are discussed in Section II-C.

A. Fundamentals of group decisions and social influence

Intuitively, the decision made by a group of agents can be much more accurate than the one made by an isolated individual. This phenomenon, known as *wisdom of the crowds*, has been acknowledged and experimentally verified in diverse contexts by researchers of economics, psychology and sociology [41]–[43]. The origin of this idea is commonly traced back to a work written in 1907 by Sir Francis Galton, who was cousin of Charles Darwin [44]. By examining the results of a "guess the weight of the ox" competition in a country fair, Galton noted that the median of all the estimations was particularly accurate, being closer to the real value than the guesses made by experts. This is somewhat related to the law of large numbers, where the process of averaging can keep the consistent part of a signal while reducing non-coherent noisy elements. It has to be noted that the trust on the accuracy of aggregated opinions is not a mere theoretical tool, being deeply rooted into the popular mind and influencing stock markets, political elections, and quiz shows [45].

Interestingly, it is also commonly acknowledged that in real life the wisdom of the crowd is far from infallible. In effect, it has been noted that the effectiveness of this phenomenon requires two main principles: decentralization (allowing diversity of opinion and independence of noise and errors) and aggregation [46]. The failure of any of these principles can severely degrade the accuracy of the wisdom of the crowds.

Already renowned philosophers, like Soren Kierkegaard and Friedrich Nietzsche, noted that aggregated behaviour can degrade into mere herd dynamics, where people follow the prominent social behaviour without judging it with critical thinking [47]. More recently, a number of studies have shown how social influence can affect individual decisions, compromising the results of estimation tasks, price determination and even music preferences [43], [48]–[50]. Social interaction undermines the independency of individual opinions, as individuals are usually aware of each other's decisions and this might induce them to review their own estimates [51]. After being aware of peer's choices, agents might want to modify their opinions due to peer pressure towards conformity or a suspect that others might have better information [52], [53]. As a simple example, many people like buying popular products and think that, if many people liked it, it cannot be bad. In particular, [50] analyses the behaviour of agents contrasting the results when they are aware or not of other's decisions. The corresponding experimental evidence shows that the knowledge of other's decisions effectively reduces the diversity of opinions, which in turns degrades the effectiveness of the wisdom of crowds. In this way, one can see how the aggregation of rational decisions can generate irrational global behaviour.

B. Preliminary definition and social implications

Social agents usually are faced with the dilemma of how to act when the information gathered from their social network contradict their private conclusion. An *information cascade* takes place when an agent actually chooses to ignore private information in order to follow the predominant social behaviour or preference, in despite of possible contradictions with his/her personal information [38]. The term "cascade" comes from the fact that once one agent has cascaded, this increases the social pressure for future agents to take a similar decision. Therefore, these events can easily spread over large portions of a social network.

Information cascades have been proposed as an explanation of how the "wisdom of the crowds" can transform into herd behaviour [54]. Moreover, it has been suggested that information cascades play crucial roles in many politic, economic and social phenomena [38]. For example, some companies provide early sales or early tests opportunities to trigger cascades of purchasing decisions [55], [56]. In markets with a monopolistic seller and buyers that are aware of each other's purchases, it has been shown that the monopolist has incentives to alter the good's prices in order to induce herding [57]. With respect to the extremely slow adoption of a clearly more convenient hybrid seed corn during the Great Depression, researchers suggest that is was due to the higher trust that farmers had for their neighbours over the information provided by the corresponding salesman [58]. Also, the dangers of information cascades over political preferences has been acknowledged by countries such as Israel and France, who have made laws to prohibit polling during the days or weeks before elections in order to avoid a cascading influence over the citizens [59].

A renewed interest about information cascades have emerged recently with respect to the social dynamics that take place in massive e-commerce and e-marketing platforms [39]. The steering or manipulation of information cascades phenomena could have a big impact over these systems, which raises concerns about their safety and resilience from malicious attacks or dishonest users. Developing a clear understanding of the triggering conditions and range of the effects of information cascades is therefore of fundamental importance, as this can enable the design of secure and trustable platforms that are crucial for a prosperous digital society.

C. Statistical approaches

The mechanisms behind the wisdom of the crowds and information cascades are of statistical nature, and hence they are not restricted to social or psychological phenomena. According to this rationale, statistical Social Learning may supply a theoretical tool to develop a deep understanding of information cascades. While Social Learning is discussed in Section III, other methodologies are introduced in the sequel.

1) Information diffusion: Empirical studies of the structural characteristics of information cascades can be done by analysing user-generated contents collected from records of online social networks. For analysing this data, some approaches (e.g. Social Learning, c.f. Section III) adopt a microscopic viewpoint where cascades are considered to be the consequence of the decision patters of specific users. In contrast, a large volume of literature in computer science avoids these complexities by adopting a macroscopic perspective, where cascading contents are analysed as a case of information diffusion over a network. In particular, most of the works focus on studying the behavior of statistical properties such as cascade length, tree size and link distribution [60]–[66]. Moreover, this approach enables the use of techniques from algebraic graph theory to characterize the diffusion process, including spectral analysis of the Laplacian or network adjacency matrices [67].

2) Dynamical systems analysis: An alternative framework to study information cascades have been presented in [68], which is based on a novel connection found between sequential decision processes and dynamical systems. In this work the authors consider a group of agents that have to sequentially make a binary decision. The decision of the n-th agent, X_n, is generated by considering the information carried by a private signal S_n and the previous decisions $X_1, \ldots, X_{n-1}$. Moreover, in order to consider irrational and stochastic aspects of human decisions, it is assumed that X_n is stochastically generated following a Bernoulli random variable with parameter $p_n = u_n(R_n, S_n)$, where $u_n(\cdot,\cdot)$ is an utility function and R_n is the output of a pre-decision filter given by

$$R_n = \frac{1}{n-1}\sum_{j=1}^{n-1} \mathcal{S}_{n,j}(X_j) \tag{1}$$

where $\mathcal{S}_{n,j}$ is a collection of stochastic functions that map $\{0,1\}$ to $\mathbb{R}$. The pre-decision filter model is motivated by the tendency of people to have selective attention and hence project high-dimensional signals into low-dimensional representations. Hence, R_n represents a lossy compressed version of the decisions of previous agents.

Interestingly, it has been shown that the evolution of the time series R_n can be traced down using tools from dynamical system theory. A very intuitive understanding of the evolution of this system can attained by considering the plot of a specific utility function, over which the evolution of R_n corresponds to a standard dynamical system. Following this rationale, stable and unstable points in the evolution can be defined, which correspond to the values towards which R_n converges almost surely (see Figure 1).

This simple model allows to develop a clear intuition over the aggregated effect of the social information over the evolution of p_n. In effect, the graphical perspective provided by the dynamical system representation enables an insight on how, depending on the structure of the variables $\mathcal{S}_{n,j}$, some situations are more likely to make R_{n+1} to be larger than R_n, or vice-versa. The decomposition of the space of possible values of R_n based on attractors, attractive regions and unstable points allows a whole new approach to the study of information cascades, whose exploration has just started.

3) Empirical approaches: Since the idea of information cascades was published in the economics literature [54], [69], a number of empirical studies of information cascades have been carried out. For example, [70]–[72] design experiments to investigate the information cascades phenomenon in sequential decision making by human test subjects. In [70], test subjects were ask to make predictions about the predominant colour of a collection of balls inside an urn, after looking to a randomly chosen one. Following the prediction provided by

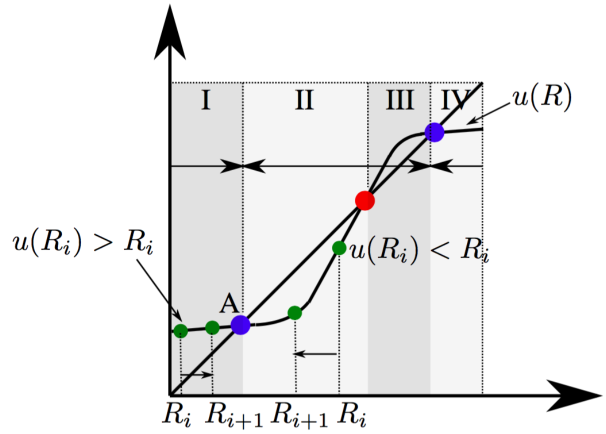


Fig. 1: Evolution of sequential decision making viewed as a dynamical system. Studying the evolution of R_n, as defined in (1), stable and unstable points can be defined, which correspond to the limiting values of this time series [68].

the information cascade theory, if some few initial decisions coincide, then the subsequent decisions tend to follow the established pattern ignoring the result of the actual withdraw.

A simple operational model for information cascades was presented in [73], being based exclusively on variables that can be observed from real data. The model was tested on data related to the adoption of electronic commerce technologies, showing that information cascades play a major role in such processes.

The report presented in [74] studies the effect of information cascades in data related to how people choose which movie to see in the theater. The results suggest that the cascading behaviour influence the box-office revenue and profits characteristics, which are governed by Levy distributions with infinite mean and infinite variance. This, in turn, might actually explain the inherent difficulty of making accurate predictions in those scenarios.

A recent research trend is to use both social learning modeling and data analytics [75], [76] from online platforms [77], [78]. Some topics considered by this literature include complex user behavior analysis and prediction [65], [66], [79], and control [80] or steering [39], [81], [82] of complex social system phenomena. The interested reader can find a detailed analysis of [39] in Section X-A.

III. How statistical learning takes place in social networks

This section provide a general introduction to social learning. The pioneer works of Bayesian social learning are discussed in Section III-A, and then Section III-B reviews the contributions of more recent works. Aspects of Non-bayesian social learning are then discussed in Section III-C, and finally Section III-D present some open question.

A. Early efforts

Social learning was initially investigated by [54], [69], [83], [84] by analyzing sequential decision-making processes in social networks. In these systems each agent has to make one

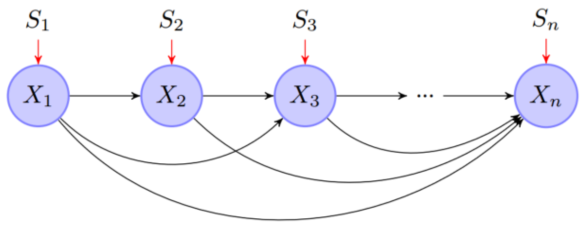


Fig. 2: The Sequential Social Learning Model. The variables S_n represent personal evidence, while the arrows between various decisions X_j illustrate how decisions influence each other in the social system.

decision following a pre-defined fixed ordering. The decision of the agent that decides in the n-th place of the sequence, denoted as $X_n \in \mathcal{X}_n$, is based on two sources of information (see Figure 2). In one hand, agents possess personal knowledge about the corresponding subject, which is represented by the random variable $S_n \in \mathcal{S}_n$ that is only available to the corresponding agent. These private signals can be discrete or continuous depending on the cardinality of the signal space $\mathcal{S}_n$. Also, these signals are correlated with a variable that represents the "state of the world", denoted as W, which is unknown to the agents. Secondly, the n-th agent is aware of the decisions made by all the previous agents, denoted as $\boldsymbol{X}^{n-1} = (X_1, \dots, X_{n-1})$. This allows each agent to learn from the examples of previous agents in order to improve their decision accuracy.

An interesting question is how the agents combine these two heterogeneous sources of information in order to optimize their decisions. A crucial assumption adopted in these works is that the agents act based on perfect rationality making *Bayesian decisions*, which maximize the average value of a cost function that quantifies their personal benefit. This cost function is traditionally assumed to be affected by the agent's own decision X_n and the state of the world as encoded by W. An interesting result is that social learning allows individuals with uninformative private signals to harvest information from other's private signals by copying their decisions. Therefore, even if S_n is not very informative, the n-th agent can still gain indirect access to the information conveyed in $S_1, \dots, S_n$ by considering $\boldsymbol{X}^{n-1}$, which represent a lossy compressed version of the other agents' private signals. This is an embodiment of the wisdom of the crowds (c.f. Section II-A), as the evidence provided by early agents' decisions can be successfully aggregated for guaranteeing a surprisingly improved accuracy for the decision made by later agents. This phenomenon is particularly remarkable for the case of binary decisions (i.e. $X_n \in \{0, 1\}$) and complex private signals, as from a distributed sensing perspective this can be seen as a distributed data fusion scheme that requires small amounts of information exchange.

However, these efforts suggest further room to enhance social learning, as under certain conditions the aggregation of rational decisions can generate irrational global behaviour and information cascades (c.f. Section II-B). In effect, information cascades arises in social learning when the social information becomes so persuasive that all subsequent agents ignore their personal knowledge and adopt an homogeneous behaviour, and hence $X_m = X_{n_c}$ for all $m > n_c$. This is usually an undesired phenomenon that stops the inclusion of new evidence in the inference process, as further agents discard their own private information to blindly follow the prevailing behavior.

In summary, the initial research succeeded in showing that information cascades effectively can take place within social systems, but could not provide a general understanding of their nature and generating causes.

B. Effect of cost functions and network topology

Motivated by these initial findings, researchers aimed to deepen the understanding of the mechanisms of social learning by extending the original models by considering more general cost metrics, assuming that the cost functions and priors possessed by different agents could disagree [85]–[88]. Their results show that the cost function plays a crucial role in generating homogeneous asymptotic decisions, or in allowing heterogeneous behaviours to coexist. The work reported in [87] shows that diversity in the agent's preferences can undermine the trust that agents give to each other, decreasing the interest in conformity and allowing different behaviours to coexist asymptotically. In fact, [86] presents some examples where the trust is undermined to such an extent that the past history does not provides valuable information to future agents, and hence they have to rely exclusively on their own private signals.

In addition, [86] proved that the asymptotic accuracy of social learning is not perfect if the information conveyed by the private signals is bounded. Concretely, let us consider binary decisions (i.e. $\mathcal{X}_n \in \{0, 1\}$) and a binary state of the world variable W. Then, under very general conditions, it can be shown that the maximization of the utility function is equivalent to making X_n as much similar as W as possible given the available information (c.f. the corresponding discussion in Section IV-C). Perfect asymptotic accuracy is hence equivalent to guarantee that

$$\lim_{n \to \infty} \mathbb{P}\{X_n \neq W\} = 0 \ . \tag{2}$$

Now, for each possible private signal realization $S_n = s \in \mathcal{S}_n$ one can compute the likelihood ratio of s taking place under the event $\{W = 1\}$ versus $\{W = 0\}$ (for a precise definition of the private signal likelihood see Section IV-A). It is well-known that in general these likelihood ratio values are sufficient statistics for estimating W based on S_n [89], and hence due to the above discussion they contain all the information relevant for generating X_n. Following this, [86] proved that (2) is not satisfied if the likelihood ratio values of the private signals are bounded above and bellow by given constants. Note that one of the consequences of this result is that social learning cannot attain perfect asymptotic accuracy when the private signals are drawn from finiten spaces.

Further important insights about social learning were achieved by studying the effects of the social network topology on the aggregated behaviour [39], [56], [90]–[92]. Hence, although the original models of social learning assumed that the n-th agent act based on the knowledge of all the previous decisions $\boldsymbol{X}^{n-1}$, these works consider a more general case where agents only have access to a limited subset of the previous decisions. The neighbourhood of the n-th agent is normally defined as the set of agents which are connected to him/her by the social network, which are denoted as $\mathcal{B}_n$. Correspondingly, in these scenarios the decision X_n is made considering the information of S_n and $\boldsymbol{X}_{\mathcal{B}_n}$, where the latter represents the vector of decisions of the neighbours of the n-th agent. Interestingly, the above works consider both deterministic and stochastic social network topologies, the later being characterized by random neighbourhoods.

In [91] the authors provide various conditions over the network topology and private signal structure that guarantee or forbid perfect asymptotic social leaning. Their results show that asymptotic learning does not take place if there is a group of agents that are "too influential", i.e. if there exist a infinite number of individuals influenced exclusively by a specific finite subset of agents. On the other hand, it is also shown that perfect asymptotic learning takes place if there are no too influential groups and the private signal likelihood is unbounded.

As a partial converse result, a number of network topology characteristics have been presented which, when combined with bounded private signal likelihood, makes it impossible to achieve perfect learning. This extends the result of [86] for the more general case of a (possibly random) social network topology. However, [91] also provides a fascinating example of a network structure that allows perfect asymptotic learning even in presence of bounded private signal likelihood. This topology is characterized by two different kinds of agents: innovators, which have few social connections and hence are likely to follow the dictate of their own private signal, and gatherers, whose highly connected network location allows them to synthesize previous opinions. This inspiring heterogeneous network showed how an adequate topology can undo the limitations of the private signal structure, allowing the system to reach a perfect asymptotic inference.

C. Non-bayesian social learning

All the works discussed so far are focused on Bayesian learning, where agents choose the actions following perfect rationality. However, important research efforts have been made in parallel in non-bayesian social learning models, where agents use simple rule-of-thumb methods to combine their private information and the one that comes from the social signals [93]–[97].

A vast part of the literature that studies non-bayesian social leaning is inspired on a model presented in [93], where agents combine the neighbours' opinions additively. In particular, this work considers a group of agents who need to make estimates of the value of a parameter θ. The prior knowledge of the n-th agent about the parameter is represented by a probability distribution over the possible values, denoted as $f_n(0)$. Hence, by considering non-negative constants $q_{i,j}$ that represent the trust of agents on each other, the process of exchanging opinions is modeled as

$$f_n(t) = \sum_{j=1}^{N} q_{n,j} f_j(t-1) \ , \tag{3}$$

where t is an non-negative integer. This method of fusing information is algebraically simpler than the one used by Bayesian agents, allowing to perform detailed analyses based on techniques based on Markov chain theory [98, Chapter 6]. In fact, the iterative process can be represented by

$$\vec{f}(t) = Q\vec{f}(t-1) = Q^n \vec{f}(0) \tag{4}$$

where $\vec{f}(t) = (f_1(t), \ldots, f_n(t))^t$ and $\boldsymbol{Q}$ is the matrix with entries $q_{n,j}$. Therefore, it is shown that the asymptotic value $\lim_{t\to\infty} \vec{f}(t)$ is governed by the properties of Q^n for large values of n. Therefore, using standard tools like the Perron-Frobenius Theorem [98], it is possible to predict when the agents achieve asymptotic agreement. Interestingly, if consensus is reached its value is an afine combination of the original opinions.

In [95] and [97], it is shown that simple ad hoc updating rules can still achieve asymptotic correct inference, even when they are unaware of important aspects like the social network topology or the signal structure of other agents. The system model of [95] closely follows the model presented in [93], although it explores the connection of Q with the social network topology and other aspects. On the other hand, [97] introduces a constant arrival of new information, which allows the study novel phenomena like asymptotic learning in finite networks. Both papers acknowledge that, although somehow surprisingly simple rules can achieve asymptotic learning, the simplicity of the rule might have a strong negative effect on the learning rate and the corresponding speed of convergence.cy

The effect of the network topology on the asymptotic performance of non-bayesian social learning was studied in [96] using a similar model than the one proposed in [93], but introducing random matrices Q in order to represent stochastic social networks. This work also considers heterogeneous agents, some of them being stubborn and hence less likely to modify their initial opinions. Their analysis shows that the evolution of the social learning is connected to the matrix $\tilde{W}$, which can be expressed as

$$\tilde{W} = T + D \ . \tag{5}$$

Above, the matrix T is governed by the probabilities of agent interaction and hence is related to the social structure of the group, while D represents the influence structure that quantifies which agents are more or less willing to adapt their decisions to follow their neighbours. Interestingly, it is shown that under general conditions $\tilde{W}^n$ converges to a matrix with equal columns, and that the row vector determines the value of the social consensus. Moreover, they provide bounds of possible deviations from asymptotic learning that can be produced by an excessive influence of some agents. One of these bounds is based on the spectral gap of T, which is a well-known measure of the social network connectivity (related to the second largest eigenvalue of the matrix [67]).

There is an interesting ongoing debate about if Bayesian or non-Bayesian frameworks are more suitable to describe human endeavour. In a nutshell, although Bayesian models are elegant and tractable, they assume that agents act always rationally [99], and consequently make unrealistic assumptions about their knowledge of posterior probabilities that are related to non-trivial aggregated social interactions [97]. However, Bayesian models provide an important benchmark, not necessarily due to their accuracy but because their simplicity allows to develop important insights about the nature of the dynamics of aggregated decisions, providing important reference points for discussing non-bayesian models as well [92]. On the other hand, the use of ad hoc information fusion methods in the non-bayesian social learning literature makes it difficult to attain general results.

Although highly desirable, it is challenging to compare the results from bayesian and non-bayesian social learning in a meaningful and fair way. In fact, while agents in Bayesian systems usually decide once and do not adapt their decision further, most of the non-bayesian learning literature consider agents that update their decision regularly, being closer to information diffusion processes (c.f. Section II-C1). One way of filling this gap is by exploring single decision non-bayesian social learning. This has been done in some recent literature [68], [100], [101], which also addresses non-asymptotic properties which were not well-explored before. For example, [100] proposes a data fusion scheme which combines Bayesian updating for processing private information and an ad-hoc combination based on a Gibbs measure for synthesising the social information. They show this scheme achieves exponential convergence, and moreover quantify the dependencies of the learning rate over different learning rules and communication constraints using large deviation theory. In [101] there is a comparison of the performance that are attainable using different kind of data aggregation rules, including additive averaging and majority rules. However, it is difficult to compare their performance results with previous works as their model focus on a final decision that uses as input all the partial decisions generated in the learning process. The main aspects of [68] are discussed in Section II-C2.

D. Asymptotic learning and information cascades

The existent literature suggests a direct connection between imperfect asymptotic learning and information cascades. In most sequential decision processes, the progressive amending of new evidence generates the eventual achievement of perfect inference, at least asymptotically. This can be verified in the case of classic hypothesis testing [102] and in various distributed hypothesis testing schemes [36], where the convergence to a perfect inference is attained exponentially fast. However, one of the distinctive characteristics of social learning —acknowledged since the early efforts— is that under some conditions perfect learning cannot be achieved and the asymptotic performance is still sub-optimal. Moreover, the literature argues that imperfect learning is a distinctive effect of information cascades, which limit the amount of new evidence that is included in the learning process [103]. Therefore, depending on the network topology and private signal structure there are two exclusive possibilities: either the social learning achieves perfect asymptotic learning in the absence of information cascades, or there are information cascades that limit the learning process and hence prevent perfect learning to be achieved. Following this idea, in combination with the insights about the effect of the network topology presented in [91], [39] explores how selective adaptation of incentives and the progressive rewiring of the network connections can help to steer information cascades. In despite of this preliminary effort, the literature presents little fundamental understanding of the relationship between information cascades and the achievable inference performance that can be attained by social learning. A first attempt to clarify this relationship is presented in Section VII.

IV. Social learning as a data aggregation scheme

This section presents our interpretation of social learning as a case of distributed signal processing. For this, first Section IV-A discusses the system model and basic assumptions and then Section IV-B focus in analyzing the decision rule used by agents. Finally, Section IV-C develops a communication theoretic interpretation of social learning, settling the bases of the framework that is developed in the next sections.

A. Preliminaries and basic assumptions

Let us consider a group of N agents that are sequentially engaged in a binary decision-making process. Each agent makes one decision, being it labeled according to the place it takes within the decisions' sequence. The decision of the n-th agent, denoted as $X_n \in \{0, 1\}$, is based on two sources of information (see Figure 3): a *private signal* $S_n \in \mathcal{S}$ that corresponds to a discrete or continuous random variable that represents personal information that the n-th agent possesses, and *social information* given by the random variable $\boldsymbol{G_n} \in \mathcal{G}_n$ that corresponds to information that the agent obtains from its own social network.

All participating agents have the same observation capabilities, and therefore the signals S_n are assumed to be identically distributed. Moreover, it is assumed that the signals are affected by the environment, which is represented by the random variable W. We focus in the case of $W \in \{0, 1\}$, as this simplifies the details of our presentation. For tractability reasons we follow the existent literature in assuming that the signals S_n are conditionally independent given W, following probability measures denoted by μ_w when conditioned on the event $\{W = w\}$. It is assumed that both μ_0 and μ_1 are absolutely continuous with respect to each other [104], which means that no particular signal completely determines the state of the world. As a consequence of this assumption, the log-likelihood ratio of these two distributions is well-defined and given by the logarithm of the corresponding Radon-Nikodym derivative $\Lambda_S(s) = \log \frac{d\mu_1}{d\mu_0}(s)^*$. It is also assumed that $\mu_0 \neq \mu_1$, so that $\Lambda_S(s)$ is not trivially equal to zero.

*When S_n takes a finite 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. $p(S_n|w)$ then $\frac{d\mu_1}{d\mu_0}(s) = \frac{p(s|w=1)}{p(s|w=0)}$.

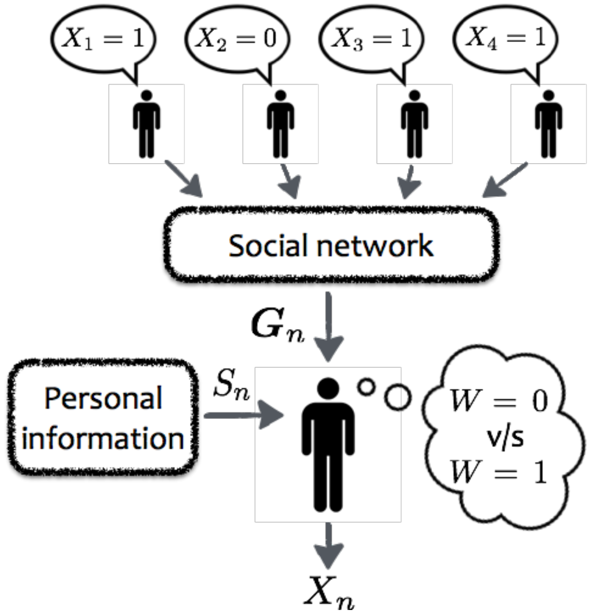


Fig. 3: A social learning problem, where an agent need to make a decision (X_n) based on personal information coming from a private signal (S_n) and social information ($\boldsymbol{G}_N$) coming from a social network.

The available social information, $\boldsymbol{G}_n$, represents what the n-th agent can observe in the social network about the decisions made by other agents, which are denoted as $\boldsymbol{X}^{n-1} = (X_1, \dots, X_{n-1})$. In effect, it is assumed that in general those decisions might not be directly observable by the agent, as they are measured through a social network that can impose observational restrictions. In general $\boldsymbol{G}_n$ can be a random variable, vector, matrix or other mathematical object. Useful examples are when $\boldsymbol{G}_n$ corresponds to:

- The k previous decisions: $\boldsymbol{G}_n = (X_{n-k+1}, \dots, X_{n-1})$.
- The average value of the all the previous decisions: $\boldsymbol{G}_n = \sum_{k=1}^{n-1} X_k/(n-1)$.
- The decisions of agents connected by an Erdos-Renyi network with parameter $q \in [0,1]$, i.e. $\boldsymbol{G}_n \in \{0,1,e\}^{n-1}$, where

$$Z_k = \begin{cases} X_k & \text{with probability } q, \\ e & \text{with probaility } 1-q. \end{cases} \tag{6}$$

Our approach is not to assume any concrete functional form for $\boldsymbol{G}_n$, but to develop a general framework with which the consequences of various properties of $\boldsymbol{G}_n$ can be explored. As a minimal requirement, we ask $\boldsymbol{G}_n$ to satisfy the following basic properties:

(i) *Causality:* it is assumed that $\boldsymbol{G}_n$ is conditionally independent given W of S_m for all $m \geq n$.
(ii) *Uniform social uncertainty:* the uncertainty present in the social media is independent of W. Therefore, $\boldsymbol{G}_n$ and W are conditionally independent given $\boldsymbol{X}^{n-1}$.

A *strategy* is a rule for generating a decision X_n based on $S_n = s$ and $\boldsymbol{G}^n = \boldsymbol{g}^n$, i.e. a collection of deterministic or random functions π_n such that $X_n = \pi_n(S_n, \boldsymbol{G}^n)$ for $n \in \{1, \dots, N\}$.

B. Decision rule

We consider rational agents that follow a *Bayesian strategy* to minimize the average cost given by $\bar{\mathcal{U}}_n\{\pi_n\} = \mathbb{E}\{u(\pi_n(S_n, \boldsymbol{G}_n), W)\}$, where $\mathbb{E}\{\cdot\}$ is the expected value operator and $u(x,w)$ is a cost function that can be engineered to match the relevance of the decision $X_n = x$ when $W = w$. For example, if $u(w,x) = 1 - \delta_{w,x}$ with $\delta_{w,x}$ the Kronecker delta, then $\bar{\mathcal{U}}_n\{\pi\} = \mathbb{P}\{W \neq \pi\}$ is the error rate of π as a predictor of W. Also, if $u(w,x) = |w-x|^2$ then the $\bar{\mathcal{U}}_n$ is the mean square error.

To find a functional description of Bayesian strategies, let us first consider the average cost of deciding $X_n = x$ given $S_n = s$ and $\boldsymbol{G}_n = \boldsymbol{g}_n$, which can be expressed as

$$\begin{aligned} \mathcal{U}_n(x|s,\boldsymbol{g}_n) &= \mathbb{E}\{u_n(x,W)|S_n = s, \boldsymbol{G}_n = \boldsymbol{g}_n\} \\ &= \sum_{w\in\{0,1\}} u(x,w)\mathbb{P}\{W = w|S_n = s_n, \boldsymbol{G}_n = \boldsymbol{g}_n\}. \end{aligned}$$

Hence, the corresponding Bayesian strategy is given by $\pi_n^{\mathsf{b}}(s,\boldsymbol{g}_n) = \text{argmin}_{x\in\mathcal{X}} \mathcal{U}_n(x|s,\boldsymbol{g}_n)$. Note that the average cost after adopting the policy π_n can then be written as

$$\bar{\mathcal{U}}_n\{\pi_n\} = \mathbb{E}\{\mathbb{E}\{\mathcal{U}_n(\pi_n(s,\boldsymbol{g})|s,\boldsymbol{g})|S_n = s, \boldsymbol{G}_n = \boldsymbol{g}_n\}\} \ ,$$

clarifying that $\bar{\mathcal{U}}_n\{\pi_n^{\mathsf{b}}\} \leq \bar{\mathcal{U}}_n\{\pi_n\}$ for any other strategy π_n.

The Bayesian strategy for the case of binary decisions can be determined by comparing $\mathcal{U}_n(0|s,\boldsymbol{g}^{n-1})$ and $\mathcal{U}_n(1|s,\boldsymbol{g}^{n-1})$, which are the relative costs associated with $X_n = 0$ and $X_n = 1$, respectively. This leads to an equivalent condition given by [89]:

$$\frac{\mathbb{P}\{W=1|S_n,\boldsymbol{G}_n\}}{\mathbb{P}\{W=0|S_n,\boldsymbol{G}_n\}} \underset{X_n=1}{\overset{X_n=0}{\lessgtr}} \frac{u(0,0)-u(0,1)}{u(1,1)-u(1,0)} \ . \tag{7}$$

Moreover, due to the causality property of $\boldsymbol{G}_n$ (c.f. Section IV-A), S_n and $\boldsymbol{G}_n$ are conditionally independent given $W = w$. Therefore, using the Bayes rule on $\mathbb{P}\{W=1|S_n,\boldsymbol{G}_n\}$ and $\mathbb{P}\{W=0|S_n,\boldsymbol{G}_n\}$, a direct calculation shows that (7) can be re-written as

$$\Lambda_S(S_n) + \Lambda_{\boldsymbol{G}_n}(\boldsymbol{G}_n) \underset{X_n=1}{\overset{X_n=0}{\lessgtr}} \nu + \eta \ , \tag{8}$$

where $\nu = \log\frac{[u(0,0)-u(0,1)]}{[u(1,1)-u(1,0)]}$, $\eta = \log\frac{\mathbb{P}\{W=0\}}{\mathbb{P}\{W=1\}}$ and $\Lambda_{\boldsymbol{G}_n}(\boldsymbol{G}_n)$ is the log-likelyhood ratio of $\boldsymbol{G}_n$. This condition supplies a simple data fusion rule for combining the information provided by S_n and $\boldsymbol{G}_n$.

C. Communication theoretic interpretation

Without loss of generality, for non-constant cost functions and adequate decision's labeling one can make the event $\{X_n = W\}$ less costly that $\{X_n \neq W\}$, or equivalently $u(1,1) \leq u(1,0)$ and $u(0,0) \leq u(0,1)$. Therefore, the Bayesian strategy is to choose X_n as similar to W as possible, according to the available state of knowledge provided by S_n and $\boldsymbol{G}_n$. Hence, decisions X_n can be considered to be noisy estimations of W in a communication theoretic signal space.

To formalize the above intuition, one can represent the decision process of each agent as data transmission over a noisy channel (for a summary of correspondences please refer to Table I). As a matter of fact, our scenario is equivalent to a single data source W that is measured over multiple noisy channels, generating the signals S_k for $k = 1, \dots, n$ that are processed in order to generate X_n (c.f. Figure 4). The decoder of the n-th node, hence, receives as input the signal S_n and uses $\boldsymbol{G}_n$ as side information for optimizing the decoding process. The social information $\boldsymbol{G}_n$ corresponds to a lossy compression of the information provided by the signals $S_1, \dots, S_{n-1}$ that is expressed in the vector of previous decisions $\boldsymbol{X}^{n-1}$, representing a bandwidth constrain for the communication between the agents.

TABLE I: Table of Correspondances

Communication theory	Social learning
Node	Social agent
Data source	"State of the world"
Noisy measurement	Private information
Communication range	Social neighbourhood
Local processing	Agent's decision
Bandwidth constraints	Social information

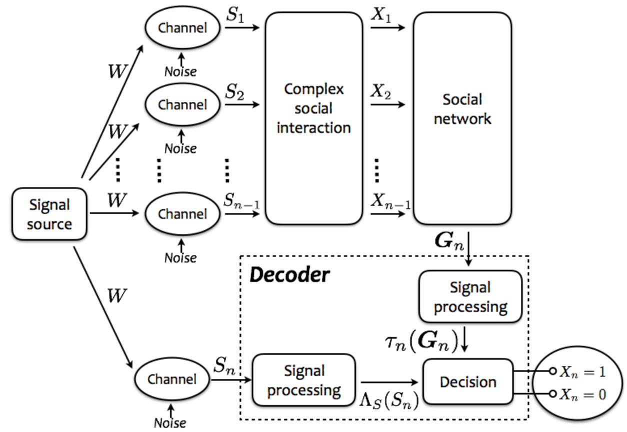


Fig. 4: Diagram that shows how social learning can interpreted as a case of distributed signal processing. The depicted decoder implements the rule given by (9), which combines the information provided by the private signal of the n-th agent (S_n) and the evidence that comes form the social network ($\boldsymbol{G}_n$). The latter can be considered as additional side-information that helps to increase the accuracy of the inference.

To further explore this perspective, let us re-formulate (8) as

$$\Lambda_S(S_n) \underset{X_n=1}{\overset{X_n=0}{\lessgtr}} \tau_n(\boldsymbol{G}_n) \ , \tag{9}$$

where $\tau_n(\boldsymbol{G}_n) = \nu + \eta - \Lambda_{\boldsymbol{G}_n}(\boldsymbol{G}_n)$ is the decision threshold. Therefore, the agent's decoder can be modeled as two independent signal processing modules that feed a decision module (see Figure 4). The first signal processing module receives as input the signal S_n —which can be a number, vector, matrix or any other mathematical object— and outputs $\Lambda_S(S_n)$, which is a real number that serves as sufficient statistic for the decision process. In this sense, this signal processing module plays a similar role to the one of matched filtering in a digital communication system [105]. The second signal processing module takes as input $\boldsymbol{G}_n$ and outputs $\tau_n(G_n)$, which corresponds to side information that is processed in order to optimize the decision threshold.

Finally, a decision module classifies the decision signal $\Lambda_S(S_n)$ based on a Vonoroi tessellation, which divides $\mathbb{R}$ in two semi-open intervals given by

$$\mathcal{K}_n^0 = (-\infty, \tau_n(\boldsymbol{G}_n)), \quad \mathcal{K}_n^1 = [\tau_n(\boldsymbol{G}_n), \infty) \ . \tag{10}$$

Therefore, the output of the decision module is provided by

$$\pi_{\mathrm{b}}^n(S_n, \boldsymbol{X}^{n-1}) = \begin{cases} 1 & \text{if } \Lambda_S(S_n) \in \mathcal{K}_n^1, \\ 0 & \text{if } \Lambda_S(S_n) \in \mathcal{K}_n^0. \end{cases} \tag{11}$$

Note that the decision module is equivalent to the last stage of a demodulator module in digital communication receivers [105], with the particular feature that the tessellation is determined by the side information provided by $\Lambda_{\boldsymbol{G}_n}(\boldsymbol{G}_n)$. This feature and subsequent consequences are analyzed in the next sections.

V. Cascading behaviour

This section presents our main contribution in the analysis of information cascades. For this purpose, first Section V-A presents a novel statistical definition of information cascades that distinguishes between local and global cascades, being valid for Bayesian and non-Bayesian social learning. Then, Section V-B characterizes the conditions that trigger local information cascades. Global information cascades are then analyzed for the case of perfect social information in Section V-D. The results are extended for non-idea social information in Section V-C. Finally, Section V-E presents a communication-theoretic interpretation of the main results of previous sections.

In the following, $X_n = \pi_n(S_n, G_n)$ corresponds to Bayesian strategies unless otherwise stated. Also, $\mathbb{P}_w\{X|Y\} = \mathbb{P}\{X|Y, W = w\}$ is used as a short-hand notation.

A. *Definitions*

Following [54], we understand as *information cascade* the phenomenon where the behaviour of a small part of the social network can trigger a herd behaviour, forcing agents to ignore their personal knowledge and act according to the social pressure. In general the decision $X_n = \pi_n(S_n, \boldsymbol{G}_n)$ depends directly on S_n and $\boldsymbol{G}_n$; however, a local cascade takes place when the interdependency between X_n and S_n is broken due to a dominant influence of $\boldsymbol{G}_n$. This intuition is formalized in the next definition.

Definition 1. *The social information $\boldsymbol{g}_n \in \mathcal{G}_n$ pushes the n-th agent into a* local information cascade *if $X_n = \pi_n(S_n, \boldsymbol{g}_n)$ is statistically independent of S_n.*

Note that, for the particular case of Bayesian strategies then $\pi_n^b : \mathcal{S} \times \mathcal{G}_n \to \{0, 1\}$ is a deterministic function, and hence the above definition states that $\boldsymbol{g}_n$ causes a local cascade if and only if $\pi(s, \boldsymbol{g}_n)$ is constant for all $s \in \mathcal{S}$. Therefore, the above definition generalises the one provided in [39], being also valid for non-Bayesian strategies.

As a next step, global information cascades are defined. Intuitively, after a particular agent experiences $\boldsymbol{G}_n = \boldsymbol{g}_n$ then a global cascade takes place if all subsequent agents also fall in local information cascades almost surely.

Definition 2. *A social network experiences a* global information cascade *if there exist a $\boldsymbol{g}_n \in \mathcal{G}_n$ such that the variables S_k and X_k are statistically independent for all $k \geq n$ when conditioned to the event $\{\boldsymbol{G}_n = \boldsymbol{g}_n\}$.*

An interesting question is when a local information cascade triggers a global one. The following sub-sections explore this issue for the case of Bayesian statistics.

B. Decision statistics and local information cascades

As a next step, we aim to analyse the behavior of local information cascades as specified by Definition 1. For studying the statistics of X_n when rational agents use Bayesian statistics, first note that all $\Lambda_S(S_n)$ are identically distributed. Therefore, for a given $\boldsymbol{G}_n = \boldsymbol{g}_n$ and $W = w$, X_n is Bernoulli distributed with parameter given by

$$\begin{aligned}\mathbb{P}_w\{X_n = 0 \,|\, \boldsymbol{G}_n = \boldsymbol{g}_n\} \\ &= \int_{\mathcal{S}} \mathbb{P}_w\left\{X_n = 0 | \boldsymbol{G}_n = \boldsymbol{g}_n, S_n = s\right\} \mathrm{d}\mu_w(s) \\ &= \int_{\mathcal{S}} \mathbb{1}\left\{\pi_n^{\mathrm{b}}(s, \boldsymbol{g}_n) = 0\right\} \mathrm{d}\mu_w(s) \\ &= \mathbb{P}_w\left\{\Lambda_S(S_n) < \tau_n(\boldsymbol{g}_n)\right\} \\ &= F_w^{\Lambda}\big(\tau_n(\boldsymbol{g}_n)\big) \ ,\end{aligned}$$

where $F_w^{\Lambda}(\cdot)$ is the c.d.f. of the variable $\Lambda_S(S_n)$ conditioned to $W = w$, whose properties are discussed in Appendix A. Note that the first equality is a consequence of the conditional independency between S_n and $\boldsymbol{G}_n$ given $W = w$, the second is due to (11) and the third to (9). The above results allows to prove an useful lemma.

Lemma 1. *$X_n - \tau_n - \boldsymbol{G}_n$ form a Markov Chain (i.e. $\tau_n(\boldsymbol{G}_n)$ is a sufficient statistic for generating X_n).*

Proof: From (V-B) one can see that $\mathbb{P}_w\left\{X_n | \tau_n, \boldsymbol{G}_n\right\}$ do not depend on $\boldsymbol{G}_n$, and therefore the conditional independency of X_n and $\boldsymbol{G}_n$ given τ_n is clear. ■

Let us introduce the notation $U_s = \operatorname{ess\,sup}_{s\in\mathcal{S}} \Lambda_S(S_n = s)$ and $L_s = \operatorname{ess\,inf}_{s\in\mathcal{S}} \Lambda_S(S_n = s)$ for the essential supermum and infimum of $\Lambda_S(S_n)^{\dagger}$. If any of these quantities diverge, then there exist signals that provide overwhelming evidence in favour of one of the hypothesis. If both are finite, the agents are said to have *bounded beliefs*. Using these definitions, we proceed to characterize local information cascades.

Proposition 1. *The social information $\boldsymbol{g}_n \in \mathcal{G}_n$ triggers a local information cascade if and only if $\tau_n(\boldsymbol{g}_n) \notin [L_s, U_s]$.*

Proof: From (V-B) it can be seen that if $\tau_n < L_s$ then is $F_0^{\Lambda}(\tau_n) = F_1^{\Lambda}(\tau_n) = 0$, while if $\tau_n > U_s$ then $F_0^{\Lambda}(\tau_n) = F_1^{\Lambda}(\tau_n) = 1$. Therefore, if $\tau_n(\boldsymbol{g}_n) \notin [L_s, U_s]$ then it determines X_n almost surely, making X_n and S_n independent.

On the other hand, if $L_s < \tau_n(\boldsymbol{g}_n) < U_s$ then (V-B) and the definition of U_s and L_s allows to conclude that $0 < \mathbb{P}_w\left\{X_n = 0 | \boldsymbol{X}^{n-1}\right\} < 1$ for any $w \in \{0,1\}$. This implies that the sets $\mathcal{S}^0(\tau) = \{s \in \mathcal{S} | \Lambda_S(s) < \tau\}$ and $\mathcal{S}^1(\tau) = \mathcal{S} - \mathcal{S}^0$ have positive probability under both μ_0 and μ_1, which in turn implies the existence of interdependency between X_n and S_n in this case. ■

In this way, we found a simple characterization of the conditions that trigger local information cascades. Intuitively, Proposition 1 states that if the social information provides more evidence than any possible signal, then a local cascade is triggered almost surely. Some consequences of this result are explored in the next section.

C. Global cascades under perfect social information

In this section we explore the conditions that trigger global information cascades (c.f. Definition 2 given in Section V-A) in the special case where $\boldsymbol{G}_n = \boldsymbol{X}^{n-1}$, i.e. when each agent has perfect access to all the previous decisions. As a first observation, note that a direct calculation shows that

$$\begin{aligned}\tau_{n+1}(\boldsymbol{X}^n) - \tau_n(\boldsymbol{X}^{n-1}) =& \Lambda_{\boldsymbol{X}^{n-1}}(\boldsymbol{X}^{n-1}) - \Lambda_{\boldsymbol{X}^n}(\boldsymbol{X}^n) \\ =& -\Lambda_{X_n|\boldsymbol{X}^{n-1}}(X_n|\boldsymbol{X}^{n-1}) \ , \quad (12)\end{aligned}$$

where the conditional log-likelihood is given by

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

This shows that τ_{n+1} decreases if X_n provides additional evidence about $W = 1$ over $W = 0$ with respect to the previous decisions, and increased if the opposite happens. A direct calculation using (V-B) shows that

$$\Lambda_{X_n|\boldsymbol{X}^{n-1}}(x_n|\boldsymbol{x}^{n-1}) = \lambda\big(x_n, \tau_n(\boldsymbol{x}^{n-1})\big) \ . \quad (13)$$

where the function $\lambda(\cdot,\cdot)$ is defined as

$$\lambda(x,\tau) = x \log \frac{F_1^{\Lambda}(\tau)}{F_0^{\Lambda}(\tau)} + (1-x) \log \frac{1 - F_1^{\Lambda}(\tau)}{1 - F_0^{\Lambda}(\tau)} \ . \quad (14)$$

This shows that $\tau_n(\boldsymbol{X}^{n-1})$ is a sufficient statistic of $\boldsymbol{X}^{n-1}$ for predicting $\tau_{n+1} - \tau_n$. Finally, using these results one can find that

$$\tau_n(\boldsymbol{X}^{n-1}) = \eta + \nu - \sum_{k=1}^{n-1} \lambda\big(X_k, \tau_k(\boldsymbol{X}^{k-1})\big) \quad (15)$$

for $n \geq 2$, while for $n = 1$ then $\tau_1 = \nu + \eta$.

From (15) it is tempting to interpret τ_n as a random walk over the decision space(for an introduction about Random walks, see [107]). Although this is true for some particular cases , this does not hold always as the steps $\tau_{n+1} - \tau_n = -\Lambda_{X_n|\boldsymbol{X}^{n-1}}$ are in general not identically distributed. However, although the process of decisions $X_1, X_2, \dots$ in general have quite complex non-Markovian statistics, the process $\tau_1, \tau_2, \dots$ posses some useful properties that are explored in the next Lemma.

Proposition 2. *The process $\tau_1, \tau_2, \dots$ is Markovian and Super- or sub- Martingale for $W = 0$ and $W = 1$, respectively.*

[†] The essential supremum is the smallest upper bound that holds almost surely, being the natural measure-theoretic extension of the supremum [106].

Proof: See Appendix B. ■

The main result of this sub-section is to state that the process τ_n get trapped in some specific areas of $\mathbb{R}$, and that these events correspond to global information cascades. Theorem 1 states that as soon as τ_n goes beyond $[L_s, U_s]$ then it gets stucked, this condition being necessary and sufficient for an information cascade to be triggered. As that region of the decision signal space characterizes local information cascades (c.f. Proposition 1), is clear that this implies that in this scenario every local information cascade triggers a global one.

Theorem 1. *Let us assume that the private signals provide bounded beliefs, i.e. both U_s and L_s are finite (c.f. Section V-B). Then, for a given $\boldsymbol{x}^{n-1} \in \{0,1\}^{n-1}$, the three following conditions are equivalent:*

(i) $\boldsymbol{X}^{n-1} = \boldsymbol{x}^{n-1}$ *triggers a local information cascade in agent n.*

(ii) $\Lambda_{X_n|\boldsymbol{X}^{n-1}}(X_n|\boldsymbol{x}^{n-1}) = 0$ *almost surely.*

(iii) $\boldsymbol{X}^{n-1} = \boldsymbol{x}^{n-1}$ *causes a global information cascade.*

Proof: See Appendix B. ■

Basically, the above theorem states that if the private signals have bounded beliefs then each local information cascade triggers a global one. Therefore, the simple graphical characterization provided by Proposition 1 for local cascades can be used to analyse global cascades as well (c.f. the analysis provided in Section V-E).

D. Global cascades under non-ideal social information

This section extends the results for global information cascades obtained in the previous section to more general scenarios. For this, let us first define the *distortion coefficients* given by

$$\alpha_n(\boldsymbol{g}_n|\boldsymbol{x}^{n-1}) = \mathbb{P}\left\{\boldsymbol{G}_n = \boldsymbol{g}_n|\boldsymbol{X}^{n-1} = \boldsymbol{x}^{n-1}\right\} \quad . \tag{16}$$

Note that $\alpha_n(\boldsymbol{g}_n|\boldsymbol{x}^{n-1})$ corresponds to the likelihood for the n-th agent to experience $\boldsymbol{G}_n = \boldsymbol{g}_n$ when the previous decisions are $\boldsymbol{X}^{n-1} = \boldsymbol{x}^{n-1}$. In the following, we use the notation $\tau_n^{\text{full}}(\boldsymbol{X}^{n-1}) = \eta + \nu - \Lambda_{\boldsymbol{X}^{n-1}}(\boldsymbol{X}^{n-1})$ to refer to the decision threshold of the case of perfect social information, studied in Section IV, distinguishing it with respect to the actual decision threshold $\tau_n(\boldsymbol{G}_n)$ related to a state of limited knowledge as introduced in Section IV-C. We also introduce the following further property, which is crucial for the rest of the section.

Definition 3. *The social information $\boldsymbol{G}_n$ is said to has consistent distortion if, for all $\boldsymbol{g}_n \in \mathcal{G}_n$, one of the following possibilities hold: either all the decision vectors $\boldsymbol{x}^{n-1} \in \{0,1\}^{n-1}$ such that $\alpha_n(\boldsymbol{g}_n|\boldsymbol{x}^{n-1}) > 0$ satisfy $\tau_n^{full}(\boldsymbol{x}^n) \notin [L_s, U_s]$, or all decision vectors such that $\alpha_n(g|\boldsymbol{x}^{n-1}) > 0$ satisfy $\tau_n^{full}(\boldsymbol{x}^n) \in [L_s, U_s]$.*

Before presenting the main results of the section, we need to state the following useful lemma.

Lemma 2. *If $a_1, \dots, a_n$ and $b_1, \dots, b_n$ are collections of non-negative numbers, then*

$$\min\left\{\frac{a_1}{b_1}, \dots, \frac{a_n}{b_n}\right\} \le \frac{\sum_{j=1}^n a_j}{\sum_{j=1}^n b_j} \le \max\left\{\frac{a_1}{b_1}, \dots, \frac{a_n}{b_n}\right\}.$$

Proof: See appendix B. ■

The following Theorem extends Theorem 1 for the case of non-ideal social information, providing a sufficient condition under which local information cascades always trigger a global one. In summary, this proves that in these scenarios the decision threshold τ_n evolves until the first time it reaches outside of the interval $[L_s, U_s]$. If $\tau_n \notin [L_s, U_s]$ then $\tau_m = \tau_n$ for all $m > n$, this being an unequivocal signal of the beginning of an global information cascade.

Theorem 2. *If $\boldsymbol{G}_n$ possess global consistent distortion, then any local information cascades trigger a global information cascade.*

Proof: See appendix B. ■

For providing further intuition and further leverage our results from Section V-C, we explore the relationship between the condition that trigger global information cascades under perfect social information and the ones that trigger cascades in the non-ideal case.

Proposition 3. *If the process $\boldsymbol{G}_n$ has consistent distortion, then each $\boldsymbol{g} \in \mathcal{G}_n$ for which exists at least one $\boldsymbol{x}^n \in \{0,1\}^{n-1}$ with $\alpha_n(\boldsymbol{g}|\boldsymbol{x}^{n-1}) > 0$ and $\tau_n^{full}(\boldsymbol{x}^{n-1}) \notin [L_s, U_s]$ triggers an global information cascade.*

Proof: Let us focus in the case that $\tau_n^{\text{full}}(\boldsymbol{x}^n) > U_s$, as the proof for the case $\tau_n^{\text{full}}(\boldsymbol{x}^n) < L_s$ is analogous. Thanks to Lemma 2 and Theorem 2, it is sufficient to prove that the above condition guarantees that $\tau_n(\boldsymbol{G}_n) > U_s$. A direct computation shows that

$$\begin{aligned}\Lambda_{\boldsymbol{G}_n}(\boldsymbol{g}) &= \log\frac{\mathbb{P}_1\left\{\boldsymbol{G}_n = \boldsymbol{g}\right\}}{\mathbb{P}_0\left\{\boldsymbol{G}_n = \boldsymbol{g}\right\}} \\ &= \log\frac{\sum_{\boldsymbol{x}^n \in \{0,1\}^n} \alpha_n(\boldsymbol{g}|\boldsymbol{x}^{n-1})\mathbb{P}_1\left\{\boldsymbol{X}^n = \boldsymbol{x}^n\right\}}{\sum_{\boldsymbol{x}^n \in \{0,1\}^n} \alpha_n(\boldsymbol{g}|\boldsymbol{x}^{n-1})\mathbb{P}_0\left\{\boldsymbol{X}^n = \boldsymbol{x}^n\right\}}\end{aligned} \tag{17}$$

Then, by applying the previous lemma to the argument of the logarithm, one can show that

$$\min_{\boldsymbol{x}^n \in \mathcal{A}_n}\left\{\Lambda_{\boldsymbol{X_n}}(\boldsymbol{x}^n)\right\} \le \Lambda_{\boldsymbol{G}_n}(\boldsymbol{g}) \le \max_{\boldsymbol{x}^n \in \mathcal{A}_n}\left\{\Lambda_{\boldsymbol{X_n}}(\boldsymbol{x}^n)\right\} ,$$

where $\mathcal{A}_n = \left\{\boldsymbol{x}^n \in \{0,1\}^n \middle| \alpha(\boldsymbol{g}|\boldsymbol{x}^n) > 0\right\}$. This condition is equivalent to

$$\min_{\boldsymbol{x}^n \in \mathcal{A}_n}\left\{\tau_n^{\text{full}}(\boldsymbol{x}^n)\right\} \le \tau_n(\boldsymbol{G}_n) \le \max_{\boldsymbol{x}^n \in \mathcal{A}_n}\left\{\tau_n^{\text{full}}(\boldsymbol{x}^n)\right\} . \tag{18}$$

From this last condition, combined with the global consistency, guarantees that if one $\boldsymbol{x}^n$ is such that $\alpha_n(g_n|\boldsymbol{x}^{n-1}) > 0$ and $\tau_n^{\text{full}}(\boldsymbol{x}^{n-1}) > U_s$, then $\tau_n(\boldsymbol{G}_n) > U_s$ as well. ■

E. Information cascades from a communication-theoretic perspective

While previous sections study the conditions that trigger local and global information cascades, here we aim to relate the achieved results with the discussion presented in Section IV-C. In this way, by following a communication-theoretic perspective, it is possible to see the evolution of τ_n as a refinement in the process of signal decoding and a information cascade as a halt on this refinement. In effect,

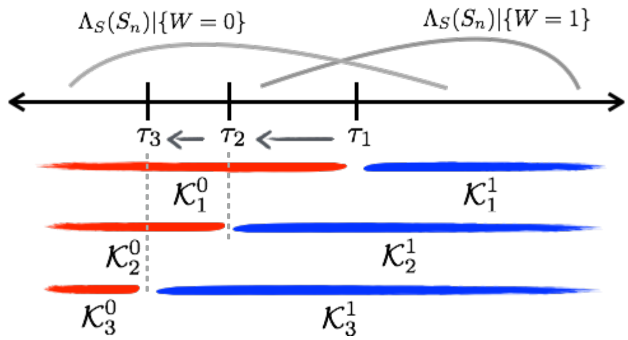


Fig. 5: The decision module compares the decision signal $\Lambda_S(S_n)$ against a threshold $\tau_n(G_n)$, which evolves with the social information.

the social information makes τ_n to grow progressively, which corresponds to a stronger side information that favours one of the two hypothesis (see Figure 5). Correspondingly, an information cascade is equivalent to side information that is so persuasive that completely determines the output of the decoder, disregarding the actual private signal realization.

It is important to notice that when τ_n stops evolving due to an information cascade, then the subsequent realizations of private signals or social information do not influence the decision process of future agents any longer, implying that the learning process has actually stopped. The error rates of all subsequent agents after an information cascade is triggered are the same, as their decoding conditions are equivalent. This explains why information cascades prevent the achievement of a perfect inference even for asymptotically large networks.

It is interesting to note that the asymptotic performance of social learning is directly related to the size of $[L_s, U_s]$. To illustrate this fact, let us first note that a value of $\mathbb{P}\left\{W=1|S_n,\boldsymbol{X}^{n-1}\right\}$ close to 0 or 1 of represent a high certainty about the "state of the world". A direct calculation using Bayes rule shows that

$$\begin{aligned}\mathbb{P}\left\{W=1|S_n,\boldsymbol{G}_n\right\} =&\phi(\Lambda_{\boldsymbol{G}_n}(\boldsymbol{G}_n)-\eta+\Lambda_S(S_n))\\ =&\phi(-\tau_n+\nu+\Lambda_S(S_n)) \ , \end{aligned}\tag{19}$$

where $\phi(x) = 1/(1+e^{-x})$ is the well-known sigmoid function, and the second equality comes from the fact that $\nu-\tau_n = \eta-\Lambda_{\boldsymbol{G}_n}(\boldsymbol{G}_n)$. Therefore, it is clear that a large value of $|\tau_n|$ corresponds to a state of high certainty about W. Moreover, according to Theorem 2 and Proposition 1, asymptotically all values of τ_n are either larger than U_s or smaller than L_s. Therefore, if those values are large then this guarantees a smal asymptotic error rate. This intuition is further explored in Section VII.

VI. Social learning with binary private signals

For illustrating the results of the previous section, we present an application of our framework to study social systems where the private signals are binary. Please note that that such binary systems are popular in the literature, being extensively discussed, e.g. [38], and experimentally validated [70]. Further applications to systems with other private signal distributions can be found in Appendix C.

In the following, first Section VI-A presents results valid for the general binary case, and then Section VI-B focuses in the case of private signals with structure similar to a binary symmetric channel.

A. General results

Let us consider focus our analysis on the general case of social systems where the agents have access to binary signals, i.e. $\mathcal{S}=\{0,1\}$. Let us denote the false alarm and miss-detection rates by $\epsilon_w = \mathbb{P}_w\left\{S_n=1-w\right\}$ for $w\in\{0,1\}$, and assume without loss of generality that $\max\{\epsilon_0,\epsilon_1\}\leq 1/2$.

A direct computation of Λ_S gives that

$$\Lambda_S(S_n) = S_n\log\frac{1-\epsilon_1}{\epsilon_0}+(1-S_n)\log\frac{\epsilon_1}{1-\epsilon_0} \ . \tag{20}$$

Note that $U_s = \Lambda_S(1) > 0 > \Lambda_S(0) = L_s$, which is consequence of the fact that $1-\epsilon_1\geq\epsilon_0$. Correspondingly, the c.d.f. of $\Lambda_S(S_n)$ for given $W=w$ is a step function given by

$$F_w^\Lambda(\tau)=\begin{cases}0 & \text{if } \tau<L_s,\\ \mathbb{P}_w\left\{S_n=0\right\} & \text{if } L_s\leq\tau<U_s,\\ 1 & \text{if } U_s\leq\tau.\end{cases}\tag{21}$$

As an inmediate observation, Proposition 1 guarantees that if $\tau_1=\eta+\nu<L_s$, then $X_n=1$ for all n, triggering a trivial global information cascade. Equivalently, if $\tau_1>U_s$ then $X_n=0$ for all n. Therefore, for avoiding trivial scenarios it is always assumed that $\tau_1\in[L_s,U_s]$.

Let us compute the log-likelihoods of the social decisions. A simple application of (11) and the fact that $L_s\leq\tau_1\leq U_s$, it can be concluded that $X_1=S_1$ almost surely. Therefore, by comparing (13) and (20), one obtains that

$$\Lambda_{X_1}(X_1)=\Lambda_S(S_1) \ , \tag{22}$$

where the equality holds almost surely. An analogous analysis reveals that

$$\Lambda_{X_n|\boldsymbol{X}^{n-1}}(X_n|\boldsymbol{X}^{n-1})=\begin{cases}\Lambda_S(S_n) & \text{if } \tau_n(\boldsymbol{x}^{n-1})\in[L_s,U_s],\\ 0 & \text{in other case.}\end{cases}$$

Therefore, using (13) and (15) is clear that

$$\tau_n(\boldsymbol{X}^{n-1})=\nu+\eta-\sum_{k=1}^{n_c-1}\Lambda_S(S_k) \tag{23}$$

for all $n\geq n_c$, where n_c is the first agent who experiences a local information cascade. Therefore, the evolution of τ_n can be described as follows: it starts at $\nu+\eta$ and evolves taking upward steps of size $-L_s$ or downward steps of size U_s, stoping as soon as it reaches beyond $[L_s,U_s]$.

The above derivation illustrates the result stated by Theorem 1, showing how τ_n stops evolving after reaching beyond the range of values of the private evidence. Moreover, it also shows how the global information cascade corresponds to the fact that the social network stops processing new data after the first agent suffers a local information cascade.

B. Binary symmetric channel

Let us now specify our analisys by considering a case where the false alarm rate is equal to the miss-detection rates, which correspond to a binary symmetric channel with cross-over probability $\epsilon_0 = \epsilon_1 := \epsilon$. The symmetry of this scenario allows to relate the evolution of the decision threshold with a random walk and find close-form expressions for the cascading probabilities.

As a first step, by realizing that in this case $\Lambda_S(1) = -\Lambda(0) = \log \frac{1-\epsilon}{\epsilon}$, it is clear that (23) can be re-written as

$$\tau_n = \nu+\eta-k_0\Lambda(0)-k_1\Lambda(1) = \nu+\eta+(k_0-k_1)\log\frac{1-\epsilon}{\epsilon} \ ,$$

where k_0 and k_1 are the number of decisions equal to 0 or 1 made by the agents before n_c. Therefore, (24) shows that in this scenario τ_n evolves following a random walk over the set $\{\nu+\eta+k\log\frac{1-\epsilon}{\epsilon}; k \in \mathbb{Z}\}$, which gets trapped as soon as $|\tau_n| > \log\frac{1-\epsilon}{\epsilon}$ (c.f. Figure 6). A direct inspection shows that k might only take four possible values: $-1, 0, 1$ and 2 if $\nu+\eta < 0$ or $-2, -1, 0$ and 1 if $\nu+\eta > 0$. An example of a realization of one decision sequence is illustrated in Figure 7, where one can observe an information cascade being triggered as soon as the decision threshold leaves the area demarcated by red lines.

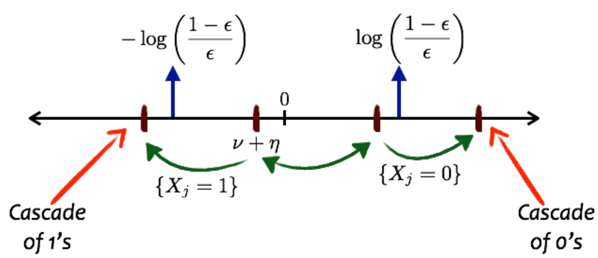


Fig. 6: The evolution of the decision threhold for fully connected social networks with binary private signals can be characterized as a random walk in the decision signal space. Every decision $X_n = 1$ or $X_n = 0$ causes steps to the left or right, respectively. When random walk moves away of the interval defined by the private believes of each agent, then the random walk stops and an information cascade is triggered.

With this characterization, the probability of information cascade of 1's or 0's, denoted as $\mathbb{P}\{C_1\}$ and $\mathbb{P}\{C_0\}$, can be found using the theory of *Random Walks and Ruin Problems* [107]. For this, let us consider a Random Walk over the numbers $0, 1, 2$ and 3, which starts either at 1 or 2 and stops when it first hits 0 or 3. Without loss of generality (due to the symmetry of the scenario), we assume that each upward step occurs with probability ϵ while each downward step take place with probability $1-\epsilon$. Let us denote by q_z the probability of hitting 0 before 3, and p_z the probability of hitting the 3 before the 0, both when the random walk starts from position z. Then, a classic result of random walk theory (c.f. [107, pg. 345]) states that $q_z = 1 - p_z$ and

$$q_z = \frac{1-\left(\frac{\epsilon}{1-\epsilon}\right)^{3-z}}{1-\left(\frac{\epsilon}{1-\epsilon}\right)^{3}} \ , \qquad (24)$$

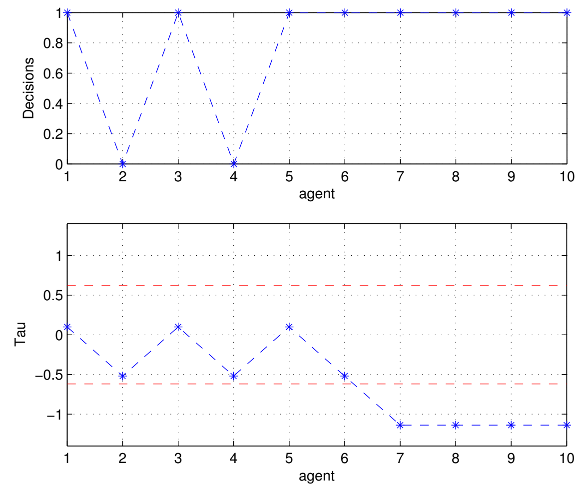


Fig. 7: *Up*- Sequence of decisions in a social system with binary symmetric signals, cross-over probability $\alpha = 0.35$ and $\nu+\eta = 0.1$. *Down*- Evolution of the decision threshold τ_n for the same system. As Theorem 2 predicted, a global information cascade is triggered after the decision threshold reach beyond the limits imposed by the bounded private beliefs (marked by the red lines), which takes place after the decision of the sixth agent.

where $\lambda_\epsilon = (1-\epsilon)/\epsilon$. Therefore, q_1 can be understood as the probability of having a correct cascade (e.g. a cascade of 1's when $W = 1$) when $\nu+\eta \in [-\log\frac{1-\epsilon}{\epsilon}, 0]$, i.e. when having favorable priors (c.f. Figure 6). On the other hand, q_2 corresponds to the case where $\nu+\eta \in [0, \log\frac{1-\epsilon}{\epsilon}]$, i.e. to the rate of correct cascades when the priors are not favorable. The accuracy of these expressions for predicting the cascade rates have been verified by numerical simulations (see Figure 8).

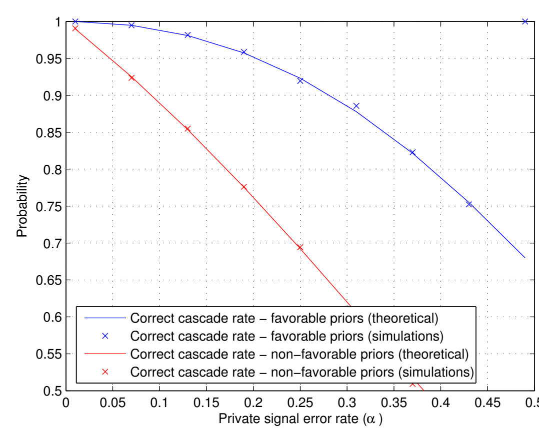


Fig. 8: Rate of the correct cascade (e.g. $\mathbb{P}\{C_1\}$ if $W = 1$) for different priors, as described in Section VI-B. Results confirm that the evolution of social learning for binary signals can be accurately described by a random walk model, and that (24) effectively predict the corresponding cascading rates.

The characterization of the threshold's evolution in terms of a random walk introduces clear insights about the social learning process. For example, the above results show how two agreed consecutive decisions are sufficient to trigger a global information cascade (c.f. Figure 6). In contrast, Appendix C shows how more complex private signals can make it more difficult to trigger global cascades.

VII. Analysis of the average cost

In this section we study the evolution of the agent's cost, which is a natural metric to evaluate the performance of social learning. First, Section VII-A provide general close-form formulas that enable efficient numerical evaluations. Then, Section VII-B use these results to provide upper bounds to the asymptotic cost, which clarify the impact of information cascades over the learning process.

A. Computing the average payoff

A direct calculation shows that the average payoff of the n-th agent under a Bayesian strategy, as defined in Section IV-B, can be expressed as

$$\begin{aligned}\bar{\mathcal{U}}_n(\pi_\text{b}) &= \sum_{\substack{w\in\{0,1\}\\ g\in\mathcal{G}_n}} \mathbb{P}_w\left\{\boldsymbol{G}_n=\boldsymbol{g}_n\right\}\mathbb{P}\left\{W=w\right\} \\ &\qquad\times\int_{\mathcal{S}} u_n(w,\pi_\text{b}\{s,\tau_n(\boldsymbol{g}_n)\})\mathrm{d}\mu_w(s) \qquad (25)\\ &= \sum_{\substack{w\in\{0,1\}\\ g\in\mathcal{G}_n}} \mathbb{P}_w\left\{\boldsymbol{G}_n=\boldsymbol{g}_n\right\}\mathbb{P}\left\{W=w\right\} \\ &\quad\times\Big(u_{w,0}F_w^\Lambda(\tau_n(\boldsymbol{g}_n))+u_{w,1}\left[1-F_w^\Lambda(\tau_n(\boldsymbol{g}_n))\right]\Big)\ , \qquad (26)\end{aligned}$$

where the first equality is consequence of the conditionally independency of S_n and $\boldsymbol{G}_n$ given $W=w$, and the second equality of (V-B) and the fact that π_b is a binary variable. Note that a direct evaluation of (26) is possible using the algorithm to compute $\mathbb{P}_w\left\{\boldsymbol{G}_n=\boldsymbol{g}\right\}$ given by Appendix E.

With the above results it is possible to perform numerical evaluations of the exact performance of social learning in diverse scenarios. This is illustrated in Section VIII for the case of different private signal statistics.

B. Asymptotic performance analysis

In this section we derive lower bounds for the asymptotic payoff $\lim_{n\to\infty}\bar{\mathcal{U}}_n(\pi_\text{b}) := \bar{\mathcal{U}}_\infty(\pi_\text{b})$. This quantity is a key performance metric for large social networks. Although the exact value can be estimated by numerical evaluations using the formulas presented in Section VII-A, in most cases the complexity of the computations that are required to reach the point of convergence. Therefore, the simple upper bounds presented in this Section provide valuable insights of this important metric.

For this, let us first present Proposition 4, which considers the statistics of τ_n. Note that, because τ_n is a deterministic function of $\boldsymbol{G}_n$, is possible to express its statistics as

$$\mathbb{P}_w\left\{\tau_n=t\right\} = \sum_{\substack{g\in\mathcal{G}_n\\ \tau_n(g)=t}} \mathbb{P}_w\left\{\boldsymbol{G}_n=\boldsymbol{g}\right\}. \qquad (27)$$

Proposition 4. *For arbitrary social systems with partial social information, it holds that*

$$\bar{\mathcal{U}}_\infty(\pi_b) = \lim_{n\to\infty}\mathbb{E}\left\{\Phi(\tau_n)\right\}\ , \qquad (28)$$

where $\Phi(\cdot)$ is defined as

$$\begin{aligned}\Phi(t) =& \Big(u_{0,0}F_0^\Lambda(t)+u_{0,1}\left[1-F_0^\Lambda(t)\right]\Big)\phi(t-\nu)\\ &+\Big(u_{1,0}F_1^\Lambda(t)+u_{1,1}\left[1-F_1^\Lambda(t)\right]\Big)\phi(\nu-t).\end{aligned}$$

Proof: See Appendix D. ■

Note that $\Phi(\cdot)$ measures the contribution of each $t\in\mathcal{T}_n$ to the average costs. In particular, following the results of Sections IV, let us consider a possible cascade of 0's generated by $t_0\geq U_s$. Note that, under that condition, then $F_0^\Lambda(t_0)=F_1^\Lambda(t_0)=1$. Then, using the previous proposition, the contribution of this potential cascade to the asymptotic average cost can be found to be $\Phi(t_0)=u_{1,0}-(u_{1,0}-u_{0,0})\phi(t_0-\nu)$. Interestingly, when $t_0-\nu\approx 0$ then $\Phi(t_0)\approx(u_{1,0}+u_{0,0})/2$, which is the payoff of a random guess. On the other hand, due to $u_{0,0}\leq u_{1,0}$ (c.f. Section IV-C), when $t_0\to\infty$ then $\Phi(t_0)$ decreases monotonically towards a limit given by $\Phi(t_0)\to u_{0,0}$, being this the payoff of a perfect prediction. An equivalent analysis can be done for the case of cascades of 1's where $t\leq L_s$, showing that for that case $\Phi(t)=u_{0,1}+(u_{1,1}-u_{0,1})\phi(\nu-t)$ and hence where a more negative value of t reduces the impact of cascades. These insights motivate us in calling the quantity $|t-\nu|$ the "accuracy" or predictive power of a cascade. Then, the brief analysis provides the following important insight: *cascades with a higher accuracy have a smaller impact over the asymptotic cost.*

As a next step, we use these insights to develop an upper bound to the average cost. The main idea is that U_s and L_s impose natural restrictions over the threshold values of information cascades, and hence can be used to consider the performance of a worst-case-scenario.

Theorem 3. *Let us consider a social network with consistent distortion and bounded beliefs. Moreover, assume that the network end up in a information cascade almost surely. Then, the following lower bound holds:*

$$\bar{\mathcal{U}}_\infty(\pi_b)\leq h_0(U_s-\nu)\mathbb{P}\left\{C_0\right\}+h_1(\nu-L_s)\mathbb{P}\left\{C_1\right\}\ . \qquad (29)$$

where $h_w(x):=u_{1-w,w}-(u_{1-w,w}-u_{w,w})\phi(x)$, and C_0 and C_1 denote the events of cascades of 0's or 1's, respectively.

Proof: See Appendix D. ■

Although we provide closed-form expressions for $\mathbb{P}\left\{C_1\right\}$ and $\mathbb{P}\left\{C_0\right\}$ for some special cases (c.f. Section VI-B), in general these terms are difficult to compute. Nevertheless, they can be calculated by Monte Carlo simulations. Furthermore, one can use the following simpler bounds, whose proofs are direct and left to the interested reader.

Corollary 1. *For a social network that satisfies the conditions of Theorem 3, the following upper bound holds:*

$$\bar{\mathcal{U}}_\infty(\pi_b)\leq\max\Big\{h_0(U_s-\nu),h_1(\nu-L_s)\Big\}\ . \qquad (30)$$

Finally, let us study the case of $u_{0,0}=u_{1,1}=0$ and $u_{0,1}=u_{1,0}=1$, for which

$$\begin{aligned}\bar{\mathcal{U}}_n(\pi_\text{b}) &= \mathbb{P}\left\{X_n\neq W\right\} \qquad (31)\\ &= \sum_{w\in\{0,1\}}\mathbb{P}_w\left\{X_n=1-w\right\}\mathbb{P}\left\{W=w\right\}\ . \qquad (32)\end{aligned}$$

Let us introduce the shorthand notation $P_\infty(\text{FA}) = \lim_{n\to\infty} \mathbb{P}_0\{X_n = 1\}$ and $P_\infty(\text{MD}) = \lim_{n\to\infty} \mathbb{P}_1\{W = 0\}$ for the asymptotic false alarms and miss-detection rates, respectively. The next corollary provides useful expressions for those quantities.

Corollary 2. *For a social network that satisfies the conditions of Theorem 3, then the following formulae hold:*

$$P_\infty(\mathit{FA}) \le \phi(-U_s)\mathbb{P}_0\{C_0\} + \phi(L_s)\mathbb{P}_0\{C_1\},$$
$$P_\infty(\mathit{MD}) \le \phi(-U_s)\mathbb{P}_1\{C_0\} + \phi(L_s)\mathbb{P}_1\{C_1\}.$$

Proof: From (32) is clear that

$$\bar{\mathcal{U}}_\infty(\pi_\text{b}) = P_\infty(\text{FA})\mathbb{P}\{W=0\} + P_\infty(\text{MD})\mathbb{P}\{W=1\}. \quad (33)$$

Now let us consider the case of $u_{0,0} = u_{1,1} = 0$ and $u_{0,1} = u_{1,0} = 1$, for which $\nu = 0$ and $h_0(x) = h_1(x) = \phi(-x)$. Hence, the bound provided in (29) can be re-written as

$$\bar{\mathcal{U}}_\infty(\pi_\text{b}) \le \sum_{w\in\{0,1\}} \Big[\phi(-U_s)\mathbb{P}_w\{C_0\} + \phi(L_s)\mathbb{P}_w\{C_1\}\Big]\mathbb{P}\{W=w\}. \quad (34)$$

The corollary is proven by comparing (33) and (34) and realizing that they both hold for any choice of priors $\mathbb{P}\{W=w\}$. ∎

One of the main consequences of these results is the fact that —as long as the system has consistent distortion— one can provide a bound of the asymptotic error rate based exclusively in the extreme values of the signal log-likelihood. For illustrating this point, let us consider a system where $L_s = -U_s$. Then, as a consequence of Corollary 2, one can directly guarantee that the false alarms and error rates are smaller than $\phi(-U_s) = 1/(1+e^{U_s})$. Reciprocally, one can guarantee that the asymptotic error rate is less than p_0 if both U_s and $-L_s$ are both larger than $\log\{p_0^{-1} - 1\}$.

VIII. Numerical results

This section present numerical results that illustrate the findings presented in Sections V and Section VII. Our aim is to show how our approach allows to find quantitative conclusions about the achievable performance of social learning, providing an engineering perspective that complements the more qualitative results that exist in the social science literature.

In the sequel, we use our results to show to aspects of social learning. First, Section VIII-A presents an Algorithm based on the results of Section V to simulate decision sequences, and also illustrate how the decision threshold evolves in accordance to the learning process. Then, Section VIII-B uses the results of Section VII in order to to compare the achievable quality of the inference developed by social learning in scenarios with different private signals statistics. For simplicity, through this section we focus in the case of flat priors (i.e. $\eta = 0$) and $u_{1,1} = u_{0,0} = 0$ and $u_{1,0} = u_{0,1} = 1$ —and hence $\nu = 0$.

A. Simulations of decision sequences

The results presented in Section V-B allow us to develop an efficient algorithm to simulate decision sequences following diverse signal statistics. Algorithm 1 implements this, using as inputs η, ν, the state of the world w, the network size N, the distortion coefficients and the signal statistics in the form of the c.d.f. of the signal log-likelihood $F_0^\Lambda(\cdot)$ and $F_1^\Lambda(\cdot)$.

We used Algorithm 1 to generate decision sequences under diverse private signal statistics. Please note that if the private signals follow a Natural Exponential family distribution[‡] then the corresponding signal log-likelihood is simply a linear function. In fact, a Natural Exponential Family distribution p.d.f. can be expressed as

$$\mathbb{P}_w\{S_n = s\} = \exp\{s\cdot\theta(w) + h(s) - b(w)\}, \quad (35)$$

where $\theta(w)$ is the natural parameter, $h(\cdot)$ is the carrier measure and $b(\cdot)$ is the log-normalizer, and therefore the corresponding log-likelihood can be written as

$$\Lambda_{S_n}(S_n) = S_n[\theta(1) - \theta(0)] - [b(1) - b(0)]. \quad (36)$$

Because of this reason, Gaussian signals have unbounded private beliefs, as in this case $S_n \in (-\infty, \infty)$. On the contrary, as exponential distributions $S_n \in [0,\infty)$ then the corresponding beliefs are bounded in one side and unbounded in the other, allowing only cascades of 1's if $\theta(1) - \theta(0)$ is positive and only cascades of 0's if it is negative. On the other hand, any discrete distribution with finite support has bounded beliefs, as the private signal log-likelihood can take only a finite number of different values.

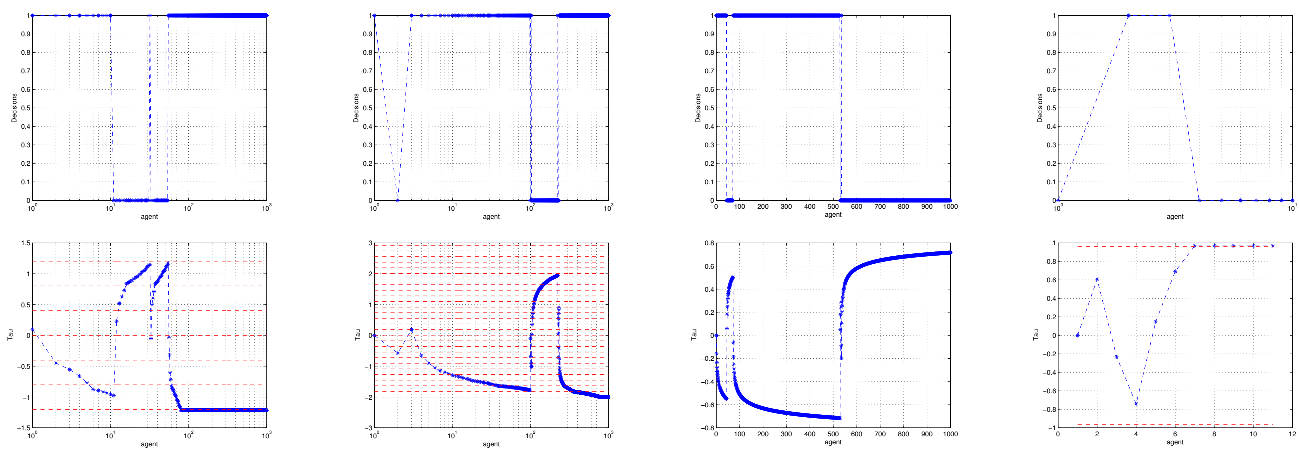
Our simulations reflect these results, and illustrate the insights discussed with respect to Proposition 1 and Theorem 2. For signals with bounded beliefs a global information cascade is triggered as soon as the evolution of the decision threshold τ_n drives it away from $[L_s, U_s]$. This is illustrated by Figure 9, where it can be seen that after τ_n goes beyond the red lines which are further away from zero (which correspond to the extreme values of $\Lambda_S(S_N)$) a sequence of homogeneous decisions is triggered. On the other hand, our simulations show that under Gaussian signals even very long sequences of equal decisions can be suddenly reversed when a signal with high enough log-likelihood is founded. As an example of this, Figure 9 shows how in a system with Gaussian private signals a sequence of 1's take place after a sequence of almost 500 consecutive 1's.

Interestingly, our simulations also show that decisions that confirm a trend have a diminishing impact (smaller step size) over the evolution of the decision threshold. For example, the plots that correspond to Binomials and Exponentials in Figure 9 show that the subsequent step-sizes decrease when the value of τ_n moves away from zero. On the contrary, the same figures show how decisions that go against the majority of previous choices induce important jumps, which corresponds to a high amount of new information that is included in the inference process. Correspondingly, the fact that τ_n is constant after a global information cascade is trigger corresponds to the fact that no new information is being processed by the social learning —and hence the accuracy of future agents does not increase further but stays constant.

[‡]Many well-known distributions belong to the Natural Exponential Family, including Bernoulli, Binomial, Poisson, Negative Binomial, Gaussian with known variance and Exponential distributions among others.

Algorithm 1 Simulation of social decisions

function DECISION_VECTOR(N, η, ν, w)
  $\tau_1 = \nu + \eta$.
  $\mathbb{P}_0\{X_1 = 0\} = F_0^{\Lambda}(\tau_1)$.
  $\mathbb{P}_1\{X_1 = 0\} = F_1^{\Lambda}(\tau_1)$.
  Generate $x_1 \sim$ Bernoulli($\mathbb{P}_w\{X_1 = 0\}$).
  for $n = 2, \dots, N$ **do**
    for $\forall \boldsymbol{g} \in \mathcal{G}_n$ **do**
      $\mathbb{P}_0\{\boldsymbol{G}_n = \boldsymbol{g}\} = \alpha_n(\boldsymbol{g}|\boldsymbol{x}_1)\mathbb{P}_0\{\boldsymbol{X}^{n-1} = \boldsymbol{x}^{n-1}\}$.
      $\mathbb{P}_1\{\boldsymbol{G}_n = \boldsymbol{g}\} = \alpha_n(\boldsymbol{g}|\boldsymbol{x}_1)\mathbb{P}_1\{\boldsymbol{X}^{n-1} = \boldsymbol{x}^{n-1}\}$.
      $\Lambda_{\boldsymbol{G}_n}(\boldsymbol{g}) = \log \frac{\mathbb{P}_1\{\boldsymbol{G}_n = \boldsymbol{g}\}}{\mathbb{P}_0\{\boldsymbol{G}_n = \boldsymbol{g}\}}$.
      $\tau_n(\boldsymbol{g}) = \nu + \eta - \Lambda_{\boldsymbol{G}_n}(\boldsymbol{g})$.
    $\mathbb{P}_0\{X_n = 0|\boldsymbol{X}^{n-1} = \boldsymbol{x}^{n-1}\} = \sum_{\boldsymbol{g} \in \mathcal{G}_n} \alpha(\boldsymbol{g}_n|\boldsymbol{x}^{n-1}) F_0^{\Lambda}(\tau_n(\boldsymbol{g}_n))$
    $\mathbb{P}_1\{X_n = 0|\boldsymbol{X}^{n-1} = \boldsymbol{x}^{n-1}\} = \sum_{\boldsymbol{g} \in \mathcal{G}_n} \alpha(\boldsymbol{g}_n|\boldsymbol{x}^{n-1}) F_1^{\Lambda}(\tau_n(\boldsymbol{g}_n))$
    Generate $x_n \sim$ Bernoulli($\mathbb{P}_w\{X_n = 0|\boldsymbol{X}^{n-1} = \boldsymbol{x}^{n-1}\}$).
    $\mathbb{P}_0\{\boldsymbol{X}^n = \boldsymbol{x}^n\} = \mathbb{P}_0\{X_n = x_n|\boldsymbol{X}^{n-1} = \boldsymbol{x}^{n-1}\} \cdot \mathbb{P}_0\{\boldsymbol{X}^{n-1} = \boldsymbol{x}^{n-1}\}$.
    $\mathbb{P}_1\{\boldsymbol{X}^n = \boldsymbol{x}^n\} = \mathbb{P}_1\{X_n = x_n|\boldsymbol{X}^{n-1} = \boldsymbol{x}^{n-1}\} \cdot \mathbb{P}_1\{\boldsymbol{X}^{n-1} = \boldsymbol{x}^{n-1}\}$.
  return $\boldsymbol{x}^N$



(a) Binomial (b) Poisson (c) Gaussian (d) Cauchy

Fig. 9: *Above*- Sequence of decisions in a social system generated using Algorithm 1, assuming perfect social information and diverse signals statistics. *Bellow*- Evolution of the corresponding decision threshold τ_n. Confirming Proposition 1 and Theorem 2, as soon as the decision threshold —*Tau*, Y-axis— reaches beyond the the possible values of the private signals (marked by red lines), the decision threshold stops evolving and the network falls into a global information cascade.

For completeness, simulations over signals with Cauchy distributions with fixed scale and variable location coefficients were included in order to illustrate that these results also hold under continuous signal with bounded private beliefs. The interested reader can find an analysis of Cauchy signals in Appendix C.

B. Evaluation of payoffs

Numerical evaluations confirmed the accuracy of the exact expressions and bounds derived in Section IV for the error rates. Figure 10 compares the performance attained by social networks with perfect social information when they are driven by various signal statistics. In accordance to the results presented in Section III, the error rates converge to a value larger than zero when the signals have bounded beliefs and hence global information cascades take place, while they converge to zero in other case (e.g. Gaussian signals in Figure 10). Interestingly, results suggest that the error rates do not converge to zero when the signals are bounded in one side —and hence only allow cascades of either 0's or 1's, which in Figure 10 corresponds to the Poisson distribution. It is also interesting how Cauchy distributions achieve a very poor performance, although being a continuous distribution (c.f. Appendix C).

Our results also confirmed the intuition that the asymptotic error rate depends mainly on the extreme values of the signal likelihood (c.f. the corresponding discussion in Section IV). In effect, if the signal parameters are chosen in a way that they possess the same values of U_s and L_s, then the asymptotic value is generally close to the value predicted by Theorem 3 (see Figure 11). It is to be noted that, if Poisson distributions are chosen in such a way that it's single extreme value of the

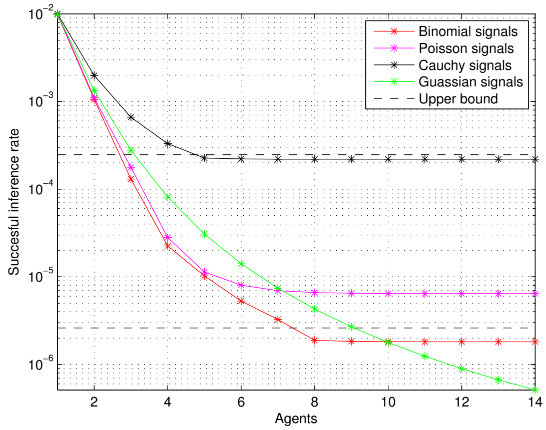


Fig. 10: Evolution of the error rate for various signal statistics, as given by (26), and the upper bound presented in (30). The parameters of the different distributions were chosen in order to provide a error rate of 10^{-2} for the first agent, while the binomial distribution has a support of $n = 6$.

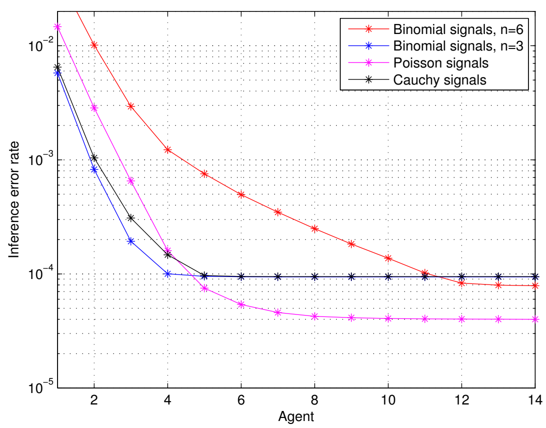


Fig. 11: Evolution of the error rate for various signal statistics, as given by (26), and the upper bound presented in (30). The parameters of the different distributions were chosen in order to provide a error rate of 10^{-2} for the first agent, while the binomial distribution has a support of $n = 6$.

signal log-likelihood coincide with one of the values of the other signals, it's asymptotic error rate is approximately half than the other; this corresponds with the fact that only half of the information cascades happen (c.f. the corresponding discussion about Poisson distributions in Section VIII-A).

Finally, although different signal statistics with similar extreme values converge to the same value, Figure 11 shows different convergence speeds. It is reasonably to postulate that the convergence speed is related to the Total Variation distance of the corresponding probability density functions [104], but the proof of this conjecture remains open.

C. Discussion

An important insight that springs out of our analysis is that information cascades are not intrinsically prejudicial for the performance, as their impact is conditioned over their accuracy. In effect, in a case where cascades would be always correct then the asymptotic performance would be perfect in despite of the cascading behaviour. Therefore, the asymptotic performance of social learning is not limited by the cascading itself, but by the corresponding "cascade accuracy" (c.f. Section VII-B). In the case of Bayesian strategies, our framework shows that the cascade accuracy is directly related with the extreme values of the private signal log-likelihood. Interestingly, this allows to conclude that social learning can provide error rates as small as desired if the system designer can engineer the private signal statistics appropriately. Therefore, this data aggregation method does not impose, by itself, any lower bounds on the achievable performance.

The proposed framework is based on two important simplifying assumptions: the conditional independency of the private signals and the assumption of perfect rationality of the agents. Please note how these assumptions allow a detailed analytic approach, which allows to generate a first understanding on such a challenging topic.

IX. Application to cyber-physical system security

This section reviews an application of information cascades in the field of cyber-physical security, following the work reported in [108]. This application is novel in taking advantage of some of characteristics of information cascades that are traditionally regarded as undesired. In the following, first Section IX-A provides the necessary context and discusses the application. Then, Section IX-B presents the system model and fundamental assumptions. The leveraging of information cascades in this scenario is discussed in Section IX-C, and finally some numerical results are presented in Section IX-D.

A. CPS security as a dilemma in sensor networks design

Security plays a critical role in cyber-physical systems (CPSs), particularly for those involved in public utilities whose safety is critical [21]. In effect, recent attacks to public CPSs that created significant damages have been widely reported. There exist many technological challenges for the security in nowadays's cyber-physical systems. Moreover, as more cyber-control automation is progressively entering our daily life, guaranteeing CPS security will become even a more challenging subject. As the level of security is determined by the weakest element of the entire system, one major dilemma lies in the information fusion that takes place on the sensor networks that supply vital information to control and manage the CPS. The main weakness of these sensor networks are described in the sequel.

- In such networks the number of sensors is significantly large, and hence sensors are usually deployed randomly (i.e. the location of each sensor might be unknown to the system controller). A precise management of them is therefore challenging or even unfeasible. Furthermore, a good portion of the sensors might be deployed in geographical regions where it is not possible to warrant physical or cyber security (e.g. war zones or regions controlled by an adversary party).

- Typical low-end (i.e. low-cost) sensors are only able to perform low-complexity computing and networking. Therefore, it is nearly impossible to implement reliable security functions and protocols of high complexity.
- Sensor data from these low-complexity sensors may be unreliable due to malfunctions in terms of measurement, computing, or battery. Data can also be corrupted or delayed because of link outages or packet errors resulted from noise and interference in wireless communications, or also due to spectrum-sharing or energy-harvesting opportunities.

The vulnerability of these systems makes it reasonable to expect that they might be victims of cyber/physical attacks by intelligent adversaries. This is particularly critical to low-end sensor networks that are usually at the edge of large and complex CPS.

Due to above reasons, although the technical challenges of the design of secure wireless sensor networks have been widely studied [109], there remain open problems of both theoretical and engineering nature [110]. Attacks to wireless sensor networks are commonly categorized into outside attacks and insider attacks. Outside attacks include (distributed) denial of services (DoS) attacks facilitated by cyber- or physical-means, which are facilitated by the broadcasting nature for wireless communications [109]. Insider attacks can create potentially more severe harm to CPS, where the adversary can recruit low-complexity sensors by malware through cyber/wireless means, or by physical substitution of targeted sensor nodes. These compromised sensor nodes can report false data in order to create harmful results and malfunctions, which is related to the well-known Byzantine Generals problem [111]. Low-complexity sensors at the edge of CPS are natural targets of such insider attacks.

A traditional sensor network consists of a fusion center and regular sensors nodes. Most of the literature assume that the fusion center is capable of executing secure coding and protocols. It is to be noted that in large CPS, the sensors at the edge of the network may require another kind of mediator devices known as data aggregators (DAs), which have the capability to access the cloud through high-bandwidth communication [112]. DAs are surely attractive to insider attacks. Differing from almost all existing research, this project assumes that DAs and fusion centers are possible to be recruited and compromised.

In response to the technology challenges related to the development of resilient low-end sensor networks that are capable of facing an intelligent adversary, the desirable sensor fusion mechanism shall be robust and adaptive to the existence of compromised sensor nodes and false data. In this way, the adapted operation can be subsequently reconstituted [113] if there exists a reliable and secure management. This is similar to the concept of cyber resilience, which was officially proposed in 2012 World Economy Forum. Therefore, the goal of the proposed methodology is

(1) to establish a resilient operation with the presence of an unknown number of compromised sensors, and
(2) to enable a resilient sensor fusion even in the presence of false data from compromised sensors.

B. Scenario description and main assumptions

Consider a sensor network composed of N sensor nodes, which are deployed over an area for the purpose of monitoring or surveillance. Based on the sensor's signals, the network shall infer the value of the binary random variable W, with events $\{W = 1\}$ and $\{W = 0\}$ corresponding to the presence or absence of an intrusion, respectively. No knowledge about of the prior distribution of W is assumed, as intrusions are rare and have unknown/unpredictable patterns.

Battery limitations impose severe restrictions on the communication between sensors, and hence each node is assumed to forwards data to others by broadcasting only a binary variable. Under a medium access control mechanism, and without loss of generality, sensor nodes are assumed to transmit their signals sequentially according to their indices. Due to the nature of wireless broadcasting, nearby transmissions can be overheard. Therefore, it is assumed that the n-th node can generate its decision based on its own sensor output and the signal exchanged by other sensors.

A data aggregator or fusion center collects the transmitted data and is recognized as a specific node denoted as $n_{\text{FC}} \in \{1, ..., N\}$. The performance of the entire sensor network is quantified by the corresponding miss-detection and false alarm rates in the fusion center, given by $\mathbb{P}_{\text{MD}} = \mathbb{P}\{X_{n_{\text{FC}}} = 0 | W = 1\}$ and $\mathbb{P}_{\text{FA}} = \mathbb{P}\{X_{n_{\text{FC}}} = 1 | W = 0\}$ respectively.

1) Powerful insider attack: During the attack, it is assumed that there are N^* Byzantine nodes controlled by an adversary, while the network management does not know this situation. It is important to notice that these Byzantine nodes may include DAs or FCs. The adversary can therefore freely define the values of the binary signals transmitted by Byzantine nodes in order to degrade the sensor network performance, which might be viewed as a man-in-the-middle attack or false data inject attack. It is further assumed that the adversary is topology-aware, knowing the sensor sequence and the strategy in use. In other words, this is a very powerful attack from inside of the sensor network.

2) Defense without knowledge of attack: In most (surveillance) sensor networks, miss-detections are more important than false alarms. Furthermore, it is difficult to estimate the cost structure under the worst-case scenario. Therefore, the Neyman-Pearson criteria shall be selected by setting an allowable false alarm rate and focusing on the achievable miss-detection rate. Most signal processing techniques for distributed detection rely on a FC(s) that gather data and generate estimators [114]. In order to guarantee diversity, traditional distributed detection schemes choose to ignore previously broadcasted signals. However, as regular sensors do not perform any data aggregation, each of the overheard signals cannot serve as a good estimator of the target variable. When there exist Byzantine nodes in sensor network, various techniques have been proposed, such as identification and removal of Byzantine sensors, cryptography, secure protocols [115], but none of them consider potential Byzantine DAs or FCs. Consequently, it is very desirable for sensor network management to find an appropriate network-resilient strategy to mitigate the effect from this powerful topology-aware adversary, especially when the network manager (i.e. defender) has no knowledge of the number of Byzantine nodes or other statistics of the attack.

Inspired by collective behavior and social learning, a totally different philosophy is undertaken to face this problem. Each sensor node can be considered to be a rational agent that decides sequentially about the presence of attacks, based on a Bayesian data fusion of their measurements and overheard signals from other nodes. Of course, the Byzantine sensors do not follow this strategy as their goal is to bring down the sensor fusion performance. Let $\mathcal{B}$ denote the set of indices of Byzantine nodes in the sensor network, where $N^* = |\mathcal{B}|$. In the example of intrusion detection, as events $\{W = 0\}$ are much more frequent than $\{W = 1\}$, any abnormal increase of the false alarm rate would be quickly noted by the operator, which is undesirable to the adversary. Consequently, the adversary strategy is to increase the miss-detection rate as much as possible, which is achieved by forcing null signals for all $n \in \mathcal{B}$.

C. Leveraging information cascade

Intuitively, the accuracy of the n-th sensor grows with n, and hence n_{FC} is usually chosen as one of the last nodes in the decision sequence. However, as the number of shared signals grow, the increasing social belief can make the nodes to ignore their individual measurements and fall into and information cascade. Interestingly, one unique aspect of this approach is to identify a positive effect of information cascades, which has been overlooked before. In effect, information cascades make a large number of nodes to hold equally qualified estimators, generating a large number of locations in the network where the network operator can collect aggregated data. This property avoids single points of failure, providing robustness against topology-aware false data injection attacks.

On the other hand, an attacker can also leverage the information cascade phenomenon. In fact, a rational attacking strategy is to tamper the first N^* nodes of the decision sequence, setting their signals in order to push the networked decisions towards a misleading cascade or a misleading public belief. If N^* is large enough, an information cascade can be triggered almost surely, making the learning process to fail. However, if N^* is not large enough then the network may undo the initial pool of wrong opinions and end up triggering a correct cascade anyway. Therefore, to achieve resilient sensor fusion against false data attacks, the trustworthy sensor networking shall be conducted in clusters for any large sensor network. Luckily, since sensors are usually equipped with very short-range radios and sequentially multiple access, such a clustering strategy is consistent with engineering reality.

More precisely, the proposed sensor fusion does not solely rely on the security mechanisms to against the attacks, but take advantage of public belief to enhance its capability against attacks. As DA or FC usually serves as the last node to transmit the measurement (i.e. announce the decision), both measurement and public belief are simultaneously disclosed to report fusion result for CPS operation and implicit security status for the trust management in the CPS. Please note that a social learning data fusion mechanism is generally compatible to any cryptograph and secure networking protocol.

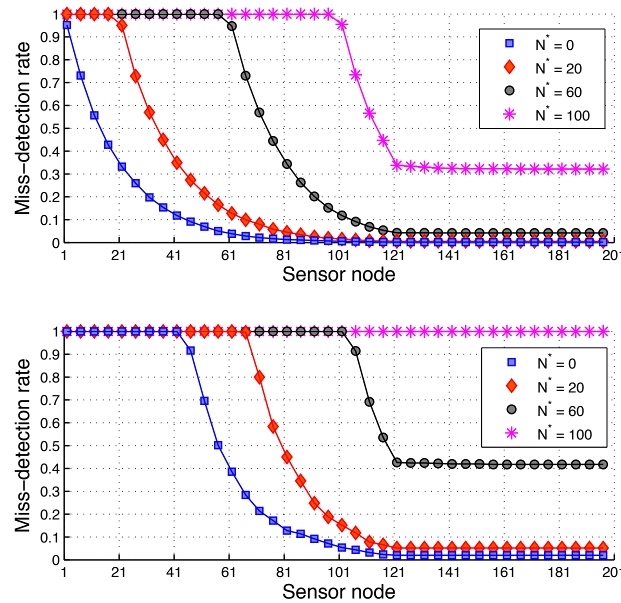


Fig. 12: The top figure presents the performance (i.e. miss-detection rate for fusion) of a surveillance sensor network of 200 nodes, while the lower figure shows the same quantity when only the 10% less favorable cases are considered. Different curves correspond to different number of compromised sensors. When considering total miss-detection rate, up to 60 compromised sensors —that is 30% of sensor nodes— the fusion still functions well. The performance floor for 100 compromised sensors implies exceeding the bound of the Byzantine Generals Problem.

D. Numerical results

To illustrate the application of social learning against topology-aware data falsification attacks, a network of randomly distributed sensors over a sensitive area following a Poisson Point process (PPP) was considered. The ratio of the total area that falls under the range of each sensor is denoted by r. It is assumed that intrusions can occur uniformly over the surveilled area, and hence the probability of an intrusion taking place under the coverage area of a particular sensor is equal to r. It is further assumed that each node is equipped with a binary sensor (i.e. $S_n \in \{0, 1\}$) that might generate wrong measurements due to electronic and other imperfections. Figure 12 shows the social learning based fusion successfully against Byzantine data attacks.

X. Further applications

This section presents a brief exploration to other applications of information cascades and social learning. In the following, Section X-A explains how the steering of information cascades can be supervised and controlled, and discusses promising applications to e-marketing and e-commerce. Then, Section X-B introduces the potential applications of information cascades to consensus building problems. Finally, Section X-C gives a vision on the way of connecting social learning and machine learning applications.

A. Steering information cascades

This subsection presents a discussion about design considerations for steering information cascades in a social system where agents perform Bayesian social learning, following the work reported in [39].

1) Design of social systems: Information cascades play a sensible role in e-commerce and e-marketing [39]. In effect, it is intuitive that customers might tend to choose a given product when they find many positive comments about it online. Conversely, customers might be prone not to choose the product when facing a significant number of negative reviews. Therefore, for marketing purposes, an information cascade of positive reviews is highly desirable as it can increase sales volume. It is clear that information cascades can usually be steered by manipulating social information, e.g. by creating biased reviews or introducing fake statistics. Therefore, if people have the ability to selfishly manipulate information cascades, then the integrity and trust of the e-commerce system will be compromised. Therefore, the design of tools for preventing fake information cascades is a fundamental challenge for the well-being of a digital society [5].

To approach this issue, a first step is to analize (8) using a new perspective, from which Λ_{S_n} is interpreted as *personal evidence*, $\Lambda_{\boldsymbol{G}_n}$ as the *social observation*, η_n as the *bias* and ν_n as the *incentive structure* of the n-th agent. In most scenarios it is not possible to control the personal evidence and bias, which can only be estimated experimentally. However, in many cases the scope of the social observation and the incentive structure are susceptible to be engineered, as the system manager usually can modify the portion of the social network that is available to the inspection of other agents (e.g. determining a group of desirable reviews that shall be shown to new customers). Similarly, the incentive structure can also be engineered, say, by acknowledging good reviews with special bonus, coupon, or discounts.

Following this rationale, selective rewiring is proposed as a method to control or steer information cascades when the social observations can be controlled by the system administrator. On the other hand, incentive seeding is served for scenarios where the incentive structure can be engineered.

2) Selective rewiring: Let us denote by $\mathcal{H}_k$-cascade as a cascade where $X_m = k$ for all $m \geq n_c$, and by $\mathcal{B}_n$ the set of neighbours of the n-th agent. The intuition behind the proposed selective rewiring approach is that, in order to steer a $\mathcal{H}_k$-cascade, one can restructure the observation neighbourhood of the n-th agent $\mathcal{B}_n$ to connect it with previous decisions such that $X_j = k$, or to disconnect it previous decisions where $X_j \neq k$, where $1 \leq j < n$. The number of connections to be added or deleted, denoted as N_z and N_w respectively, are parameters to be defined by the system controller. A pseudo code for implementing this idea is presented in Algorithm 2.

3) Incentive Seeding: In the case when the incentive structure can be engineered, the proposed approach is to first identify the most influential agents within the system, and set their incentive structure to the value ν^* in order to make their decisions consistent with the desired cascade. Although there exist multiple ways of defining "social influence", in this approach individuals were selected according to their *out-degree* number, denoted as d_n^{out}, which correspond to agents whose decisions is seen by the larger number of other agents [67]. The pseudo-code of the proposed algorithm is presented in algorithm 3.

Algorithm 2 Selective Rewiring

1: **Input:** $N_z, N_w, \mathcal{H}_k$
2: **for** $n = 1, \dots, N$ **do**
3: Find $\mathcal{Z}_n = \{z | X_z \neq \mathcal{H}_k\}$, $\mathcal{Z}_n \subset \mathcal{B}_n$, $|\mathcal{Z}_n| = N_z$
4: Find $\mathcal{W}_n = \{w | X_w = \mathcal{H}_k\}$, $\mathcal{W}_n \cap \mathcal{B}_n = \varnothing$, $|\mathcal{W}_n| = N_w$
5: $\mathcal{B}_n \leftarrow (\mathcal{B}_n \backslash \mathcal{Z}_n) \cup \mathcal{W}_n$

Algorithm 3 Incentive Seeding

1: **Input:** $d_{min}^{out}, \nu^*, \mathcal{H}_k$
2: **for** $n = 1, \dots, N$ **do**
3: **if** $d_n^{out} \geq d_{min}^{out}$ **then**
4: $\nu_n \leftarrow \nu^*$

4) Numerical results: The algorithms were tested over various networks of different topologies, including fully-connected networks, Erdös-Renyi random networks, small-world networks, and scale-free networks [38], [67], [116]. The private signals were assumed to follow a Binary Symmetric channel with respect to the state of the world variable, similar to the case analyzed in Section VI-B. The parameter of the Binary Symmetric channel, denoted here as p, corresponds to the agent's private signal quality. Numerical simulations were perform to compare how agents can form information cascades under various setting, and to investigate the ways to steer cascades efficiently.

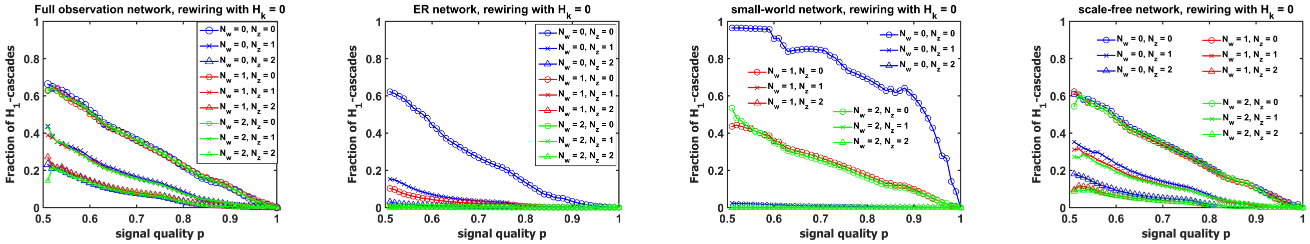
Figures 13 and 14 demonstrate the ratio of $\mathcal{H}_1$-cascades that can be steered by a system controller when the true state of the world is $W = 0$, and see the relation to the signal quality. Interestingly, results show that the social network topology significantly affect the steering capability, which suggests some practical system design considerations with respect to particular types of network topology:

- For Erdös-Renyi networks, adding more links progressively is an effective way of steering cascades.
- For small-world and scale-free networks, it is better to combine the use of adding more new links and deleting bad links.
- Small-world networks have small diameters, which helps to resist control.
- Scale-free networks are similar to full observation networks.

Moreover, following these results, practical advices for dealing with a network of an unknown type can be formulated:

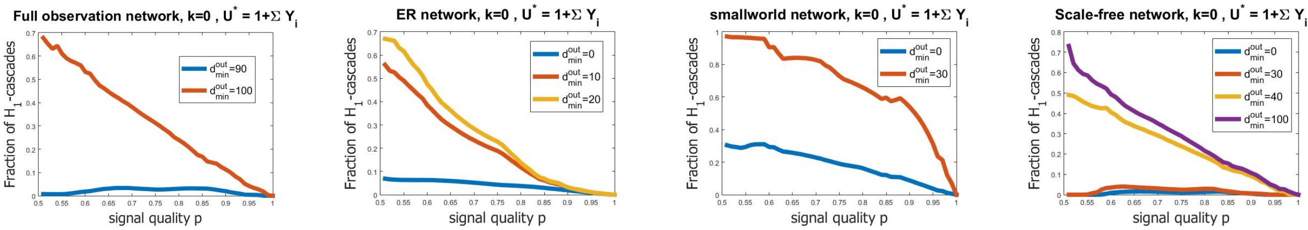
- Deleting links is less effective than adding links.
- Merely connecting correct decisions do not perform well.
- It is better to start by discarding some wrong decisions, and then to continue connecting correct decisions.

5) Discussion: The proposed methods can serve as a simple guideline to the design of social systems based on manageable



(a) Full observation network (b) ER network (c) Small-world network (d) Scale-free network

Fig. 13: Ratio of $\mathcal{H}_1$-cascades that can be steered by selective rewiring when $W = 0$ under various network topologies. Intuitively, when the private signal is very good then all cascades follow the state of the world, hence reducing the possibilities of steering other cascades.



(a) Full observation network (b) ER network (c) Small-world network (d) Scale-free network

Fig. 14: Ratio of $\mathcal{H}_1$-cascades that can be steered using incentive seeding under different social network topologies. The value of d_{min}^{out} determines the minimal outdegree needed to be considered a influential agent whose incentives should be modified. Selecting an appropriate d_{min}^{out} has a important impact over the algorithm performance [39].

online digital platforms. Moreover, it was demonstrated that it feasible for a system designer to identify control variables of a complex social system and use them to devise effective unsophisticated algorithms for steering user behavior.

The presented numerical simulations illustrate the idiosyncrasies of different network topologies. These results leads to the question of how to design efficient algorithms for steering information cascades under more general constraints in the network topology. Finally, another important issue is how the incentive can be embedded in the network structure, which points out interesting further research developments.

B. Consensus building in multi-agent system

Multi-agent systems abound in our modern world, including groups of mobile robots and unmanned aerial vehicles (UAVs), aircraft and satellites, and many more [19]. In these systems agents need to make autonomous decisions, but at the same time the controller need to ensure that the group as a whole can achieve some desired goal. A primary challenge in the design multi-agent systems is the optimization of control and computing aspects considering trade-offs between performance, delay and communication costs or constraints [117].

Consensus building plays a crucial role in most of multi-agent system protocols, being the subject of many works in the control engineering literature. In particular, the flocking problem [118] considers a set of autonomous agents that can observe their neighbors dynamics (position, velocity, etc.) in a 3-dimensional space, and use this data to adjusts their movements accordingly to flock together with other agents. In general the connectivity of a network of flocking agents depends on the geometry of the space. The simplest connectivity model may be the one where two agents i, j are said to be connected if their distance is smaller than a radius r, that is,

$$\|\vec{q}_i - \vec{q}_j\| < r,$$

where $\vec{q}_i$ and $\vec{q}_j$ are vectors that denote the position of the agents in space. The design of an efficient flocking protocol is full of challenges; we conjecture that information cascades and social learning could be applied to the analysis and design of flocking protocol, which is an open problem to be investigated.

Another interesting scenario, more related to complex social systems modeling [80], is the one provided by the Hegselmann-Krause opinion dynamics model [119], [120]. As in the DeGroot model [93], the Hegselmann-Krause opinion dynamics model assumes that the opinion of the i-th agent can be described as the weighted average of other agents' opinions, i.e.

$$x_i(t+1) = a_{i1}x_1(t) + a_{i2}x_2(t) + ... + a_{in}x_n(t),$$

which can also be written in matrix form as $\mathbf{x}(t+1) = A\ \mathbf{x}(t)$. It is noted that, in either the flocking problem or the Hegselmann-Krause opinion dynamics model, agents are assumed to be able to observe their neighbor's characteristics. Hence, agents' actions are based on these informed observations from their counterparts, and therefore information cascades are likely to play an important role on their dynamics.

We emphasize that consensus problems belong to the regime of social learning problems. Definitely, consensus protocols in multi-agents systems are a special type of non-Bayesian social learning strategy. How to utilize the characteristics of information cascades to the study of flocking and consensus problem in multi-agent system is of paramount interest for a wide range of applications related to large-scale CPS.

C. Social learning for machine intelligence

Our interpretation of social learning as a data aggregation scheme can be applied to machine learning [121], [122]. In effect, by considering a scenario where there are a number of interconnected learning machines, we can view each machine as a social agent and the whole system as a social network, where the neighbourhoods are determined by the degree of interaction between different machines.

The primary objective of learning machines is to learn from the patterns in their observations about the environment, and hence minimize their prediction error with respect to a given loss function. When the observation includes both the data generated from the state of the world and the prediction made by other learning machines, this scenario fit in the social learning framework presented in this work. With the detailed analysis of the average cost function in Section VII, inspired by the work of [123], [124], it is noted that one can characterize the interdependency among the loss functions of learning machines. Therefore, the problem is how to design a mechanisms for a networked machine learning system, which can be very relevant for problems of *Ensemble Learning*. Analytic techniques of social learning can be applied to study the complex feedback that exists between each learning machine and the whole networked system. Do groups of learning machines generate information cascades? Is the knowledge about information cascades suitable to be used for improving the achievable performance of systems of networked learning machines? These issues constitute a promising application of our proposed framework.

XI. Concluding remarks

Although this paper presents a systematic analytical methodology and an extensive literature survey with illustrating examples, many aspects of collective behavior in human society and online platforms remain open to human knowledge. Moreover, the impact they have on the technological development and our human society urge for further explorations. Many possibilities to assist future technological systems design are still on the horizon, and rely on the readers of IEEE Access to make them come true.

Appendix A
Properties of F_w^Λ

In general $F_w^\Lambda(x)$ can be directly computed from the statistics of the signal distribution. For simplicity let us consider the case of real-value signals, i.e. $S_n \in \mathbb{R}$. In this case, the c.d.f. of the signal likelihood is given by

$$F_w^\Lambda(y) = \int_{\mathcal{S}^y} \mathrm{d}\mu_w \tag{37}$$

where $\mathcal{S}^y = \{x \in \mathbb{R} | \Lambda_s(x) \leq y\}$. If Λ_s is an increasing function, then $\mathcal{S}^y = \{x \in \mathbb{R} | x \leq \Lambda_s^{-1}(y)\} = (-\infty, \Lambda_s^{-1}(y)]$ and hence

$$F_w^\Lambda(y) = \int_{-\infty}^{\Lambda_s^{-1}(y)} \mathrm{d}\mu_w = H_w(\Lambda_s^{-1}(y)) \ , \tag{38}$$

where $H_w(s)$ is the cumulative density function (c.d.f.) of S_n for $W = w$. For the general case where Λ_s is an arbitrary (piece-wise continuous) function, then $\mathcal{S}^y$ can be expressed as the union of intervals. Then $\cup_{j=1}^{\infty}[a_j(y), b_j(y)] = \mathcal{S}^y$ (note that $\Lambda_s(a_j(y)) = \Lambda_s(b_k(y)) = y$) and hence from (37) is clear that

$$F_w^\Lambda(y) = \sum_{j=1}^{\infty} \int_{a_j(y)}^{b_j(y)} \mathrm{d}\mu_w = \sum_{j=1}^{\infty} [H_w(b_j(y)) - H_w(a_j(y))] \ .$$

Appendix B
Proofs of Section V

Proof of Proposition 2: To prove the Markovianity, let us first note that (V-B) implicitly show that X_n is conditionally independent of $\boldsymbol{X}^{n-1}$ given τ_n (in this case $\boldsymbol{G}_n = \boldsymbol{X}^{n-1}$). Therefore, (12) shows that $\tau_{n+1} = \tau_n - \lambda(X_n, \tau_n)$, and therefore if τ_n is given then τ_{n+1} depend only on X_n and hence is conditionally independent on $\boldsymbol{X}^n$ as well. Finally, the Markovianity is proven by realizing that τ_k is a deterministic function of $\boldsymbol{X}^n$, and hence also conditionally independent of τ_{n+1} given τ_n.

To prove that $\{\tau_n\}_{n=1}^{\infty}$ is still a sub- or super- martingale with respect to $\boldsymbol{X}^n$, depending on the realization of $W = w$. In fact, it is direct to see that

$$\mathbb{E}\left\{\tau_n | W = w, \boldsymbol{X}^{n-1}\right\} = \tau_{n-1} - D_{X_j|\boldsymbol{x}^{j-1}} \tag{39}$$

where $D_{X_j|\boldsymbol{x}^{j-1}}$ is the conditional mean value of $\Lambda_{X_j|\boldsymbol{X}^{j-1}}(X_j|\boldsymbol{x}^{j-1})$, which can be computed as

$$\begin{aligned} D_{X_j|\boldsymbol{x}^{j-1}} &= \mathbb{E}\left\{ \log \frac{\mathbb{P}_1\left\{X_j|\boldsymbol{x}^{j-1}\right\}}{\mathbb{P}_0\left\{X_j|\boldsymbol{X}^{j-1}\right\}} \Big| W = w, \right\} \\ &= \begin{cases} D(\, p_{1|1,\boldsymbol{x}^{j-1}} || \, p_{1|0,\boldsymbol{x}^{j-1}}\,) & \text{if } w = 1 \text{ , and} \\ -D(\, p_{1|0,\boldsymbol{x}^{j-1}} || \, p_{1|1,\boldsymbol{x}^{j-1}}\,) & \text{if } w = 0 \text{ ,} \end{cases} \end{aligned}$$

where $p_{x_j|w,\boldsymbol{x}^{j-1}} = \mathbb{P}_w\left\{X_j = x_j | \boldsymbol{X}^{j-1} = \boldsymbol{x}^{j-1}\right\}$ and $D(p||q) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}$ is the Kullback-Leiver divergence between two Bernoulli distributions with parameters p and q, respectively. The Lemma is finally proven by the well-known non-negativity of the Kullback-Leiver divergence [102]. ■

Proof of Theorem 1: To start, let us note that for any $x \in \mathbb{R}$ the function $\Lambda_S(s)$ introduces a partition over the signal space $\mathcal{S} = \mathcal{S}^0(x) \cup \mathcal{S}^1(x)$, where $s \in \mathcal{S}^0(x)$ if $\Lambda_S(s) < x$ and $s \in \mathcal{S}^1(x)$ if $\Lambda_S(s) \geq x$. By comparing this with (10), one can see that $\mathcal{S}^j(\tau_n)$ are the signals that cause $\Lambda_S(S_n) \in \mathcal{K}_n^j$, and therefore is clear that

$$\mathbb{P}\left\{S_n \in \mathcal{S}^j(\tau_n)\right\} = \mathbb{P}\left\{X_n = j | \tau_n\right\} \ . \tag{40}$$

Let us prove first that (i) ⇔ (ii). If $\tau_n > U_s$ then $\mathcal{S}^1(\tau_n) = \varnothing$ and therefore $X_n = 0$ for all possible signals S_n. Similarly, if $\tau_n < L_s$ then $X_n = 1$ for all possible signals. On the other hand, if $tau_n \in [L_s, U_s]$ then is direct to see that there exist signals such that $X_n = 0$ or $X_n = 1$, and hence X_n and S_n are not independent. Therefore, the condition $\tau_n \in (-\infty, L_s] \cup [U_s, \infty)$ holds if and only if X_n and S_n are conditionally independent given τ_n.

Moreover, from the discussion in Section V-A is clear that in a Bayesian setup $X_n = \Phi(S_n, \boldsymbol{X}^{n-1})$ is a deterministic function of the private signal and the previous decisions.

Hence, the conditional independency of X_n and S_n given τ_n —which is a function of $\boldsymbol{X}^{n-1}$— is equivalent to X_n to be a deterministic function of $\boldsymbol{X}^{n-1}$. This is equivalent for X_n and W to be conditionally independent given $\boldsymbol{X}^{n-1}$, which in turns is equivalent to $\mathbb{P}\left\{X_n|W=1,\boldsymbol{X}^{n-1}\right\} = \mathbb{P}\left\{X_n|W=0,\boldsymbol{X}^{n-1}\right\}$, which guarantees (ii).

Let us show that (i) $\Rightarrow$ (iii). The previous discussion showed that (i) implies that X_n and S_n to be conditionally independent given $\boldsymbol{X}^{n-1}$. Moreover, the fact that (i) implies (ii) and $\tau_{n+1} = \tau_n - \Lambda_{X_n|\boldsymbol{X}^{n-1}}(X_n|\boldsymbol{X}^{n-1})$ shows that $\tau_m = \tau_n$ for all $m > n$. This, in turn, implies that (i) holds for all $m > n$. Therefore, it can be seen that X_m is a deterministic function of $\boldsymbol{X}^{n-1}$, and hence X_m and S_m are conditionally independent on $\boldsymbol{X}^{n-1}$ for all $m > n$, proving (iii).

To prove (iii) $\Rightarrow$ (i), let us assume that (i) does not hold and show this implies that (iii) also doesn't. If $\tau_n \in (L,U)$ implies that both $\mathcal{S}^0(\tau_n)$ and $\mathcal{S}^1(\tau_n)$ are both not empty, and hence $\min\{\mathbb{P}\{X_n=1|\tau_n\}, \mathbb{P}\{X_n=0|\tau_n\} > 0$. The fact that both probabilies are positive implies that there are signals $s \in \mathcal{S}$ that make $X_n = 1$ and others that make $X_n = 0$, and hence X_n and S_n are not conditionally independent given τ_n, or equivalently $\boldsymbol{X}^{n-1}$. Therefore the system is not in a cascade. ∎

Proof of Lemma 2: The proof can be done directly using an induction over n, and is left to the interested reader. ∎

Proof of Theorem 2: Let us assume that G_n has consistent distortion and consider $g \in \mathcal{G}_n$ is such that $\tau_n(g) > U_s$. Then, considering the fact that

$$\begin{aligned}\Lambda_{G_n}(g) &= \log\frac{\mathbb{P}_1\{G_n=g\}}{\mathbb{P}_0\{G_n=g\}}\\ &= \log\frac{\sum_{\boldsymbol{x}^{n-1}\in\mathcal{A}_n}\alpha_n(g|\boldsymbol{x}^{n-1})\mathbb{P}_1\left\{\boldsymbol{X}^{n-1}=\boldsymbol{x}^{n-1}\right\}}{\sum_{\boldsymbol{x}^{n-1}\in\mathcal{A}_n}\alpha_n(g|\boldsymbol{x}^{n-1})\mathbb{P}_0\left\{\boldsymbol{X}^{n-1}=\boldsymbol{x}^{n-1}\right\}},\end{aligned}$$

it can be shown using Lemma 0 that there exist at least one $\boldsymbol{x}^{n-1}$ such that $\alpha_n(g|\boldsymbol{x}^{n-1}) > 0$ and $\tau_n^{\text{full}}(\boldsymbol{x}^{n-1}) > U_s$. Above, $\mathcal{A}_n = \{\boldsymbol{x}^{n-1} \in \{0,1\}^{n-1}|\alpha(g|\boldsymbol{x}^{n-1}) > 0\}$. Then, due to the consistent distortion, all $\boldsymbol{x} \in \{0,1\}^{n-1} \in \mathcal{A}_n$ also satisfy $\tau_n^{\text{full}}(\boldsymbol{x}^{n-1}) > U_s$.

A similar derivation shows that, for any $\boldsymbol{x}^{n-1} \in \mathcal{A}_n$, all $g' \in \mathcal{G}_n$ such that $\alpha_n(g'|\boldsymbol{x}^{n-1}) > 0$ also satisfy $\tau_n(g') > U_s$. Therefore, is clear that

$$\begin{aligned}\mathbb{P}_w\{X_n=0|\boldsymbol{X}^{n-1}=\boldsymbol{x}^{n-1}\} &= \sum_{g'\in\mathcal{G}_n}\alpha_n(g'|\boldsymbol{x}^{n-1})\mathbb{P}_w\{X_n=0|G_n=g'\}\\ &= \sum_{g'\in\mathcal{G}_n}\alpha_n(g'|\boldsymbol{x}^{n-1})F_w^{\Lambda}(\tau_n(g'))\\ &= 1 \ ,\end{aligned} \tag{41}$$

where the last equality is due to the fact that if $\tau_n > U_s$ then $F_w^{\Lambda}(\tau_n) = 1$ and also that for a fixed $\boldsymbol{x}^{n-1}$ the terms $\alpha_n(g'|\boldsymbol{x}^{n-1})$ form a p.d.f. With this result, it can be shown that

$$\begin{aligned}\Lambda_{\boldsymbol{X}^n}(\boldsymbol{X}^n) &= \log\frac{\mathbb{P}_1\left\{X_n|\boldsymbol{X}^{n-1}\right\}}{\mathbb{P}_0\left\{X_n|\boldsymbol{X}^{n-1}\right\}} + \Lambda_{\boldsymbol{X}^{n-1}}(\boldsymbol{X}^{n-1})\\ &= \Lambda_{\boldsymbol{X}^{n-1}}(\boldsymbol{X}^{n-1})\end{aligned} \tag{42}$$

almost surely, which in turns shows that $\tau_{n+1}^{\text{full}}(\boldsymbol{X}^n) > U_s$ almost surely as well. Finally, the fact that all decision vectors $\boldsymbol{x}^n$ that have positive probability guarantee $\tau_n > U_s$, combined with the consistent distortion condition and Lemma 2 guarantee that $\tau_n(G_{n+1}) > U_s$ almost surely. This, combined with Proposition 1, proves the desired result.

The proof for the the case $\tau_n(g) < L_s$ is analogous and is not included. ∎

Appendix C
Analysis of systems with various private signal structure

A. Binomial distribution

Let us consider the case where the signals follow a Binomial distribution with parameters q_w and n, i.e.

$$p_w(s) = \binom{n}{s} q_w^s (1-q_w)^{(n-s)} \ . \tag{43}$$

By assuming without lack of generality that $q_1 > q_0$, then the signal log-likelihood is a linear function given by

$$\Lambda_S(s) = s\log\frac{q_1}{q_0} + (n-s)\log\frac{1-q_1}{1-q_0} \ . \tag{44}$$

Note that $\Lambda_S(s)$ is bounded with $L_s = \Lambda_S(0) = n\log\frac{1-q_1}{1-q_0}$ and $U_s = n\log\frac{q_1}{q_0}$. Moreover, using XX and XX and XX one can find that

$$\mathbb{P}_w\left\{X_n=0|\boldsymbol{X}^{n-1}\right\} = \begin{cases}0 & \text{if } \tau_n < L_S,\\ \sum_{k=0}^{k^*(\tau_n)} p_w(k) & \text{if } \tau_n \in [L_S,U_S],\\ 1 & \text{if } \tau_n > U_S.\end{cases}$$

where $k^*(\tau_n)$ is the largest integer k such that $\Lambda_S(k) \leq \tau_n$. By defining $k^* = -1$ when $\tau_n < L_s$ and taking the convention $\sum_{k=0}^{-1} f(k) = 0$, then

$$\begin{aligned}\Lambda_{X_n|\boldsymbol{X}^{n-1}}(X_n|\boldsymbol{X}^{n-1}) =& X_1\log\frac{\sum_{j=k^*+1}^{n}p_1(j)}{\sum_{j=k^*+1}^{n}p_0(j)} &(45)\\ &+ (1-X_n)\log\frac{\sum_{j=1}^{k^*}p_1(j)}{\sum_{j=1}^{k^*}p_0(j)} &(46)\\ :=& f(X_1,k^*) \ . &(47)\end{aligned}$$

Therefore, a partition of $n+1$ regions is introduced in the decision signal space by the signal log-likelihood function, each of which indexed by a k^*. Therefore, the step sizes can be of $n+1$ different sizes, according to which region the τ_n belongs to.

B. Poisson signals

Let us now consider a system where agents receive discrete but infinite signals $s \in \{0,1,\dots\}$ that follow a Poisson distribution with parameter L_w, i.e.

$$\mathbb{P}_w\{S_n=s\} = \frac{(L_w)^s}{n!}e^{-L_w} \tag{48}$$

Let us assume without lack of generality that $L_1 > L_0$. Then, the signal log-likelihood is given by

$$\Lambda_S(s) = s\log\frac{L_1}{L_0} + L_0 - L_1 \ , \tag{49}$$

where $\lfloor s \rfloor$ is the greatest integer that is smaller that s. Interestingly, as $s \geq 0$ then the private beliefs are bounded from bellow by $L_s = L_0 - L_1$ but not from above.

Therefore, one can find that

$$\Lambda_S^{-1}(l) = \left\lceil \frac{l + L_1 - L_0}{\log \frac{L_1}{L_0}} \right\rceil \tag{50}$$

Then, using XX and XX and XX one can find that

$$\mathbb{P}\left\{X_n = 0 | W = w, \boldsymbol{X}^{n-1}\right\} = \begin{cases} 0 & \text{if } \tau_n^{\text{social}} < L_S, \\ \frac{\Gamma(k^*, L_1)}{(k^*)!} & \text{otherwise,} \end{cases}$$

where $\Gamma(x, L) = \int_0^x u^{L-1} e^{-u} \mathrm{d}u$ is the incomplete Gamma function and

$$k^*(\tau_n) = \left\lceil (\tau_n + L_1 - L_0) \left(\log \frac{L_1}{L_0}\right)^{-1} \right\rceil \tag{51}$$

Finally, then log-likelihood can be expressed as

$$\begin{aligned} \Lambda_{X_n|\boldsymbol{X}^{n-1}}(X_n|\boldsymbol{X}^{n-1}) =& X_1 \log \frac{k^* - \Gamma(k^*, L_1)}{k^* - \Gamma(k^*, L_0)} \\ &+ (1 - X_n) \log \frac{\Gamma(k^*, L_1)}{\Gamma(k^*, L_0)} \ . \end{aligned}$$

As in the case of Binomial distributions, we see that the step sizes are determined according to k^* which introduces a numerable partition in the semiplane given by $[L_0 - L_1, \infty)$.

C. Guassian signals

Let us assume that, for given $W = w$, S_n are continuous real variables which are absolutely continuous with respect to the Lebesgue measure, i.e. their statistics can be described using a p.d.f. $h_w(s)$. Then, the log-likelihood ratio can be expressed as

$$\Lambda_S(s) = \log \frac{h_1(s)}{h_0(s)} \ . \tag{52}$$

As a particular case, we will study Gaussian channels where, for given $W = w$, S_n distributes as a Gaussian random variable with mean value m_w and variance σ that does not dependent on $W = w$. Without loose of generality, let us assume that $m_1 = -m_0 := m$. Then, a direct computation shows that for this case the log-likelihood is linear, as shown by a direct computation:

$$\begin{aligned} \Lambda_S(s) &= \frac{1}{2\sigma^2} \left[(s+m)^2 - (s-m)^2\right] & (53) \\ &= \frac{2m}{\sigma^2} s \ . & (54) \end{aligned}$$

This shows that Gaussian signals provide non-bounded beliefs, as a large signal can provided a arbitrarly strong evidence in favor of any of the two states of the world. Note that this also shows that the log-likelihood is an increasing function, and its inverse can be expressed as

$$\Lambda_S^{-1}(x) = \frac{\sigma^2}{2m} x \ . \tag{55}$$

Moreover, it is useful to recall that the c.d.f. of Gaussian variables can be written in terms of the Q-function as $H_w(s) = 1 - Q(\frac{s - m_w}{\sigma})$, where $Q(s) = 1/(2\pi) \int_s^\infty e^{-u^2/2} \mathrm{d}u$. Hence, one can find that

$$H_w(\Lambda_S^{-1}(x)) = H_w\left(\frac{\sigma^2}{2m} x\right) = 1 - Q\left(\frac{\sigma}{2m} x + (-1)^w \frac{m}{\sigma}\right).$$

Using this and (23), one can find that

$$\mathbb{P}\left\{X_n = 1 | W = w, \boldsymbol{X}^{n-1}\right\} = Q\left(\frac{\sigma}{2m} \tau_n^{\text{social}} + (-1)^w \frac{m}{\sigma}\right).$$

Finally, noting that $1 - Q(x) = Q(-x)$, the conditional log-likelihood functions can be written as

$$\begin{aligned} \Lambda_{X_1}(X_1) =& X_1 \log \frac{Q\left(\frac{\sigma(\nu+\eta)}{2m} - \frac{m}{\sigma}\right)}{Q\left(\frac{\sigma(\nu+\eta)}{2m} + \frac{m}{\sigma}\right)} & (56) \\ &+ (1 - X_1) \log \frac{Q\left(-\frac{\sigma(\nu+\eta)}{2m} + \frac{m}{\sigma}\right)}{Q\left(-\frac{\sigma(\nu+\eta)}{2m} - \frac{m}{\sigma}\right)} & (57) \end{aligned}$$

and

$$\begin{aligned} \Lambda_{X_{n+1}|\boldsymbol{X}^n}(X_{n+1}|\boldsymbol{X}^n) = & X_{n+1} \log \frac{Q\left(\frac{\sigma \tau_{n+1}^{\text{social}}}{2m} - \frac{m}{\sigma}\right)}{Q\left(\frac{\sigma \tau_{n+1}^{\text{social}}}{2m} + \frac{m}{\sigma}\right)} \\ &+ (1 - X_{n+1}) \log \frac{Q\left(-\frac{\sigma \tau_{n+1}^{\text{social}}}{2m} + \frac{m}{\sigma}\right)}{Q\left(-\frac{\sigma \tau_{n+1}^{\text{social}}}{2m} - \frac{m}{\sigma}\right)} \ . \end{aligned}$$

D. Cauchy signals

Let us consider now the case in which agents have access to Cauchy signals, whose p.d.f. is given by

$$h_w(s) = \frac{1}{\pi\gamma \left[1 + \left(\frac{s - m_w}{\gamma}\right)^2\right]} \tag{58}$$

where m_w and γ are the location and shape parameters. As in the case of Gaussian signals, we will assume that $m_1 = -m_0 := m$ while γ do not depend on W. With this, a direct calculation gives that

$$\Lambda_S(s) = \log \frac{\gamma^2 + (s+m)^2}{\gamma^2 + (s-m)^2} \ . \tag{59}$$

By inverting the equation $\Lambda_S(s) = l$, one finds the following second order polinomial

$$s^2 + 2\frac{1 + e^l}{1 - e^l} m s + \gamma^2 + m^2 = 0 \ , \tag{60}$$

whose solution is given by

$$s_\pm(l) = -m \frac{1 + e^l}{1 - e^l} \pm \sqrt{\frac{4m^2 e^l}{(1 - e^l)^2} - \gamma^2} \ . \tag{61}$$

Therefore, for each $l \in \mathbb{R} - \{0\}$ there exists two signals such that $\Lambda_s(s_\pm) = l$, which are the ones given in (61) (see Figure 15). With this and (62), one can find that

$$F_w^\Lambda(l) = \begin{cases} H_w(s_+(l)) - H_w(s_-(l)) & \text{if } l \leq 0 \\ H_w(s_-(l)) + 1 - H_w(s_+(l)) & \text{if } l > 0, \end{cases} \tag{62}$$

where $H_w(s) = \frac{1}{\pi}\arctan\left(\frac{s+(-1)^w m}{\gamma}\right) + \frac{1}{2}$ is the c.d.f. of a Cauchy distribution. Using this and (23), one finds that

$$\begin{aligned}
&\mathbb{P}_w\{X_{n+1} = 1|\boldsymbol{X}^n\} \\
&= \begin{cases} 1 - H_w(s_+(\boldsymbol{X}^n)) + H_w(s_-(\boldsymbol{X}^n)) & \text{if } \tau_n(\boldsymbol{X}^n) \le 0 \\ H_w(s_+(\boldsymbol{X}^n)) - H_w(s_-(\boldsymbol{X}^n)) & \text{if } \tau_n(\boldsymbol{X}^n) > 0, \end{cases} \\
&= \begin{cases} 1 - \frac{1}{\pi}\left[\arctan\left(\frac{c_+}{\gamma}\right) - \arctan\left(\frac{c_-}{\gamma}\right)\right] & \text{if } \tau_n(\boldsymbol{X}^n) \le 0 \\ \frac{1}{\pi}\left[\arctan\left(\frac{c_+}{\gamma}\right) - \arctan\left(\frac{c_-}{\gamma}\right)\right] & \text{if } \tau_n(\boldsymbol{X}^n) > 0. \end{cases}
\end{aligned}$$

Above, the last equality introduces the shorthand notation $c_+ = s_+(\boldsymbol{X}^n) + (-1)^w m$ and $c_- = s_-(\boldsymbol{X}^n) + (-1)^w m$.

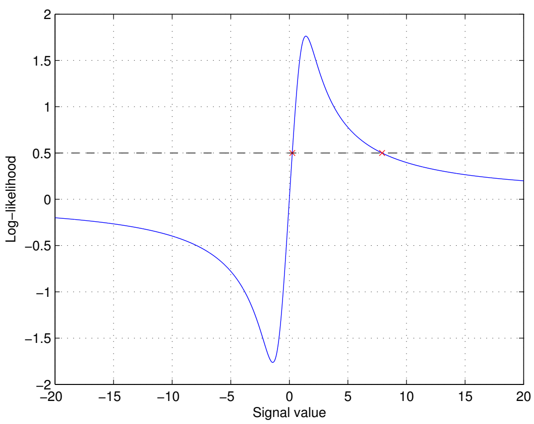


Fig. 15: In contranst to the Gaussian case, Cauchy signals provide bounded beliefs. This implies that a very high signal in the Gaussian case is a clear indicator that $w = 1$, while in the Cauchy case such a high signal is not really informative. Moreover, the inverse is no 1-1 but 1-2. Red asterisks show s_+ and s_- as given in (61) for $l = 1/2$.

APPENDIX D
PROOFS OF SECTION VII

Proof of Proposition 4: Let us first notice that a direct calculation using Bayes rule shows that

$$\mathbb{P}\{W = 1|\boldsymbol{G}_n = \boldsymbol{g}\} = \frac{1}{1 + \frac{\mathbb{P}_0\{\boldsymbol{G}_n=\boldsymbol{g}\}\mathbb{P}\{W=0\}}{\mathbb{P}_1\{\boldsymbol{G}_n=\boldsymbol{g}\}\mathbb{P}\{W=1\}}} \quad (63)$$

$$= \phi(\nu - \tau_n(\boldsymbol{g})), \quad (64)$$

where the last equality uses the definition of the logistic function and the fact that $\eta - \Lambda_{\boldsymbol{G}_n} = \nu - \tau_n$. Therefore, one can show that

$$\mathbb{P}\{W = w, \boldsymbol{G}_n = \boldsymbol{g}\} = \phi([2w-1][\nu - \tau_n])\mathbb{P}\{\boldsymbol{G}_n = \boldsymbol{g}\}$$

With this, one can re-write the average utility in a slightly different decomposition:

$$\bar{\mathcal{U}}_n(\pi_\text{b}) = \sum_{\substack{w\in\{0,1\} \\ g\in\mathcal{G}_n}} \phi([2w-1][\nu - \tau_n(\boldsymbol{g})])\mathbb{P}\{\boldsymbol{G}_n = \boldsymbol{g}\} \times \left(u_{w,0}F_w^\Lambda(\tau_n(\boldsymbol{g}_n)) + u_{w,1}\left[1 - F_w^\Lambda(\tau_n(\boldsymbol{g}_n))\right]\right).$$

Let us denote as $\mathcal{T}_n = \{t \in \mathbb{R}|\text{exists } g \in \mathcal{G}_n \text{ such that } \tau_n(g) = t\}$ the set of all values that τ_n can adopt. Then, one can re-write the above expression as

$$\bar{\mathcal{U}}_n(\pi_\text{b}) = \sum_{\substack{w\in\{0,1\} \\ t\in\mathcal{T}_n}} \phi([2w-1][\nu - t])\left(\sum_{\substack{\boldsymbol{g}\in\mathcal{G}_n \\ \tau_n(\boldsymbol{g})=t}} \mathbb{P}\{\boldsymbol{G}_n = \boldsymbol{g}\}\right) \times \left(u_{w,0}F_w^\Lambda(t) + u_{w,1}\left[1 - F_w^\Lambda(t)\right]\right) .$$

Finally, the Proposition is proven by expanding the sum over W and using (27). ■

Proof of Theorem 3: Due to the theorem's assumptions and Theorem 2, one can find that, for large values of n, $\mathcal{T}_n$ ends up being composed by numbers that are either larger than U_s or smaller than L_s. Now, by recalling that if $t > U_s$ then $F_w^\Lambda(t) = 1$ while if $t < L_s$ then $F_w^\Lambda(t) = 0$, then is clear that (28) can be re-written as

$$\bar{\mathcal{U}}_\infty(\pi_\text{b}) = \lim_{n\to\infty} \sum_{\substack{t\in\mathcal{T}_n \\ t>U_s}} \Big[u_{1,0} - (u_{1,0} - u_{0,0})\phi(t-\nu)\Big]\mathbb{P}\{\tau_n = t\} + \lim_{n\to\infty} \sum_{\substack{t\in\mathcal{T}_n \\ t<L_s}} \Big[u_{0,1} - (u_{0,1} - u_{1,1})\phi(\nu - t)\Big]\mathbb{P}\{\tau_n = t\} .$$

If $t \ge U_s$ then due to $u_{1,0} \ge u_{0,0}$ (c.f. Section IV-C) and the monotonicity of $\phi(\cdot)$ is clear that

$$(u_{1,0} - u_{0,0})\phi(t-\nu) \ge (u_{1,0} - u_{0,0})\phi(U_s - \nu) \ . \quad (65)$$

Equivalently, for the case of $t \le L_s$ one finds that $(u_{0,1} - u_{1,1})\phi(\nu - t) \ge u_{0,1}\phi(L_s) + u_{1,1}\phi(\nu - L_s)$. Using these inequalities the theorem can be proven, taking into account that

$$\mathbb{P}\{C_0\} = \lim_{n\to\infty} \sum_{\substack{t\in\mathcal{T}_n \\ t>U_s}} \mathbb{P}\{\tau_n = t\} = \lim_{n\to\infty} \mathbb{P}\{\tau_n \ge U_s\}, \quad (66)$$

$$\mathbb{P}\{C_1\} = \lim_{n\to\infty} \sum_{\substack{t\in\mathcal{T}_n \\ t<L_s}} \mathbb{P}\{\tau_n = t\} = \lim_{n\to\infty} \mathbb{P}\{\tau_n \le L_s\}. \quad (67)$$

■

APPENDIX E
COMPUTING $\mathbb{P}_w\{\boldsymbol{G}_n = \boldsymbol{g}_n\}$

For finding an expression for $\mathbb{P}_w\{\boldsymbol{G}_n = \boldsymbol{g}_n\}$, first note that one can use the distortion coefficients (c.f. Section V-D) to obtain

$$\mathbb{P}_w\{\boldsymbol{G}_n = \boldsymbol{g}_n\} = \sum_{\boldsymbol{x}^{n-1}} \alpha_n(\boldsymbol{g}_n|\boldsymbol{x}^{n-1})\mathbb{P}_w\left\{\boldsymbol{X}^{n-1} = \boldsymbol{x}^{n-1}\right\} \ .$$

Then, for computing $\mathbb{P}_w\left\{\boldsymbol{X}^{n-1} = \boldsymbol{x}^{n-1}\right\}$, note that first that $X_n = \pi_b(S_n, \boldsymbol{G}_n)$, and hence the fact that S_n and $\boldsymbol{X}^{n-1}$ are conditionally independent given $W = w$ makes $\boldsymbol{X}^{n-1} - G_n - X_n$ a Markov chain for a given $W = w$. Therefore one can show that $\mathbb{P}_w\{X_n = 0|G_n = g, \boldsymbol{X}^{n-1}\} = \mathbb{P}_w\{X_n = 0|G_n = g\}$ and therefore it can be found using (V-B) that

$$\mathbb{P}_w\left\{X_n = 0|\boldsymbol{X}^{n-1} = \boldsymbol{x}^{n-1}\right\} = \sum_{\boldsymbol{g}_n\in\mathcal{G}_n} \alpha(\boldsymbol{g}_n|\boldsymbol{x}^{n-1})F_w^\Lambda(\tau_n(\boldsymbol{g}_n)) \ .$$

Finally, using this result the distribution of a decision vector can be found using the fact that

$$\mathbb{P}_w \left\{ \boldsymbol{X}^n = \boldsymbol{x}^n \right\} = \prod_{j=1}^{n} \mathbb{P}_w \left\{ X_j = x_j | \boldsymbol{X}^{j-1} = \boldsymbol{x}^{j-1} \right\} \quad , \tag{68}$$

with the convention that $\mathbb{P}_w \left\{ X_1 = x_1 | \boldsymbol{X}^0 \right\} = \mathbb{P}_w \left\{ X_1 = x_1 \right\}$.

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] J. P. Crutchfield, "The calculi of emergence: computation, dynamics and induction," *Physica D: Nonlinear Phenomena*, vol. 75, no. 1, pp. 11–54, 1994.

[3] S. Wolfram, *A new kind of science*. Wolfram media Champaign, 2002, vol. 5.

[4] R. K. Sawyer, *Social emergence: Societies as complex systems*. Cambridge University Press, 2005.

[5] K.-C. Chen, M. Chiang, and H. V. Poor, "From technological networks to social networks." *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9-Supplement, pp. 548–572, 2013.

[6] M. Del Vicario, A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H. E. Stanley, and W. Quattrociocchi, "The spreading of misinformation online," *Proceedings of the National Academy of Sciences*, vol. 113, no. 3, pp. 554–559, 2016.

[7] F. Turner, *From counterculture to cyberculture: Stewart Brand, the Whole Earth Network, and the rise of digital utopianism*. University Of Chicago Press, 2010.

[8] C. R. Sunstein and A. Vermeule, "Conspiracy theories: Causes and cures," *Journal of Political Philosophy*, vol. 17, no. 2, pp. 202–227, 2009.

[9] R. S. Nickerson, "Confirmation bias: A ubiquitous phenomenon in many guises." *Review of general psychology*, vol. 2, no. 2, p. 175, 1998.

[10] R. K. Garrett, "Echo chambers online?: Politically motivated selective exposure among internet news users1," *Journal of Computer-Mediated Communication*, vol. 14, no. 2, pp. 265–285, 2009.

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

[12] M. Castells, *The rise of the network society: The information age: Economy, society, and culture*. John Wiley & Sons, 2011, vol. 1.

[13] "Cisco visual networking index: Global mobile data traffic forecast, 2015–2020," Cisco, Tech. Rep., February 2016.

[14] N. A. Christakis and J. H. Fowler, *Connected: The surprising power of our social networks and how they shape our lives*. Little, Brown, 2009.

[15] W. Quattrociocchi, G. Caldarelli, and A. Scala, "Opinion dynamics on interacting networks: media competition and social influence," *Scientific reports*, vol. 4, 2014.

[16] R. Kumar, M. Mahdian, and M. McGlohon, "Dynamics of conversations," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2010, pp. 553–562.

[17] F. Provost and T. Fawcett, "Data science and its relationship to big data and data-driven decision making," *Big Data*, vol. 1, no. 1, pp. 51–59, 2013.

[18] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, "Big data: The next frontier for innovation, competition, and productivity," McKinsey Global Institute, Tech. Rep., 2011.

[19] G. Weiss, *Multiagent systems: a modern approach to distributed artificial intelligence*. MIT press, 1999.

[20] M. Wooldridge, *An introduction to multiagent systems*. John Wiley & Sons, 2009.

[21] K.-D. Kim and P. R. Kumar, "Cyber–physical systems: A perspective at the centennial," *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1287–1308, 2012.

[22] J. N. Tsitsiklis *et al.*, "Decentralized detection," *Advances in Statistical Signal Processing*, vol. 2, no. 2, pp. 297–344, 1993.

[23] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors i. fundamentals," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 54–63, 1997.

[24] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors i. advanced topics," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 64–79, 1997.

[25] J. Tsitsiklis and M. Athans, "On the complexity of decentralized decision making and detection problems," *IEEE Transactions on Automatic Control*, vol. 30, no. 5, pp. 440–446, 1985.

[26] D. Warren and P. Willett, "Optimum quantization for detector fusion: some proofs, examples, and pathology," *Journal of the Franklin Institute*, vol. 336, no. 2, pp. 323–359, 1999.

[27] J.-F. Chamberland and V. V. Veeravalli, "Asymptotic results for decentralized detection in power constrained wireless sensor networks," *IEEE Journal on selected areas in communications*, vol. 22, no. 6, pp. 1007–1015, 2004.

[28] J. D. Papastavrou and M. Athans, "Distributed detection by a large team of sensors in tandem," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 3, pp. 639–653, 1992.

[29] P. F. Swaszek, "On the performance of serial networks in distributed detection," *IEEE transactions on aerospace and electronic systems*, vol. 29, no. 1, pp. 254–260, 1993.

[30] B. Chen, L. Tong, and P. K. Varshney, "Channel aware distributed detection in wireless sensor networks," in *IEEE Signal Processing Mag*. Citeseer, 2006.

[31] J.-F. Chamberland and V. V. Veeravalli, "Wireless sensors in distributed detection applications," *IEEE signal processing magazine*, vol. 24, no. 3, pp. 16–25, 2007.

[32] Z.-Q. Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," *IEEE Transactions on information theory*, vol. 51, no. 6, pp. 2210–2219, 2005.

[33] J.-J. Xiao and Z.-Q. Luo, "Decentralized estimation in an inhomogeneous sensing environment," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3564–3575, Oct 2005.

[34] E. Masazade, R. Rajagopalan, P. K. Varshney, C. K. Mohan, G. K. Sendur, and M. Keskinoz, "A multiobjective optimization approach to obtain decision thresholds for distributed detection in wireless sensor networks," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, no. 2, pp. 444–457, 2010.

[35] S. Haykin and K. R. Liu, *Handbook on array processing and sensor networks*. John Wiley & Sons, 2010, vol. 63.

[36] V. V. Veeravalli and P. K. Varshney, "Distributed inference in wireless sensor networks," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 370, no. 1958, pp. 100–117, 2012.

[37] S. Barbarossa, S. Sardellitti, and P. Di Lorenzo, "Distributed detection and estimation in wireless," *Academic Press Library in Signal Processing: Communications and Radar Signal Processing*, vol. 2, p. 329, 2013.

[38] D. Easley and J. Kleinberg, *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press, 2010.

[39] J.-H. Hsiao and K.-C. Chen, "Steering information cascades in a social system by selective rewiring and incentive seeding," in *Communications (ICC), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1–6.

[40] J. Leskovec, L. A. Adamic, and B. A. Huberman, "The dynamics of viral marketing," *ACM Transactions on the Web (TWEB)*, vol. 1, no. 1, p. 5, 2007.

[41] I. Lorge, D. Fox, J. Davitz, and M. Brenner, "A survey of studies contrasting the quality of group performance and individual performance, 1920-1957." *Psychological bulletin*, vol. 55, no. 6, p. 337, 1958.

[42] C. Hommes, J. Sonnemans, J. Tuinstra, and H. Van de Velden, "Coordination of expectations in asset pricing experiments," *Review of Financial studies*, vol. 18, no. 3, pp. 955–980, 2005.

[43] I. Yaniv and M. Milyavsky, "Using advice from multiple sources to revise and improve judgments," *Organizational Behavior and Human Decision Processes*, vol. 103, no. 1, pp. 104–120, 2007.

[44] F. Galton, "Vox populi (the wisdom of crowds)," *Nature*, vol. 75, no. 7, pp. 450–451, 1907.

[45] J. Surowiecki, *The wisdom of crowds*. Anchor, 2005.

[46] H. Oinas-Kukkonen, "Network analysis and crowds of people as sources of new organisational knowledge," *Knowledge Management: Theoretical Foundation*, pp. 173–189, 2008.

[47] H. N. Tuttle, *The crowd is untruth: The existential critique of mass society in the thought of Kierkegaard, Nietzsche, Heidegger, and Ortega y Gasset*. Peter Lang, 1996, vol. 176.

[48] M. J. Salganik, P. S. Dodds, and D. J. Watts, "Experimental study of inequality and unpredictability in an artificial cultural market," *science*, vol. 311, no. 5762, pp. 854–856, 2006.

[49] R. L. Goldstone and T. M. Gureckis, "Collective behavior," *Topics in cognitive science*, vol. 1, no. 3, pp. 412–438, 2009.

[50] J. Lorenz, H. Rauhut, F. Schweitzer, and D. Helbing, "How social influence can undermine the wisdom of crowd effect," *Proceedings of the National Academy of Sciences*, vol. 108, no. 22, pp. 9020–9025, 2011.

[51] W. A. Mason, F. R. Conrey, and E. R. Smith, "Situating social influence processes: Dynamic, multidirectional flows of influence within social networks," *Personality and social psychology review*, vol. 11, no. 3, pp. 279–300, 2007.

[52] S. E. Asch, "Opinions and social pressure," *Readings about the social animal*, vol. 193, pp. 17–26, 1955.

[53] H. J O'Gorman, "The discovery of pluralistic ignorance: An ironic lesson," *Journal of the History of the Behavioral Sciences*, vol. 22, no. 4, pp. 333–347, 1986.

[54] 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.

[55] D. Sgroi, "Optimizing information in the herd: Guinea pigs, profits, and welfare," *Games and Economic Behavior*, vol. 39, no. 1, pp. 137–166, 2002.

[56] D. Gill and D. Sgroi, "Sequential decisions with tests," *Games and economic Behavior*, vol. 63, no. 2, pp. 663–678, 2008.

[57] S. Bose, G. Orosel, M. Ottaviani, and L. Vesterlund, "Dynamic monopoly pricing and herding," *The RAND Journal of Economics*, vol. 37, no. 4, pp. 910–928, 2006.

[58] C. Carboneau, "Using diffusion of innovation and academic detaling to spread evidence-based practices," *Journal for Healthcare Quality*, vol. 27, no. 2, pp. 48–52, 2005.

[59] W. Farnsworth, *The legal analyst: A toolkit for thinking about the law*. University of Chicago Press, 2008.

[60] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins, "Information diffusion through blogspace," in *Proceedings of the 13th international conference on World Wide Web*. ACM, 2004, pp. 491–501.

[61] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic, "The role of social networks in information diffusion," in *Proceedings of the 21st international conference on World Wide Web*. ACM, 2012, pp. 519–528.

[62] D. M. Romero, B. Meeder, and J. Kleinberg, "Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter," in *Proceedings of the 20th international conference on World wide web*. ACM, 2011, pp. 695–704.

[63] J. Yang and J. Leskovec, "Modeling information diffusion in implicit networks," in *2010 IEEE International Conference on Data Mining*. IEEE, 2010, pp. 599–608.

[64] A. Anderson, D. Huttenlocher, J. Kleinberg, J. Leskovec, and M. Tiwari, "Global diffusion via cascading invitations: Structure, growth, and homophily," in *Proceedings of the 24th International Conference on World Wide Web*. ACM, 2015, pp. 66–76.

[65] J. Cheng, L. Adamic, P. A. Dow, J. M. Kleinberg, and J. Leskovec, "Can cascades be predicted?" in *Proceedings of the 23rd international conference on World wide web*. ACM, 2014, pp. 925–936.

[66] J. Cheng, L. A. Adamic, J. M. Kleinberg, and J. Leskovec, "Do cascades recur?" in *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2016, pp. 671–681.

[67] M. Newman, *Networks: an introduction*. Oxford university press, 2010.

[68] S. L. Huang and K. C. Chen, "Information cascades in social networks via dynamic system analyses," in *2015 IEEE International Conference on Communications (ICC)*, June 2015, pp. 1262–1267.

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

[70] L. R. Anderson and C. A. Holt, "Information cascades in the laboratory," *The American economic review*, pp. 847–862, 1997.

[71] J. E. Alevy, M. S. Haigh, and J. A. List, "Information cascades: Evidence from a field experiment with financial market professionals," *The Journal of Finance*, vol. 62, no. 1, pp. 151–180, 2007.

[72] S. Mori, M. Hisakado, and T. Takahashi, "Collective adoption of max–min strategy in an information cascade voting experiment," *Journal of the Physical Society of Japan*, vol. 82, no. 8, p. 084004, 2013.

[73] E. Walden and G. Browne, "Information cascades in the adoption of new technology," *ICIS 2002 Proceedings*, p. 40, 2002.

[74] A. De Vany and W. D. Walls, "Uncertainty in the movie industry: Does star power reduce the terror of the box office?" *Journal of cultural economics*, vol. 23, no. 4, pp. 285–318, 1999.

[75] R. A. Baños, J. Borge-Holthoefer, and Y. Moreno, "The role of hidden influentials in the diffusion of online information cascades," *EPJ Data Science*, vol. 2, no. 1, p. 1, 2013.

[76] Y. Wang, H. Shen, S. Liu, and X. Cheng, "Learning user-specific latent influence and susceptibility from information cascades," in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*. AAAI Press, 2015, pp. 477–483.

[77] K. Burghardt, E. F. Alsina, M. Girvan, W. M. Rand, and K. Lerman, "The myopia of crowds: A study of collective evaluation on stack exchange," *Robert H. Smith School Research Paper No. RHS*, vol. 2736568, 2016.

[78] L. E. Celis, P. M. Krafft, and N. Kobe, "Sequential voting promotes collective discovery in social recommendation systems," in *Tenth International AAAI Conference on Web and Social Media*, 2016.

[79] S. Sreenivasan, K. S. Chan, A. Swami, G. Korniss, and B. Szymanski, "Information cascades in feed-based networks of users with limited attention," *IEEE Transactions on Network Science and Engineering*, vol. PP, no. 99, pp. 1–1, 2016.

[80] N. E. Friedkin, "The problem of social control and coordination of complex systems in sociology: A look at the community cleavage problem," *IEEE Control Systems*, vol. 35, no. 3, pp. 40–51, 2015.

[81] A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec, "Steering user behavior with badges," in *Proceedings of the 22nd international conference on World Wide Web*. ACM, 2013, pp. 95–106.

[82] F. Chierichetti, J. Kleinberg, and A. Panconesi, "How to schedule a cascade in an arbitrary graph," *SIAM Journal on Computing*, vol. 43, no. 6, pp. 1906–1920, 2014.

[83] I. H. Lee, "On the convergence of informational cascades," *Journal of Economic theory*, vol. 61, no. 2, pp. 395–411, 1993.

[84] S. Bikhchandani, D. Hirshleifer, and I. Welch, "Learning from the behavior of others: Conformity, fads, and informational cascades," *The Journal of Economic Perspectives*, vol. 12, no. 3, pp. 151–170, 1998.

[85] V. Bala and S. Goyal, "Learning from neighbours," *The review of economic studies*, vol. 65, no. 3, pp. 595–621, 1998.

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

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

[88] A. Banerjee and D. Fudenberg, "Word-of-mouth learning," *Games and Economic Behavior*, vol. 46, no. 1, pp. 1–22, 2004.

[89] H. V. Poor, *An introduction to signal detection and estimation*. Springer Science & Business Media, 2013.

[90] D. Gale and S. Kariv, "Bayesian learning in social networks," *Games and Economic Behavior*, vol. 45, no. 2, pp. 329–346, 2003.

[91] 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.

[92] D. Acemoglu and A. Ozdaglar, "Opinion dynamics and learning in social networks," *Dynamic Games and Applications*, vol. 1, no. 1, pp. 3–49, 2011.

[93] M. H. DeGroot, "Reaching a consensus," *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118–121, 1974.

[94] P. M. DeMarzo, J. Zwiebel, and D. Vayanos, "Persuasion bias, social influence, and uni-dimensional opinions," *Social Influence, and Uni-Dimensional Opinions (November 2001). MIT Sloan Working Paper*, no. 4339-01, 2001.

[95] B. Golub and M. O. Jackson, "Naive learning in social networks and the wisdom of crowds," *American Economic Journal: Microeconomics*, vol. 2, no. 1, pp. 112–149, 2010.

[96] D. Acemoglu, A. Ozdaglar, and A. ParandehGheibi, "Spread of (mis) information in social networks," *Games and Economic Behavior*, vol. 70, no. 2, pp. 194–227, 2010.

[97] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi, "Non-bayesian social learning," *Games and Economic Behavior*, vol. 76, no. 1, pp. 210–225, 2012.

[98] J. Lamperti, *Probability: a survey of the mathematical theory*, 2nd ed. Wiley, 1966.

[99] R. J. Shiller, "Conversation, information, and herd behavior," *The American Economic Review*, vol. 85, no. 2, pp. 181–185, 1995.

[100] A. Lalitha, A. Sarwate, and T. Javidi, "Social learning and distributed hypothesis testing," in *2014 IEEE International Symposium on Information Theory*. IEEE, 2014, pp. 551–555.

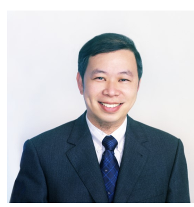
[101] J. B. Rhim and V. K. Goyal, "Distributed hypothesis testing with social learning and symmetric fusion," *IEEE Transactions on Signal Processing*, vol. 62, no. 23, pp. 6298–6308, 2014.

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

[103] J. B. Rhim and V. K. Goyal, "Social teaching: Being informative vs. being right in sequential decision making," in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*. IEEE, 2013, pp. 2602–2606.
[104] M. Loeve, *Probability Theory I*. Springer, 1978.
[105] B. Sklar, *Digital communications: fundamentals and applications*, 2nd ed. Prentice Hall, 2001.
[106] J. Dieudonne, *Treatise on Analysis*. Associated Press, New York, 1976, vol. II.
[107] W. Feller, *An introduction to probability theory and its applications: volume I*, 3rd ed. John Wiley & Sons, Inc., 1968.
[108] F. Rosas and K. C. Chen, "Social learning against data falsification in sensor networks," *IEEE Signal Processing Letters*, under revision 2016.
[109] E. Shi and A. Perrig, "Designing secure sensor networks," *IEEE Wireless Communications*, vol. 11, no. 6, pp. 38–43, 2004.
[110] W. Trappe, R. Howard, and R. S. Moore, "Low-energy security: Limits and opportunities in the internet of things," *IEEE Security Privacy*, vol. 13, no. 1, pp. 14–21, Jan 2015.
[111] L. Lamport, R. Shostak, and M. Pease, "The byzantine generals problem," *ACM Transactions on Programming Languages and Systems (TOPLAS)*, vol. 4, no. 3, pp. 382–401, 1982.
[112] K.-C. Chen and S.-Y. Lien, "Machine-to-machine communications: Technologies and challenges," *Ad Hoc Networks*, vol. 18, pp. 3–23, 2014.
[113] P. Ramuhalli, M. Halappanavar, J. Coble, and M. Dixit, "Towards a theory of autonomous reconstitution of compromised cyber-systems," in *Technologies for Homeland Security (HST), 2013 IEEE International Conference on*. IEEE, 2013, pp. 577–583.
[114] R. Rajagopalan and P. K. Varshney, "Data-aggregation techniques in sensor networks: A survey," *Commun. Surveys Tuts.*, vol. 8, no. 4, pp. 48–63, Oct. 2006.
[115] A. Vempaty, L. Tong, and P. K. Varshney, "Distributed inference with byzantine data: State-of-the-art review on data falsification attacks," *IEEE Signal Processing Magazine*, vol. 30, no. 5, pp. 65–75, 2013.
[116] M. O. Jackson *et al.*, *Social and economic networks*. Princeton university press Princeton, 2008, vol. 3.
[117] J. Ferber, *Multi-agent systems: an introduction to distributed artificial intelligence*. Addison-Wesley Reading, 1999, vol. 1.
[118] R. Olfati-Saber, "Flocking for multi-agent dynamic systems: Algorithms and theory," *IEEE Transactions on automatic control*, vol. 51, no. 3, pp. 401–420, 2006.
[119] R. Hegselmann, U. Krause *et al.*, "Opinion dynamics and bounded confidence models, analysis, and simulation," *Journal of Artificial Societies and Social Simulation*, vol. 5, no. 3, 2002.
[120] ——, "Truth and cognitive division of labour: First steps towards a computer aided social epistemology," *Journal of Artificial Societies and Social Simulation*, vol. 9, no. 3, p. 10, 2006.
[121] Y. S. Abu-Mostafa, M. Magdon-Ismail, and H.-T. Lin, *Learning from data*. AMLBook Singapore, 2012, vol. 4.
[122] P. Domingos, "A few useful things to know about machine learning," *Communications of the ACM*, vol. 55, no. 10, pp. 78–87, 2012.
[123] M.-F. Balcan, A. Blum, J. D. Hartline, and Y. Mansour, "Mechanism design via machine learning," in *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)*. IEEE, 2005, pp. 605–614.
[124] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, *Algorithmic game theory*. Cambridge University Press Cambridge, 2007, vol. 1.

Fernando Rosas received the B.A. degree in music composition and philosophy (Minor), the B.Sc. degree in mathematics, and the M.S. and Ph.D. degree in engineering sciences from the Pontificia Universidad Catlica de Chile, Santiago, Chile, in 2003, 2007, and 2012, respectively. He is currently working as a Postdoctoral Researcher with the Graduate Institute of Communication Engineering, at the National Taiwan University. His research interests lay in the interface between communication theory and complexity science. He was the recipient of the Marie Slodowska-Curie Individual Fellowship from the European Union in 2016, the F+ Scholarship from the KU Leuven University in 2013, and the CONICYT Doctoral Fellowship from the Chilean Ministry of Education in 2008.

Jen-Hao Hsiao received B.S. degree in department of electrical engineering from National Taiwan University, Taipei, in 2015. He is currently working toward his M.S. degree in the Graduate Institute of Communication Engineering, National Taiwan University, Taipei. His research interests include social networks, information dynamics, and collective behavior of interactive multi-agents.



Kwang-Cheng Chen (M89-SM94-F07) received the B.S. from the National Taiwan University in 1983, and the M.S. and Ph.D from the University of Maryland, College Park, United States, in 1987 and 1989, all in electrical engineering. From 1987 to 1998, Dr. Chen worked with SSE, COMSAT, IBM Thomas J. Watson Research Center, and National Tsing Hua University, working on the mobile communications and networks. From 1998 to 2016, he was a Distinguished Professor with the National Taiwan University, Taipei, Taiwan, ROC, also served as the Director, Graduate Institute of Communication Engineering, Director, Communication Research Center, and the Associate Dean for Academic Affairs, College of Electrical Engineering and Computer Science during 2009-2015. Since 2016, Dr. Chen has been with the Department of Electrical Engineering, University of South Florida, Tampa, United States. He has been actively involving in the organization of various IEEE conferences as General/TPC chair/co-chair, and has served in editorships with a few IEEE journals. Dr. Chen also actively participates in and has contributed essential technology to various IEEE 802, Bluetooth, and LTE and LTE-A wireless standards. Dr. Chen is an IEEE Fellow and has received a number of awards including the *2011 IEEE COMSOC WTC Recognition Award*, *2014 IEEE Jack Neubauer Memorial Award* and *2014 IEEE COMSOC AP Outstanding Paper Award.* His recent research interests include wireless networks, cybersecurity, cyber-physical systems, social networks and data analytics.