Teaching and Learning in Uncertainty

Varun Jog and Po-Ling Loh

***Abstract*— We investigate a simple model for social learning with two agents: a teacher and a student. The teacher's goal is to teach the student the state of the world; however, the teacher himself is not certain about the state of the world and needs to simultaneously learn this parameter and teach it to the student. We model the teacher's and student's uncertainties via noisy transmission channels, and employ two simple decoding strategies for the student. We focus on two teaching strategies: a "low-effort" strategy of simply forwarding information, and a "high-effort" strategy of communicating the teacher's current best estimate of the world at each time instant, based on his own cumulative learning. Using tools from large deviation theory, we calculate the exact learning rates for these strategies and demonstrate regimes where the low-effort strategy outperforms the high-effort strategy. Finally, we present a conjecture concerning the optimal learning rate for the student over all joint strategies between the student and the teacher.**

***Index Terms*— Large deviations theory, random walks, social learning.**

I. Introduction

INDIVIDUALS in a society may learn about their environment directly through their own experiences, or vicariously via communication with other members of the society. Such interactions drive the exchange of ideas, technology, news, and opinions, and are critically important to the social and economic development of a society. However, understanding and predicting the effects of social interaction on society is a hard problem: each individual's opinion is dynamic and depends on his or her own biases, observations, and social interactions. The theoretical question of how agents learn through social interactions has consequently received much attention in the past few decades, and a number of mathematical models have been proposed to analyze social learning phenomena, such as those detailed in Chamley [1] and Mossel and Tamuz [2].

Social learning models generally comprise an unknown state of the world and a number of agents. These agents have private observations of the state and use it to take actions to achieve a certain goal, such as maximizing their utility functions. Often, agents are able to observe the actions of some or all of the other agents, which they can use to glean more information about the state of the world and thereby play better actions. Broadly speaking, one is interested in analyzing the following questions in these models: (a) Convergence: Do the agents' actions eventually converge? (b) Agreement: Given convergence, do the agents agree? (c) Learning: Given agreement, is the unanimous opinion the true state of the world? and (d) Given learning, how fast does learning take place?

In this paper, we focus on analyzing the *rate of learning* posed in question (d) for a simple model with two agents. Our work is most closely related to the work of Vives [3], Jadbabaie *et al.* [4], [5], and Harel *et al.* [6], which also investigate learning rates in different social learning models. The specific learning model we consider is motivated by the work of Harel *et al.* [6], which we will describe in detail in Section II and contrast with our model.

A brief overview of our social learning model is as follows: The unknown state of the world Θ is drawn uniformly from $\{-1, +1\}$. The first agent, whom we call the teacher, receives repeated observations of Θ through a binary symmetric channel with flipping probability p. At each time instant, the teacher can transmit one bit over another binary symmetric channel with flipping probability q to the second agent, whom we call the student. The teacher's goal is to ensure that the student learns the state of the world. Our goal is to understand how the rate of learning of the student depends on the joint strategies of the teacher and the student, where we assume that both the teacher and student are aware of each other's strategies. We analyze two particular student strategies: (i) A strategy where the student simply takes an average (or majority vote, when the state of the world is binary) over all observations received from the teacher; and (ii) a strategy where the student—shrewdly cognizant of the fact that the teacher is also learning from his own observations over time—only averages a final segment of observations, which she assumes to more accurately reflect the state of the world than the initial observations.

The main mathematical tools we use in this paper are derived from the theory of large deviations, specifically concerning the large deviation properties of the sign of a transient Markov random walk on $\mathbb{Z}$. We show that the rate function of this process can be explicitly calculated; moreover, it has a surprisingly neat closed-form expression that can be used in our subsequent analysis. When the state of the world is binary, our results show that when the student employs either of the two learning strategies discussed above, the relative ordering of teacher strategies in terms of the learning rate of the student generally depends on the noise parameters of the teacher's

Manuscript received May 25, 2019; revised April 9, 2020; accepted October 5, 2020. Date of publication October 13, 2020; date of current version December 21, 2020. This work was supported by the NSF under Grant CCF-1841190. The work of Po-Ling Loh was supported by the NSF under Grant DMS-1749857. This article was presented in part at the 2019 IEEE International Symposium on Information Theory (ISIT). *(Corresponding author: Varun Jog.)*

Varun Jog is with the Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53706 USA (e-mail: vjog@wisc.edu).

Po-Ling Loh is with the Department of Statistics, University of Wisconsin-Madison, Madison, WI 53706 USA (e-mail: ploh@stat.wisc.edu).

Communicated by E. Abbe, Associate Editor for Machine Learning.

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2020.3030842

and student's channels, in a manner we can make explicit. We also analyze a setting where the state of the world is a continuous parameter and the learning rate is quantified by the variance of the student's estimator, and reach a markedly different conclusion that the laziest strategy of the teacher already leads to an optimal rate of learning for the student.

The remainder of the paper is organized as follows: In Section II, we describe our model in detail and discuss the strategies employed by the teacher and student. In Section III, we develop the main technical tools used in analyzing the learning rate of the student when the state of the world is binary; the rates of learning for the two student strategies are subsequently analyzed in Sections IV and V. In Section VI, we discuss the somewhat different case of Gaussian learning with continuous-valued parameters. Finally, we conclude the paper in Section VII with a discussion of open problems and teaching philosophy.

Notation: For a real parameter $p \in [0,1]$, we write $\bar{p}$ to denote the quantity $1-p$. For a random variable X_n, a set A, and a function $f(A)$, we write $\mathbb{P}(X_n \in A) \approx e^{-nf(A)}$ to mean that $\mathbb{P}(X_n \in A) = e^{-n(f(A)+o(1))}$. We write $D(a||b) = a\log(a/b) + \bar{a}\log(\bar{a}/\bar{b})$ to denote the Kullback-Leibler divergence between the Bernoulli(a) and Bernoulli(b) distributions. We write $\mathbf{1} \in \mathbb{R}^n$ to denote the n-dimensional vector of all 1's.

II. Problem Statement and Related Work

We consider a simple model of social learning with two agents: a teacher and a student. Suppose both agents are trying to learn an unknown binary random variable Θ, which is called the state of the world. We assume that Θ takes values in the set $\{-1,+1\}$, uniformly at random. At each time $i \geq 1$, the teacher observes a noisy version of Θ through a binary symmetric channel with parameter $p \in [0,1/2)$; i.e.,

$$\mathbb{P}(Y_i = \Theta) = 1-p, \quad \text{and} \quad \mathbb{P}(Y_i = -\Theta) = p.$$

Conditioned on Θ, the random observations $\{Y_i\}_{i\geq 1}$ are independent and identically distributed, as above. The student does not make any direct observations (noisy or otherwise) of Θ, and may only learn its value from the teacher.

At each time i, the teacher communicates a binary random variable $\hat{X}_i$, which is a (possibly random) function of the history of observations $\{Y_j\}_{1\leq j\leq i}$, and the student receives a noisy version of $\hat{X}_i$, which we call Z_i. The communication channel from the teacher to the student is assumed to be a binary symmetric channel with parameter $q \in [0,1/2)$. The student's estimate of Θ after observing $\{Z_j\}_{j\leq i}$ is denoted by $\hat{\Theta}_i \in \{-1,+1\}$. We refer to the sequence of random variables $\{\hat{X}_i\}$ as the teacher's strategy, and the decoding rules $\{\hat{\Theta}_i\}$ as the student's learning strategy. For fixed teaching and learning strategies, the student's rate of learning is defined as follows:

$$\mathcal{R} = \limsup_{n\to\infty} \left\{ -\frac{1}{n} \log \mathbb{P}\left(\hat{\Theta}_n \neq \Theta\right) \right\}. \tag{1}$$

Notice that the teacher is guaranteed to the learn the state of the world eventually, owing to his repeated noisy observations of Θ.

Comparison to Harel et al. [6]: Despite the differences between our model and that of Harel *et al.* [6], we also uncover certain counterintuitive phenomena in our two-agent setting. The model in Harel *et al.* [6][1] also considers two agents A and B and an unknown binary state of the world Θ. The agents receive i.i.d. observations $\{A_i\}_{i\geq 1}$ and $\{B_i\}_{i\geq 1}$ of Θ through their respective binary symmetric channels. At each time i, the agents also form their binary-valued best estimates $\widehat{\theta}_A$ and $\widehat{\theta}_B$ of Θ. The information available to the agents is modeled in two settings: (i) A can observe B's estimates of Θ, but not vice versa; and (ii) A and B can both observe each other's estimates. The authors made the surprising observation that agent A's learning rate is higher in setting (i) than in (ii)—contrary to the intuition that setting (ii) involves a greater exchange of information, so one might expect A to learn faster. The authors attribute this counterintuitive result to a phenomenon they call "rational groupthink."

The model studied in our paper differs from that of Harel *et al.* [6] in three key ways: First, the second agent, the student, does not have any private observations that allow her to learn. Any information she receives about the state of the world follows from a noisy interaction with the teacher. Second, the agents are not necessarily Bayesian, but instead perform heuristic calculations to form their opinions. Rich bodies of literature studying both Bayesian and non-Bayesian models of social learning exist, which we shall describe in more detail in Section II-D1. Third, the model proposed in Harel *et al.* [6] involves *pure information externalities*, where each agent receives a payoff which depends only on their action and the state of the world. In contrast, our model does not include payoff functions for agents; rather, the teacher and student are jointly working to optimize the student's learning rate. Similar features appear in team decision theory and control, which we briefly comment on in relation to our setting in Section II-D2.

Despite the differences between our model and that of Harel *et al.* [6], we also uncover certain counterintuitive phenomena in our two-agent setting. In particular, we observe that "helpful" social interactions, where the teacher always tries to transmit his best guess to the student, may actually hinder the student's rate of learning.

In Section VI below, we will introduce an alternative model for teacher/student learning when the state of the world is a continuous real number. The rate of learning of the student will then be quantified by the variance of the student's estimate, rather than the probability of error. We will introduce the appropriate terminology later; in the remainder of this section and throughout Sections IV and V, we will adopt the setting and notations for the binary model described above.

A. Student Strategies

The student's learning rate depends on both the teacher's strategy and her own decoding strategy. When the student is aware of the teacher's strategy, the optimal learning rate

[1]See version 1 of the arXiv manuscript. In later versions, the model was extended to more than two agents, but the key ingredients of the analyses may be found in the analysis of the two agent model.

is achieved when the student uses a maximum likelihood decoder to arrive at her estimate of $\hat{\Theta}_n$. However, as is well-documented in the literature on social learning, a fully rational model often places unreasonable computational demands on Bayesian agents [2]. Assuming that agents are non-Bayesian serves two goals: it makes the model more realistic by reducing its complexity; and in some cases, it also helps make the model mathematically tractable. In this paper, we will consider two simple non-Bayesian student strategies, which we now describe.

Majority rule: This is perhaps the simplest possible strategy for the student, defined as a majority vote over her observations:

$$\hat{\Theta}_n = \begin{cases} +1 & \text{if } \frac{\sum_{i=1}^n \mathbb{1}(Z_i=+1)}{n} \geq \frac{1}{2}, \\ -1 & \text{otherwise.} \end{cases}$$

Note that in some cases (e.g., when the teacher uses the simple forwarding strategy defined in the next section), the majority strategy for the student corresponds to the MLE; however, this is not generally true when the teacher employs other strategies.

ϵ-majority rule: This is a generalization of the majority learning rule, where the student's estimate is a majority vote among the latter ϵn observations, for a parameter $\epsilon \in (0, 1]$:

$$\hat{\Theta}_n = \begin{cases} +1 & \text{if } \frac{\sum_{i=\lceil (1-\epsilon)n \rceil}^n \mathbb{1}(Z_i=+1)}{\lfloor \epsilon n \rfloor} \geq \frac{1}{2}, \\ -1 & \text{otherwise.} \end{cases}$$

The rationale is that the student is aware that the teacher is learning as time progresses, so she may be skeptical of the teacher's initial transmissions and prefer to place more weight on the most recent observations. However, since analyzing the learning rate of the student becomes rather complicated for arbitrary weighting strategies, we will restrict our analysis to a strategy that places zero weight on the first $(1-\epsilon)n$ observations and equally weights the remaining observations.

Note that the majority rule is a special case of the ϵ-majority strategy when $\epsilon = 1$. However, as our analysis will reveal, the optimal choice of ϵ depends in a nontrivial manner on the channel parameters p and q. Thus, we will generally be interested in the behavior of ϵ-majority learning when ϵ is optimized to the parameters p and q, operating under the assumption that such a strategy would only be employed if the student had some knowledge of the channel parameters. We will also analyze the majority learning strategy on its own.

B. Teacher Strategies

We now turn to describing several teaching strategies that we will analyze in our paper.

Simple forwarding: A "lazy" strategy for the teacher that requires no learning on his part is to put $\hat{X}_i = Y_i$; i.e., simply forward his observation at each time step directly to the student.

Cumulative teaching: In contrast to the lazy strategy of simply forwarding information, the teacher might follow a strategy of always transmitting his current best estimate of Θ, obtained by applying the majority rule to his observations $\{Y_i\}_{1 \leq i \leq n}$.

Note that the cumulative teaching strategy clearly satisfies the property that after some finite time, the process $\{\hat{X}_n\}$ is identically equal to Θ. In contrast, the simple forwarding strategy never converges in this manner, so the teacher is correct more often in the cumulative teaching strategy if n is sufficiently large. This is the intuitive reason why one might expect the cumulative teaching strategy to dominate the simple forwarding strategy; however, as we will see later, the relative merits of the two strategies are intricately linked to the values of p and q.

ϵ-teaching: We will also analyze a teaching strategy that is tailored to the ϵ-majority student strategy. In this strategy, the teacher transmits no information during the first $(1-\epsilon)n$ time steps (e.g., always transmitting a default value of 1), and then repeatedly transmits his best guess based on the first $(1-\epsilon)n$ observations in the remaining time steps:

$$\hat{X}_i = \begin{cases} +1 & \text{if } \frac{\sum_{i=1}^{\lfloor (1-\epsilon)n \rfloor} \mathbb{1}(Y_i=+1)}{\lfloor (1-\epsilon)n \rfloor} \geq \frac{1}{2}, \\ -1 & \text{otherwise,} \end{cases} \quad \text{for } i \geq \lceil (1-\epsilon)n \rceil.$$

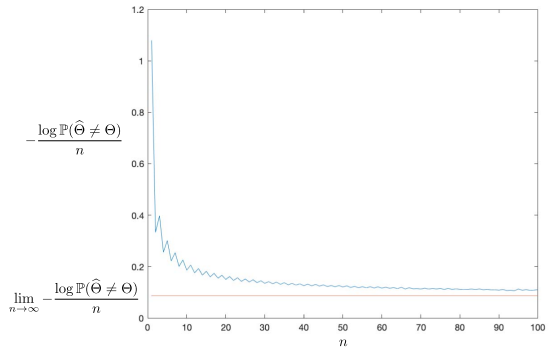
Evidently, the ϵ-teaching strategy could potentially be improved if the teacher transmitted more information in the first $(1-\epsilon)n$ time steps, or transmitted a more sophisticated estimator based on cumulative learning in the last ϵn time steps. However, we again focus on analyzing this simpler strategy since it leads to closed-form expressions for learning rates that may be more easily compared to other student/teacher strategies.

Remark: The definition of learning rate in equation (1) is motivated, at least in part, for reasons of mathematical tractability. It is natural to wonder how large n should be so that the probability of error is close to the approximation resulting from an asymptotic calculation. Calculating the exact probability of error for large n is not computationally feasible for $n > 20$; however, the error probability can be approximated using Monte Carlo simulations for moderate values of n. Based on our experiments for the "cumulative teaching + majority learning" strategy, we observe that although the values of p and q generally affect the outcome, in most cases, the asymptotic approximation becomes reasonably accurate when $n \approx 100$. Figure 1 displays plots for the error probability vs. n when $(p, q) = (0.1, 0.3)$ and $(p, q) = (0.3, 0.1)$.

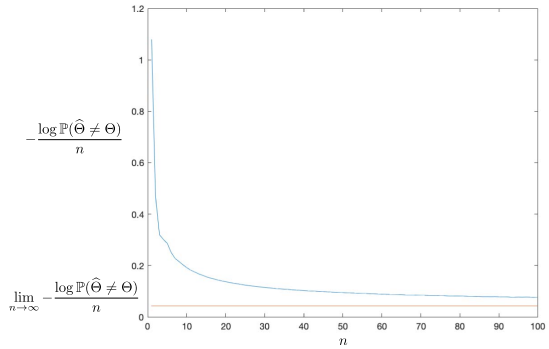
C. Overview of Results

To aid readability, we now provide a preview of our main results. As mentioned earlier, we will focus on characterizing the learning rate of different strategies when the student's strategy is fixed. The main message of our paper is that certain teacher strategies are better than other strategies for particular regimes of the channel parameters (p, q):

- When the student employs the majority rule strategy, neither the simple forwarding strategy nor the cumulative teaching strategy strictly dominates the other. In Section IV, we calculate the analytical expressions for the learning rates in both cases. Comparing them for different values of p and q, we observe that for small values of q (which one may interpret as sharp and attentive students), the simple forwarding strategy



(a) $p = 0.1$ and $q = 0.3$



(b) $p = 0.3$ and $q = 0.1$

Fig. 1. Plots showing the convergence of the non-asymptotic error probability of the student in the "cumulative teaching + majority learning" strategy to the exact learning rate defined by equation (1).

dominates. However, we also note that for all q larger than ≈ 0.15, the simple forwarding strategy is worse than the cumulative teaching strategy for all values of p; i.e., no matter how "bad" the teacher (corresponding to large values of p), a moderately attentive student still benefits from a cumulative teaching strategy.

- When the student employs the ϵ-majority strategy, neither the simple forwarding strategy nor the cumulative teaching strategy strictly dominates the other. However, we can show that the cumulative teaching strategy uniformly dominates the ϵ-teaching strategy for all values of p and q. We can obtain closed-form expressions for the learning rate of the latter teaching strategy; for the former teaching strategy, we obtain an expression that can be computed by solving an optimization problem using computer software. Comparing the cumulative teaching strategy with the simple forwarding strategy, we observe that the cumulative teaching + ϵ-majority learning strategy dominates the simple forwarding + majority learning strategy for almost all values of p and q, except when q is very small.

In the case of Gaussian teaching and learning, we show that rather surprisingly, the majority learning rule for the student and simple forwarding strategy for the teacher is jointly dominant. Notably, neither of these teaching or learning strategies requires knowledge of the Gaussian channel parameters. This underscores a fundamental difference between the social learning problem in discrete vs. continuous state spaces.

D. Related Literature

The problem described in the section above possesses additional connections to various research threads spanning a diverse range of topics. We briefly highlight some of these below.

1) Social Learning and Signaling Games: The economics literature contains a vast body of work concerning social learning. We refer the reader to the survey articles by Mossel and Tamuz [2] and Mobius and Rosenblat [7] for a broad coverage, and only highlight a subset of this literature that is most relevant to our work.

Aumann's Agreement Theorem [8] may be viewed as an early example of social learning. Aumann showed that if two agents agree on the prior of state of the world and their posteriors are "common knowledge," their posteriors must be identical; i.e., they cannot agree to disagree. Geanakoplos and Polemarchakis [9] studied the speed of convergence of the agents' posteriors when they exchange and update their posteriors at each time step. Social learning situations where agents update their actions repeatedly—based on their private signals and the (continuously updating) actions of other agents—have been studied in numerous popular models in economics. Banerjee [10], Bikhchandani *et al.* [11], and Smith and Sorenson [12] proposed models of social learning which exhibit the phenomenon of herding, where agents "herd" to the same (possibly wrong) action.

Gale and Kariv [13] proposed a social network model, wherein agents occupy vertices of a graph and are able to observe the actions of their neighbors. Akin to Aumann's result [8], Gale and Kariv showed that fully Bayesian (or fully rational) agents converge to the same action under suitable conditions. Mossel *et al.* [14] studied learning in this model by incorporating a state of the world that dictates private signals of the agents. The model studied in our paper may be considered as operating on a social network graph with one edge: the student can noisily observe the teacher's actions. Nonetheless, the information structure is somewhat different in our model, which is more closely related to the model proposed in Harel *et al.* [6].

Another notable aspect of our model is that the two agents are non-Bayesian. Learning models with non-Bayesian agents have been studied by various authors, e.g., DeGroot [15], Ellison and Fudenberg [16], Bala and Goyal [17], Rahimian and Jadbabaie [18], Hazła, Jadbabaie, Mossel, and Rahimian [19] and Molavi *et al.* [5]. Non-Bayesian models are relevant for two reasons: (1) humans are not Bayesian; and (2) it can be challenging to analyze the behavior of Bayesian agents. This is remarked upon in Gale and Kariv [13] in the study of a graph with just three nodes and also explored further in Kanoria and Tamuz [20]. Consequently, we have chosen to consider a model in which agents make heuristic rather than Bayesian decisions.

2) Team Decision Theory and Control: Team decision theory, as described in Ho [21], considers agents with correlated sources of private information who take coordinated actions to optimize a payoff function. The agents may communicate beforehand to agree on any protocol of their choice.

Similarly, our model involves a teacher and student who share a common goal and are free devise a joint strategy; however, the specifics of the mathematical model in Ho [21] differ from ours in several ways, particularly with respect to the repeated observations and actions in our setting.

Our model also shares several similarities with Witsenhausen's counterexample from stochastic control [22]. This counterexample, which involves two agents—one with a perfect observation of the state but expensive control, and the other with a noisy observation of the state (after it has been modified by the first agent) but free control—demonstrates how non-classical information structures can alter the nature of optimal control solutions. The goal of both agents is to drive the state to 0. The problem may be reduced to finding the optimal strategy of the first agent, since the second agent can use a simple minimum mean-square-error decoder. In our model, the teacher has, in some sense, direct access to the state of the world; the teacher can then communicate this state to the student in a noisy manner. Just as in the optimal control problem, our goal is to determine the teacher's optimal strategy, since the student's optimal strategy is simply the MAP decoder once the teacher's strategy is fixed. (A notable difference is the absence of a "control" aspect in our setting.) Nonetheless, the difficulty in identifying the optimal strategies in Witsenhausen [22] and our paper exemplifies how asymmetric information structures can lead to very hard problems that might initially appear innocuous.

3) Fluctuation of Random Walks: Analyzing the sign of the random walk appearing in our paper may be seen as a part of the broader literature concerning fluctuation theory in probability [23]. These works explore the distributions of various quantities such as the time to reach the maximum or minimum, and (of relevance to us) the sojourn time, which is the duration of time when the random walk is positive. Fluctuation theory is also closely related to ballot theorems in probability, where the probability of a random walk being positive, negative, or lying within certain bounds is studied [24]. Our contribution differs due to its focus on the large deviation properties of the sign, as opposed to characterizing its distribution as in Chung and Feller [25], Andersen [26], and several works along these lines. We note that the paper of Andersen [26] contains a result concerning the sign of a random walk that may be used to prove our Theorem 1 more directly. However, we have retained our original analysis (which is a brute force computation, as opposed to the specialized result from Anderson [26]), because the technique extends to non-binary random walks and the calculation of rate functions for the duration of time the random walk lies within a certain interval. A possible generalization of our model to continuous random walks may be possible to study using arcsine laws [27], [28], but we do not explore that in this paper.

4) Communication Theory: A model identical to ours was proposed independently and concurrently in Huleihel *et al.* [29], who studied how to reliably send one bit across a cascade of binary symmetric channels. For simplicity, Huleihel *et al.* [29] assume that all the binary symmetric channels in the cascade are identical, with flipping probability δ. Just as in this paper, the authors examine the learning rate (which they call the "information velocity") as a function of the number of channels k in the cascade. The paper establishes that when $\delta \to 1/2$, the ratio of the learning rates for $k = 2$ and $k = 1$ is at least $3/4$, and proposes an interesting conjecture that the ratio should be arbitrarily close to 1. Translated into our setting, this would mean that when the teacher and student both have very noisy channels, the student learns as fast as the teacher; i.e., the channel from the teacher to the student can be made "clean." We shall further discuss the conjecture from Huleihel *et al.* [29] in our concluding section along with other conjectures of interest.

III. Technical Results

The following results, related to the sign sequence of a random walk and its derived properties, will be used throughout our paper. They provide the main technical results underlying our calculations of explicit learning rates. Throughout our analysis, we assume without loss of generality that $\Theta = +1$.

The random process $X_n := \sum_{i=1}^{n} Y_i$, recording the cumulative observations of the teacher, may be modeled as a random walk $\mathbb{Z}$ with the following transition probabilities:

$$p(X_{n+1} = i+1 | X_n = i) = 1-p, \qquad \text{and}$$
$$p(X_{n+1} = i-1 | X_n = i) = p,$$

where $p < 1/2$. Notice that this random walk is transient; i.e., for every $i \in \mathbb{Z}$, the random walk visits state i finitely many times, with probability 1. Since $p < 1/2$, the random walk eventually runs off to $+\infty$. Now define the following process:

$$\hat{X}_n = \begin{cases} +1 & \text{if } X_n > 0, \\ -1 & \text{if } X_n < 0, \\ \hat{X}_{n-1} & \text{if } X_n = 0, \end{cases}$$

which is the teacher's best guess about the state of the world at time n.

Let $M_n := \sum_{i=1}^{n} \mathbb{1}(\hat{X}_i = +1)$ denote the number of times that the teacher's majority guess is correct up to time n. In order to determine learning rates for the cumulative teaching strategy, we will need to explore the large deviations behavior of M_n. In particular, we are interested in the quantity $\mathbb{P}\left(\frac{M_n}{n} \approx 1-\delta\right)$. We expect this probability to be approximately equal to $e^{-nf(\delta)}$, for some suitable exponent $f(\delta)$; in Section III-B, we will pinpoint the function $f(\delta)$ in terms of p and δ.

A. Preliminary Calculations for $\{X_n\}$

Since the random walk $\{X_n\}$ is transient, it has a positive probability of never returning to state i, starting from state i. A simple calculation shows that this probability is $1-2p$, for any value of i.

Next, we focus on the sojourn time T, defined to be the time of the *first* return to 0, starting from 0. We use the convention that T is positive if the random walk is positive during the sojourn; otherwise, T is negative. Note that sojourn times can only take even values: $T = 2k$ when the random walk takes a total of k positive steps and k negative steps, with only the

endpoints of the sojourn being at 0. The probability that $T = 2k$ may be calculated by counting the number of such paths and multiplying the result by $p^k\bar{p}^k$. Furthermore, the number of these paths is the $(k-1)^{\text{st}}$ Catalan number [30], $C_{k-1} = \frac{1}{k}\binom{2k-2}{k-1}$. The distribution of T is therefore given by

$$\mathbb{P}(T=2k) = \begin{cases} p^{|k|}\bar{p}^{|k|}\frac{1}{|k|}\binom{2|k|-2}{|k|-1} & \text{if } 0<|k|<\infty, \\ 0 & \text{if } k=0, \\ 1-2p & \text{if } k=+\infty, \\ 0 & \text{if } k=-\infty. \end{cases}$$

We will be interested in the random variable T *conditioned on* the event that $|T| < \infty$, which we call $\tilde{T}$. It is easy to see that $\mathbb{E}\tilde{T} = 0$, and

$$\mathbb{E}|\tilde{T}| = \sum_{k=1}^{\infty} 2 \cdot \frac{p^k\bar{p}^k}{2p} C_{k-1} \cdot 2k = 2\sum_{k=0}^{\infty} \frac{p^{k+1}\bar{p}^{k+1}}{p} C_k \cdot (k+1).$$

Recall that the generating function of the Catalan numbers is given by

$$f(x) = \sum_{k=0}^{\infty} C_k x^k = \frac{1-\sqrt{1-4x}}{2x}. \tag{2}$$

Thus, $\sum_{k=0}^{\infty} C_k(k+1)x^k = \frac{d}{dx}(xf(x)) = \frac{1}{\sqrt{1-4x}}$. Substituting $x = p\bar{p}$, we conclude that $\mathbb{E}|\tilde{T}| = \frac{2\bar{p}}{(\bar{p}-p)}$.

Another quantity that will be critical for deriving large deviations results is the log moment generating function of the random vector $(\tilde{T}, |\tilde{T}|)$, defined by

$$L(\lambda_1, \lambda_2) = \log \mathbb{E} e^{\lambda_1 \tilde{T} + \lambda_2 |\tilde{T}|}.$$

The following lemma is proved in Appendix A:

Lemma 1: Let

$$\mathcal{D} = \{(x_1, x_2) \mid |x_1| + x_2 \le D(1/2||p)\}.$$

For $(\lambda_1, \lambda_2) \in \mathcal{D}$, we have $L(\lambda_1, \lambda_2) \le \log \frac{1}{2p}$. For $(\lambda_1, \lambda_2) \notin \mathcal{D}$, we have $L(\lambda_1, \lambda_2) = +\infty$.

Finally, the last ingredient we need is the number of returns of the random walk $\{X_n\}$ to 0. This is a geometric random variable, with distribution given by

$$\mathbb{P}(G = i) = (2p)^i(1-2p), \quad i \ge 0.$$

B. Large Deviations Properties of M_n

Let B be the random variable indicating the time of the final visit of $\{X_n\}$ to state 0. We break up the probability of interest as follows:

$$\mathbb{P}\left(\frac{M_n}{n} \le 1-\delta\right) = \left(\sum_{g=0}^{n/2} \mathbb{P}(M_n \le n(1-\delta), B \le n, G = g)\right) + \mathbb{P}(M_n \le n(1-\delta), B > n). \tag{3}$$

Note that the sum contains fewer than n terms (the maximum number of returns to 0 in n steps is $\frac{n}{2}$), and the largest among these terms will dictate the exponential growth rate of the sum.

We have the following lemma, proved in Appendix B:

Lemma 2: We have

$$\mathbb{P}(M_n \le n(1-\delta), B > n) \approx e^{-n(D(1/2||p)}.$$

We now turn to the initial $\frac{n}{2}$ terms in equation (3). We rewrite the probability as follows:

$$\sum_{g=0}^{n/2} \mathbb{P}(M_n \le n(1-\delta), B \le n, G = g) = \sum_{g=0}^{n/2} \mathbb{P}(G=g) \times \left\{ \sum_{b=2g}^{n} \sum_{a=-b}^{b} \mathbb{P}\left(\sum_{j=1}^{g} \tilde{T}_j = -a, \sum_{j=1}^{g} |\tilde{T}_j| = b\right) \mathbb{1}\left(\frac{a+b}{2} \ge n\delta\right)\right\}.$$

This is because conditioned on $G = g$, the behavior of the random walk can be broken into segments in between return times to 0, corresponding to the sojourn times $\{\tilde{T}_1, \ldots, \tilde{T}_g\}$. Furthermore, the time $B = b$ of the final sojourn time must be at least $2g$ and at most n, which is the time of the last return to 0; and the sum $-a$ of the signed sojourn times is then clearly in $[-b, b]$. Finally, the quantity M_n is a sum of $(n-b)$ (corresponding to the final $n-b$ time steps) and the sum of lengths of positive sojourn segments, which is $\frac{b-a}{2}$. Then the inequality $M_n \le n(1-\delta)$ is equivalent to $\frac{a+b}{2} \ge n\delta$.

We now substitute $\alpha := \frac{a}{n}, \beta := \frac{b}{n}$, and $\gamma := \frac{g}{n}$. We may rewrite the above expression as a summation over α, β, and γ, where we implicitly assume that these variables take values of the form $\frac{i}{n}$, for some integer i:

$$\sum_{\gamma=0}^{1/2} \mathbb{P}\left(\frac{G}{n} = \gamma\right) \times \sum_{\beta=\max(\delta,2\gamma)}^{1} \sum_{\alpha=2\delta-\beta}^{\beta} \mathbb{P}\left(\frac{\sum_{j=1}^{n\gamma} \tilde{T}_j}{n} = -\alpha, \frac{\sum_{j=1}^{n\gamma} |\tilde{T}_j|}{n} = \beta\right).$$

Our next theorem is a core technical result of this paper:

Theorem 1: Let $\delta \in [0,1]$. The following equality holds:

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{M_n}{n} \le (1-\delta)\right) = -\delta D(1/2||p).$$

Proof: Define the random vector

$$Z_n := \left(\frac{G}{n}, \frac{\sum_{j=1}^{G} \tilde{T}_j}{n}, \frac{\sum_{j=1}^{G} |\tilde{T}_j|}{n}\right).$$

Define the set

$$\mathcal{Q} := \left\{\alpha, \beta, \gamma \mid \gamma \in [0, 1/2], \beta \in [\max(\delta, 2\gamma), 1], \alpha \in [2\delta - \beta, \beta]\right\} \subseteq \mathbb{R}^3.$$

Note that we are interested in the quantity $\mathbb{P}(Z_n \in \mathcal{Q})$. We will evaluate this probability for large n using the Gärtner-Ellis theorem from large deviation theory [31]. The first step is to show that the following limit exists for every

$\lambda := (\lambda_1, \lambda_2, \lambda_3) \in \mathbb{R}^3$:

$$\begin{aligned}
\Lambda(\lambda) &:= \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E} e^{n\lambda \cdot Z_n} \\
&= \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E} \exp\left(\lambda_1 G + \lambda_2 \left(\sum_{j=1}^{G} \tilde{T}_j\right) + \lambda_3 \left(\sum_{j=1}^{G} |\tilde{T}_j|\right)\right) \\
&= \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left[\mathbb{E}\left[\exp\left(\lambda_1 G \right.\right.\right. \\
&\qquad \left.\left.\left. + \lambda_2 \left(\sum_{j=1}^{G} \tilde{T}_j\right) + \lambda_3 \left(\sum_{j=1}^{G} |\tilde{T}_j|\right)\right)\Big| G\right]\right] \\
&\overset{(a)}{=} \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left[e^{\lambda_1 G + L(\lambda_2,\lambda_3) G}\right] \\
&\overset{(b)}{=} \begin{cases} 0 & \text{if } \lambda_1 + L(\lambda_2,\lambda_3) < \log\left(\frac{1}{2p}\right), \\ +\infty & \text{otherwise,} \end{cases}
\end{aligned}$$

where equality (a) holds for $(\lambda_2, \lambda_3) \in \mathcal{D}$, using the fact that the $\tilde{T}_j$'s are conditionally i.i.d. (and the limit is $+\infty$ if $(\lambda_2, \lambda_3) \notin \mathcal{D}$); and equality (b) holds by evaluating the moment generating function of a geometric random variable. To summarize, we have

$$\Lambda(\lambda) = \begin{cases} +\infty & \text{if } (\lambda_2,\lambda_3) \in \mathcal{D}^c, \\ +\infty & \text{if } (\lambda_2,\lambda_3) \in \mathcal{D} \text{ and } \lambda_1 \geq \log\left(\frac{1}{2p}\right) - L(\lambda_2,\lambda_3), \\ 0 & \text{otherwise.} \end{cases}$$

Let the domain of Λ be $\mathcal{D}_\Lambda$, and let Λ^* denote the convex conjugate of Λ. A direct application of the Gärtner-Ellis theorem then gives the following upper bound:

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(Z_n \in \mathcal{Q}) \leq -\inf_{z\in\mathcal{Q}} \Lambda^*(z). \tag{4}$$

We now evaluate the convex conjugate of Λ at the point $(\gamma, \alpha, \beta) \in \mathbb{R}^3_+$:

$$\begin{aligned}
\Lambda^*(\gamma,\alpha,\beta) &= \sup_{\lambda\in\mathcal{D}_\Lambda} \{\lambda_1\gamma + \lambda_2\alpha + \lambda_3\beta\} \\
&\overset{(a)}{=} \sup_{(\lambda_2,\lambda_3)\in\mathcal{D}} \left\{ \left(\log\left(\frac{1}{2p}\right) - L(\lambda_2,\lambda_3)\right)\gamma \right. \\
&\qquad \left. + \lambda_2\alpha + \lambda_3\beta \right\},
\end{aligned}$$

where in equality (a), we have used the fact that $\gamma \geq 0$ in $\mathcal{Q}$. We now make the following crucial observations: First, Lemma 1 implies that the coefficient of γ in the above expression is positive, so $\Lambda^*(\gamma, \alpha, \beta)$ is a monotonically increasing function of γ for fixed α and β. Second, the set $\mathcal{Q}$ is such that the possible values of the pair (α, β) can only increase as γ becomes smaller. This implies that if $\gamma_1 < \gamma_2$, then

$$\inf_{(\alpha,\beta):(\gamma_1,\alpha,\beta)\in\mathcal{Q}} \Lambda^*(\gamma_1,\alpha,\beta) < \inf_{(\alpha,\beta):(\gamma_2,\alpha,\beta)\in\mathcal{Q}} \Lambda^*(\gamma_2,\alpha,\beta),$$

since not only is the left-hand objective smaller than the right-hand objective for every fixed (α, β), but also the range of possible values of (α, β) on the left-hand side contains the range of values on the right-hand side. Thus, the infimum of Λ^* over $\mathcal{Q}$ must occur when $\gamma = 0$; i.e.,

$$\inf_{(\gamma,\alpha,\beta)\in\mathcal{Q}} \Lambda^*(\gamma,\alpha,\beta) = \inf_{\beta\in[\delta,1],\alpha\in[2\delta-\beta,\beta]} \Lambda^*(0,\alpha,\beta).$$

When $\beta \in [\delta, 1]$ and $\alpha \in [\beta - 2\delta, \beta]$, we have

$$\Lambda^*(0,\alpha,\beta) = \sup_{(\lambda_2,\lambda_3)\in\mathcal{D}} \{\lambda_2\alpha + \lambda_3\beta\} = \beta D(1/2||p).$$

Hence, we conclude that

$$\begin{aligned}
\inf_{(\gamma,\alpha,\beta)\in\mathcal{Q}} \Lambda^*(\gamma,\alpha,\beta) &= \inf_{\beta\in[\delta,1],\alpha\in[2\delta-\beta,\beta]} \beta D(1/2||p) \\
&= \delta D(1/2||p).
\end{aligned}$$

To complete the proof, we need to establish a lower bound counterpart to inequality (4). This is established via the following lemma, proved in Appendix C:

Lemma 3: The following inequality holds:

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(Z_n \in \mathcal{Q}) \geq -\delta D(1/2||p).$$

The proof follows by constructing a set of paths that satisfy $M_n < (1-\delta)n$ and explicitly computing the combined probability of these paths. This completes the proof of Theorem 1. □

C. Large Deviations for a Bernoulli Mixture

Finally, we first prove the following large deviations result for a mixture of Bernoulli random variables. Recall [31] that the rate function I of a sequence of random variables $\{W_n\}$ has the property that for any closed subset $F \subseteq \mathbb{R}$ and any open subset $G \subseteq \mathbb{R}$, we have

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(W_n \in F) &\leq -\inf_{w\in F} I(w), \\
\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(W_n \in G) &\geq -\inf_{w\in G} I(w).
\end{aligned}$$

Lemma 4: Let $\theta \in [0,1]$ and $q \in [0, 1/2]$. Consider a sequence of i.i.d. Bernoulli$(1-q)$ random variables $\{U_i\}_{1\leq i\leq n-\lfloor n\theta\rfloor}$ and a sequence of i.i.d. Bernoulli(q) random variables $\{V_j\}_{1\leq j\leq \lfloor n\theta\rfloor}$, such that the U_i's are independent of the V_j's. The random variable

$$W_n := \frac{\sum_{i=1}^{n-\lfloor n\theta\rfloor} U_i + \sum_{j=1}^{\lfloor n\theta\rfloor} V_j}{n}$$

satisfies the large deviation principle with rate function

$$I_\theta(w) = w\log(\eta) - \bar{\theta}\log(\bar{q}\eta + q) - \theta\log(q\eta + \bar{q}),$$

where

$$\eta := \frac{-\tau + \sqrt{\tau^2 + 4w\bar{w}}}{2\bar{w}}, \text{ and } \tau := \frac{\bar{q}}{q}(\bar{\theta} - w) + \frac{q}{\bar{q}}(\theta - w).$$

The proof is a direct application of the Gärtner-Ellis theorem and is detailed in Appendix D. Lemma 4 will be useful for evaluating the probability that the student makes an error when we condition on various aspects of the random walk $\{X_n\}$, such as the number M_n of $+1$ transmissions by the teacher, or the last return time of the random walk to 0.

IV. Majority Learning

We now present our main results in the case where the student employs the simplest majority rule strategy.

A. Simple Forwarding

If the teacher employs the simple forwarding strategy, the student's observations $\{Z_n\}$ may be viewed as noisy observations of Θ through a binary symmetric channel with error parameter

$$p \star q := p(1-q) + q(1-p) = p\bar{q} + q\bar{p}.$$

In this case, the majority learning strategy for the student exactly corresponds to the Bayesian strategy, producing an optimal learning rate of $D(1/2||p \star q)$ [31]. We will use this simple expression as a benchmark for comparing the learning rate of more complicated combinations of teacher/student strategies.

B. Cumulative Teaching

Note that Theorem 1 already provides the exact learning rate if $q = 0$, since the student makes an error exactly when $M_n < \frac{n}{2}$. Substituting $\delta = 1/2$, we see that the student will learn at a rate of $\frac{1}{2}D(1/2||p)$ via the cumulative teaching strategy. In contrast, the learning rate for simple forwarding is $D(1/2||p)$, since $p \star q = p$—and this rate is *higher* than that of cumulative teaching! It is natural to wonder what values of the channel parameters (p, q) make simple forwarding a better strategy than cumulative teaching, and vice versa.

To analyze the general case $q > 0$, we can use Lemma 4 to evaluate the probability that at most $n(1-\delta)$ instances of the student's received sequence $\{Z_n\}$ are equal to $+1$. The learning rate of a student using a majority learning rule can then be obtained by plugging in $\delta = \frac{1}{2}$.

Theorem 2: Let $\delta \in [q, 1-q]$. Suppose we say that the student commits an error if the fraction of received $+1$'s is at most $(1-\delta)$. Then the rate of learning is given by

$$\mathcal{R} = \inf_{\theta \in [0,1]} \left\{ \theta D(1/2||p) + \hat{I}(\theta) \right\},$$

where

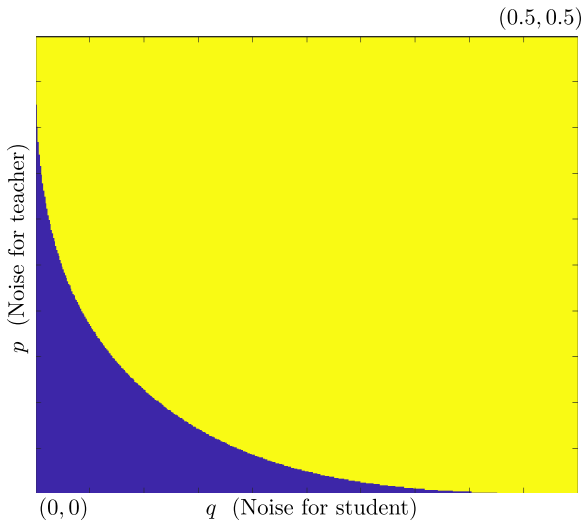
$$\hat{I}(\theta) = \inf_{w \in [0, 1-\delta]} I_\theta(w),$$

and $I_\theta(\cdot)$ is the rate function appearing in Lemma 4.

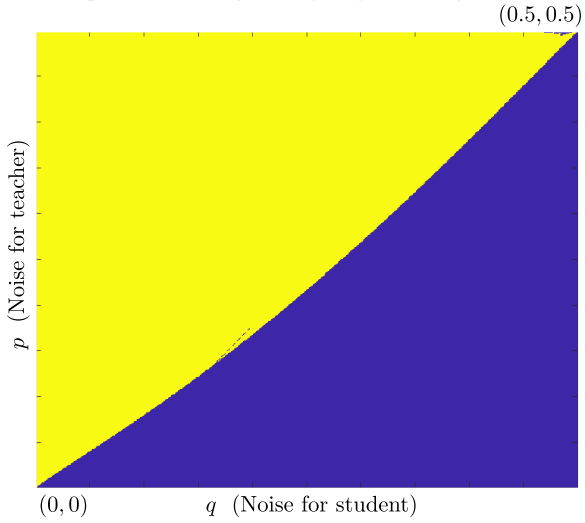
The learning rate provided in Thereom 2 can be approximately calculated using computing software; see Figure 2.

Proof: Let $\mathcal{E}_n$ be the error event that the number of $+1$'s received by the student is at most $n(1-\delta)$. Consider an integer $N > 0$, whose value will be specified later. We divide the interval $[0, 1)$ into intervals $\{L_i\}_{i=0}^{N-1}$, where $L_i := \left[\frac{i}{N}, \frac{i+1}{N}\right)$. The probability of error can then be written as

$$\mathbb{P}(\mathcal{E}_n) = \sum_{i=0}^{N-1} \mathbb{P}\left(\frac{M_n}{n} \in L_i\right) \mathbb{P}\left(\mathcal{E}_n \middle| \frac{M_n}{n} \in L_i\right). \quad (5)$$



(a) ϵ-teaching + ϵ-learning (yellow) vs. simple forwarding + majority learning (blue)



(b) ϵ-teaching + ϵ-learning (yellow) vs. cumulative teaching + majority learning (blue)

Fig. 2. Plots showing the benefits of ϵ-teaching + ϵ-learning, analyzed in Theorem 3. The figure on the right is simulated using the formula in Theorem 2. In both figures, the (lighter) yellow region indicates the parameter values where ϵ-teaching + ϵ-learning dominates the competing strategy based on majority learning. In (a), note that when the cumulative noise in the communication system is small, it is better to simply forward information. In (b), note that when the teacher's noise is large compared to the student's noise, the ϵ-teaching strategy dominates the cumulative teaching strategy, since the teacher has less time to confuse himself.

Note that

$$\mathbb{P}\left(\frac{M_n}{n} \in L_i\right) = \mathbb{P}\left(\frac{M_n}{n} < \frac{i+1}{N}\right) - \mathbb{P}\left(\frac{M_n}{n} < \frac{i}{N}\right). \quad (6)$$

Let $\epsilon_1 > 0$ be an arbitrarily small constant. By Theorem 1, we know that for all $0 \le i \le N-1$ and for all large enough n, the following inequality holds:

$$\left| -\frac{1}{n} \log \mathbb{P}\left(\frac{M_n}{n} \in L_i\right) - \frac{N-i-1}{N} D\left(1/2||p\right) \right| < \epsilon_1, \quad (7)$$

since the first term in equation (6) dominates.

Turning to the second term in the summand of equation (5), we note that the probability of error is a monotonically decreasing function of $\frac{M_n}{n}$, so

$$\mathbb{P}\left(\mathcal{E}_n \middle| M_n = \left\lceil n \cdot \frac{i+1}{N} \right\rceil\right) \leq \mathbb{P}\left(\mathcal{E}_n \middle| \frac{M_n}{n} \in L_i\right) \leq \mathbb{P}\left(\mathcal{E}_n \middle| M_n = \left\lfloor n \cdot \frac{i}{N} \right\rfloor\right). \quad (8)$$

Let $\epsilon_2 > 0$ be an arbitrarily small constant. Using the large deviations principle from Lemma 4, we know that for all large enough n and for all $0 \leq i \leq N-1$, we have the bound

$$\left|\frac{1}{n}\log \mathbb{P}\left(\mathcal{E}_n \middle| M_n = \left\lfloor n \cdot \frac{i}{N} \right\rfloor\right) - \hat{I}\left(\frac{N-i}{N}\right)\right| < \epsilon_2, \quad (9)$$

since conditioned on the fraction $\frac{M_n}{n} \approx \frac{i}{N}$ of +1's that are transmitted by the teacher, the distribution of +1's received by the student behaves exactly as the mixture of Bernoulli distributions in Lemma 4 with $\theta = 1 - \frac{i}{N}$. Furthermore, the conditional probability of the error event exactly corresponds to the infimum of the rate function over the interval $[0, 1-\delta]$. We have a similar bound for the left-hand expression in inequality (8).

Combining the bounds from inequalities (7) and (9), we then obtain

$$\mathbb{P}(\mathcal{E}_n) \leq \sum_{i=0}^{N-1} e^{-n\left(\frac{N-i-1}{N} D(1/2||p) + \hat{I}\left(\frac{N-i}{N}\right) - \epsilon_1 - \epsilon_2\right)},$$

$$\mathbb{P}(\mathcal{E}_n) \geq \sum_{i=0}^{N-1} e^{-n\left(\frac{N-i-1}{N} D(1/2||p) + \hat{I}\left(\frac{N-i-1}{N}\right) + \epsilon_1 + \epsilon_2\right)}.$$

Let $\epsilon_3 > 0$ be an arbitrarily small constant. Define the three quantities

$$u := \inf_{x\in[0,1]} \left\{ xD(1/2||p) + \hat{I}(x) \right\},$$

$$\bar{u} := \inf_{0\leq i\leq N-1} \left\{ \frac{N-i-1}{N} D(1/2||p) + \hat{I}\left(\frac{N-i}{N}\right) \right\},$$

$$\underline{u} := \inf_{1\leq i\leq N-1} \left\{ \frac{N-i-1}{N} D(1/2||p) + \hat{I}\left(\frac{N-i-1}{N}\right) \right\}.$$

Using the continuity of $\hat{I}$, we can now pick N (depending only on ϵ_3 and p) such that

$$\max(|u-\bar{u}|, |u-\underline{u}|) < \epsilon_3.$$

Then for all large enough n, we have the bounds

$$\mathbb{P}(\mathcal{E}_n) \leq Ne^{-n(u-\epsilon_1-\epsilon_2-\epsilon_3)},$$

$$\mathbb{P}(\mathcal{E}_n) \geq e^{-n(u+\epsilon_1+\epsilon_2+\epsilon_3)}.$$

Taking logarithms, dividing by n, and taking the limit, we see that

$$\left|\lim_{n\to\infty} \frac{1}{n}\log \mathbb{P}(\mathcal{E}_n) - u\right| < \epsilon_1 + \epsilon_2 + \epsilon_3.$$

Since ϵ_1, ϵ_2, and ϵ_3 are arbitrary constants, we conclude that

$$\lim_{n\to\infty} \frac{1}{n}\log \mathbb{P}(\mathcal{E}_n) = u,$$

completing the proof. □

Remark: The learning rate in Theorem 2 may be simplified further: Note that when $\bar{\theta}\bar{q} + \theta q \in [0, 1-\delta]$, the value of $\hat{I}(\theta)$ is 0, since $I_\theta(\bar{\theta}\bar{q} + \theta q) = 0$. The constraint $\bar{\theta}\bar{q} + \theta q \in [0, 1-\delta]$ is equivalent to $\theta \in \left[\frac{\bar{q}}{\bar{q}-q}, \frac{\delta-q}{\bar{q}-q}\right]$. Hence, the minimum of $\theta D(1/2||p) + \hat{I}(\theta)$ is clearly achieved over this range for $\theta = \frac{\delta-q}{\bar{q}-q}$. Further note that for $\theta \in \left[0, \frac{\delta-q}{\bar{q}-q}\right]$, we have the equality $\hat{I}(\theta) = I_\theta(1-\delta)$, since I_θ is convex and minimized at $w = \bar{\theta}\bar{q} + \theta q > 1-\delta$. Thus, we can rewrite the learning rate as

$$\mathcal{R} = \inf_{\theta\in\left[0, \frac{\delta-q}{\bar{q}-q}\right]} \left\{ \theta D(1/2||p) + I_\theta(1-\delta) \right\}.$$

This expression does not simplify further; but since the function being minimized is known in closed form, the value of $\mathcal{R}$ is easy to simulate using Matlab or similar software.

V. ϵ-Majority Learning

In the previous section, we assumed that the student is simply a majority learner; i.e., the student takes the majority of her observations to ultimately decide the value of $\hat{\Theta}$. We now explore the alternative strategy where the student's final estimate is computed to be the majority over the final ϵn observations, corresponding to a more lazy or more skeptical student. We assume that the student has access to the problem parameters p and q, and utilizes the value of ϵ that maximizes her learning rate.

A. Simple Forwarding

Suppose that the teacher is simply forwarding his observations to the student and the student is taking an ϵ-majority based on her observations. It is easy to see that the learning rate for the student in this case is simply $\epsilon D(1/2||p \star q)$, since her estimate is always a majority over some number of i.i.d. observations. Thus, the optimal choice of ϵ is 1, and the best learning rate for ϵ-majority learning simply coincides with majority learning.

B. ϵ-Teaching

We first analyze the ϵ-teaching strategy, where the teacher transmits noise in the first $(1-\epsilon)n$ time steps, and then repeatedly transmits his best guess at that point in the remaining ϵn time steps.

Theorem 3: For the ϵ-teaching strategy, the student's learning rate, optimized over all values of $\epsilon \in [0,1]$, is given by

$$\mathcal{R} = \frac{D(1/2||p)D(1/2||q)}{D(1/2||p) + D(1/2||q)}.$$

Proof: Let the teacher's estimate at time $\bar{\epsilon}n$, obtained by taking a majority over his observations until that time, be $\tilde{\Theta}$. The teacher then transmits $\tilde{\Theta}$ repeatedly for the final ϵn time slots, and the student takes a majority over her ϵn observations to determine $\hat{\Theta}$. The probability of error is thus

calculated to be

$$\begin{aligned}\mathbb{P}(\hat{\Theta} \neq \Theta) &= \mathbb{P}(\hat{\Theta} \neq \Theta, \tilde{\Theta} \neq \Theta) + \mathbb{P}(\hat{\Theta} \neq \Theta, \tilde{\Theta} = \Theta) \\ &= \mathbb{P}(\hat{\Theta} \neq \Theta | \tilde{\Theta} \neq \Theta)\mathbb{P}(\tilde{\Theta} \neq \Theta) \\ &\quad + \mathbb{P}(\hat{\Theta} \neq \Theta | \tilde{\Theta} = \Theta)\mathbb{P}(\tilde{\Theta} = \Theta) \\ &= \mathbb{P}(\hat{\Theta} = \tilde{\Theta})\mathbb{P}(\tilde{\Theta} \neq \Theta) + \mathbb{P}(\hat{\Theta} \neq \tilde{\Theta})\mathbb{P}(\tilde{\Theta} = \Theta). \end{aligned} \tag{10}$$

Notice that

$$\mathbb{P}(\tilde{\Theta} \neq \Theta) \approx \exp(-n\bar{\epsilon}D(1/2||p)),$$

and

$$\mathbb{P}(\hat{\Theta} \neq \tilde{\Theta}) \approx \exp(-n\epsilon D(1/2||q)),$$

since the teacher obtains $\tilde{\Theta}$ from $n\bar{\epsilon}$ observations of Θ through a binary symmetric channel with error probability p, and the student obtains $\hat{\Theta}$ from $n\epsilon$ observations of $\tilde{\Theta}$ through a binary symmetric channel with error probability q.

Substituting back into equation (10), we see that the learning rate is determined by

$$\min(\bar{\epsilon}D(1/2||p), \epsilon D(1/2||q)).$$

It is not hard to see that this expression is maximized when

$$\epsilon = \frac{D(1/2||p)}{D(1/2||p) + D(1/2||q)},$$

yielding the specified learning rate. □

As we will see, the rate obtained in Theorem 3 is not optimal from the point of view of the teacher; however, the ϵ-teaching strategy nonetheless provides a nice closed-form expression for the learning rate, which is convenient for comparison.

C. Cumulative Teaching

In the above strategy, the teacher essentially stops learning after time $\bar{\epsilon}n$. Intuitively, the learning rate should be higher if the teacher continues to transmit better and better estimates of the state of the world in the latter ϵn time steps, based on his own continual learning. Consequently, we now analyze the cumulative teaching strategy of the teacher, where the student continues to be an ϵ-majority learner. Note that when $\epsilon = 1$, this strategy is identical to the cumulative teaching + majority learning combination studied before. Since the best choice of ϵ will outperform $\epsilon = 1$, the learning rate for this combination will be at least as high.

We claim that when the student is an ϵ-learner, cumulative teaching is indeed at least as good as ϵ-teaching. This is stated and proved as a warm-up in Proposition 1; in Theorem 4 to follow, we provide an expression for the precise learning rate.

Proposition 1: The learning rate of cumulative teaching and ϵ-majority learning (with the optimum choice of ϵ) is at least as large as the learning rate of the ϵ-teaching strategy from Theorem 3.

Proof: Fix $\epsilon > 0$, and let B be the final time that the random walk $\{X_n\}$ visits 0. Let $\hat{\Theta}$ denote the final estimate of the student. We write

$$\begin{aligned}\mathbb{P}(\hat{\Theta} \neq \Theta) &= \mathbb{P}(\hat{\Theta} \neq \Theta \mid B \leq \bar{\epsilon}n)\mathbb{P}(B \leq \bar{\epsilon}n) \\ &\quad + \mathbb{P}(\hat{\Theta} \neq \Theta \mid B > \bar{\epsilon}n)\mathbb{P}(B > \bar{\epsilon}n),\end{aligned}$$

and compare the respective quantities in the cases of cumulative teaching vs. ϵ-teaching. The probabilities $\mathbb{P}(B \leq \bar{\epsilon}n)$ and $\mathbb{P}(B > \bar{\epsilon}n)$ have no dependence on the strategy employed by the teacher.

Note that conditioned on the event $\{B \leq \bar{\epsilon}n\}$, the ϵ-teaching and cumulative teaching strategies are identical as far as the final ϵn bits are concerned, so the values of $\mathbb{P}(\hat{\Theta} \neq \Theta \mid B \leq \bar{\epsilon}n)$ are identical. Conditioned on the event $\{B > \bar{\epsilon}n\}$, the random walk $\{X_n\}$ is equally likely to be positive or negative at time $\bar{\epsilon}n$, so in the case of ϵ-teaching, we have $\mathbb{P}(\hat{\Theta} \neq \Theta \mid B > \bar{\epsilon}n) = \frac{1}{2}$. In the case of cumulative teaching, we can lower-bound the learning rate by upper-bounding the error as $\mathbb{P}(\hat{\Theta} \neq \Theta \mid B > \bar{\epsilon}n) \leq 1$.

It is easy to see that increasing the value of the error probability $\mathbb{P}(\hat{\Theta} \neq \Theta \mid B > \bar{\epsilon}n)$ by a factor of 2 cannot make the learning rate of cumulative teaching worse than the learning rate of ϵ-teaching. Thus, we conclude that for every choice of $\epsilon > 0$, the learning rate of cumulative teaching is at least as high as the learning rate of ϵ-teaching. Optimizing the student's choice of ϵ when the teacher employs cumulative teaching can only increase the learning rate further, completing the proof. □

Using a more sophisticated argument, we can obtain an expression for the learning rate of cumulative teaching when the student is an ϵ-majority learner:

Theorem 4: The optimal learning rate for the cumulative teaching and ϵ-learning strategy is

$$\mathcal{R} = \sup_{\epsilon \in [0,1]} \left[\min \left(\epsilon D(1/2||q), \inf_{\alpha \in [0,1/2]} \{(\bar{\epsilon} + \epsilon\alpha)D(1/2||p) + \epsilon I_\alpha(1/2)\} \right) \right],$$

where $I_\alpha(\cdot)$ is the rate function defined in Lemma 4.

Proof: We will show that the learning rate for a fixed value of ϵ is given by the expression inside the objective function. Let B denote the last time at which the random walk $\{X_n\}$ corresponding to the teacher's transmissions visits 0. We write the error probability of the student as

$$\begin{aligned}\mathbb{P}(\hat{\Theta} \neq \Theta) &= \mathbb{P}(\hat{\Theta} \neq \Theta, B \leq (1-\epsilon)n) \\ &\quad + \mathbb{P}(\hat{\Theta} \neq \Theta, (1-\epsilon)n < B \leq n) \\ &\quad + \mathbb{P}(\hat{\Theta} \neq \Theta, B > n). \end{aligned} \tag{11}$$

Naturally, the largest of the three error probabilities determines the overall learning rate. For the third term in equation (11), we observe that

$$\begin{aligned}\mathbb{P}(\hat{\Theta} \neq \Theta, B > n) &= \mathbb{P}(B > n)P(\hat{\Theta} \neq \Theta | B > n) \\ &= \frac{1}{2}\mathbb{P}(B > n) \\ &\approx \exp(-nD(1/2||p)),\end{aligned}$$

where the second equality holds by the symmetry of the random walk conditioned on $\{B > n\}$, and the final approximation follows from the proof of Lemma 2 from Appendix B. For the first term in equation (11), notice that if $B \leq n(1-\epsilon)$, the teacher will transmit $+1$'s for the entire duration over

which the student is taking a majority. The student's error probability is the probability of receiving more than $\frac{n\epsilon}{2}$ observations of -1's in the final $n\epsilon$ observations, so we have

$$\begin{aligned}&\mathbb{P}(\hat{\Theta} \neq \Theta, B \leq (1-\epsilon)n)\\ &\quad = \mathbb{P}(\hat{\Theta} \neq \Theta | B \leq (1-\epsilon)n) \cdot \mathbb{P}(B \leq (1-\epsilon)n)\\ &\quad \approx \exp(-n\epsilon D(1/2||q)),\end{aligned}$$

using the fact that

$$\mathbb{P}(B > (1-\epsilon)n) \approx \exp(-(1-\epsilon)nD(1/2\|p)),$$

so $\mathbb{P}(B \leq (1-\epsilon)n) = 1 - o(1)$.

Turning to the middle term in equation (11), suppose we fix $\alpha \in [0,1]$, and suppose $B = \lfloor (1-\epsilon)n + \alpha\epsilon n \rfloor$. As we will see, the error probability is dominated by the error probability of a random walk that is negative at all time instances up to B. Thus, we define the indicator variable $I = 1$ if $\{X_n\}$ is negative for all time instances up to B, and 0 otherwise. We then write

$$\begin{aligned}&\mathbb{P}(B = \lfloor n\bar{\epsilon} + n\epsilon\alpha \rfloor, \hat{\Theta} \neq \Theta)\\ &\quad = \mathbb{P}(\hat{\Theta} \neq \Theta \mid B = n_\alpha, I = 0)\mathbb{P}(B = n_\alpha, I = 0)\\ &\qquad + \mathbb{P}(\hat{\Theta} \neq \Theta \mid B = n_\alpha, I = 1)\mathbb{P}(B = n_\alpha, I = 1),\end{aligned}$$

where we have used the shorthand $n_\alpha := \lfloor n\bar{\epsilon} + n\epsilon\alpha \rfloor$. Note that

$$\mathbb{P}(B = n_\alpha) = (1-2p)\binom{n_\alpha}{n_\alpha/2} p^{n_\alpha/2}\bar{p}^{n_\alpha/2},$$

$$\mathbb{P}(B = n_\alpha, I = 1) = (1-2p)\frac{1}{n_\alpha/2}\binom{n_\alpha - 2}{n_\alpha/2 - 1} p^{n_\alpha/2}\bar{p}^{n_\alpha/2},$$

where the second expression involves the corresponding Catalan number. Furthermore, by the same argument used in the proof of Lemma 3, we know that

$$\mathbb{P}(B = n_\alpha, I = 1) \approx \exp(-n(\bar{\epsilon} + \epsilon\alpha)D(1/2\|p)). \tag{12}$$

Since $\mathbb{P}(B = n_\alpha)$ is at most a polynomial factor of n larger, we conclude that

$$\mathbb{P}(B = n_\alpha, I = 0) \approx \exp(-n(\bar{\epsilon} + \epsilon\alpha)D(1/2\|p)),$$

as well. Turning to the conditional probability terms, note that

$$\mathbb{P}(\hat{\Theta} \neq \Theta \mid B = n_\alpha, I = 0) \leq \mathbb{P}(\hat{\Theta} \neq \Theta \mid B = n_\alpha, I = 1).$$

Hence, the error probability $\mathbb{P}(B = n_\alpha, \hat{\Theta} \neq \Theta)$ is within a poly(n) factor of

$$\mathbb{P}(\hat{\Theta} \neq \Theta \mid B = n_\alpha, I = 1)\mathbb{P}(B = n_\alpha, I = 1),$$

implying that

$$\begin{aligned}&\lim_{n\to\infty} \frac{1}{n}\log \mathbb{P}(B = n_\alpha, \hat{\Theta} \neq \Theta)\\ &\qquad = \lim_{n\to\infty} \frac{1}{n}\log\Big(\mathbb{P}(\hat{\Theta} \neq \Theta \mid B = n_\alpha, I = 1)\\ &\qquad\qquad \mathbb{P}(B = n_\alpha, I = 1)\Big).\end{aligned}$$

Now suppose $\alpha \in [0, 1/2]$. We have

$$\mathbb{P}(\hat{\Theta} \neq \Theta \mid B = n_\alpha, I = 1) \approx \exp(-n\epsilon I_\alpha(1/2)),$$

since conditioned on the event $\{B = n_\alpha, I = 1\}$, we know that the teacher transmits $\epsilon\alpha n$ values that are -1 and $(\epsilon n - \epsilon\alpha n)$ values that are +1 to the student in the last ϵn time steps, so the error probability is given by the rate function in Lemma 4. Combining the equations, we have

$$\begin{aligned}&\mathbb{P}(B = n_\alpha, \hat{\Theta} \neq \Theta)\\ &\quad \approx \exp(-n(\bar{\epsilon} + \epsilon\alpha)D(1/2||p)) \cdot \exp(-n\epsilon I_\alpha(1/2))\\ &\quad = \exp(-n((\bar{\epsilon} + \epsilon\alpha)D(1/2||p) + \epsilon I_\alpha(1/2))).\end{aligned}$$

Finally, for $\alpha > \frac{1}{2}$, note that

$$\mathbb{P}(B = n_\alpha, I = 1) \leq \mathbb{P}(B = n_{1/2}, I = 1)$$

by equation (12), and

$$\begin{aligned}&\mathbb{P}(\hat{\Theta} \neq \Theta \mid B = n_\alpha, I = 1) \leq 1\\ &\qquad = 2\mathbb{P}(\hat{\Theta} \neq \Theta \mid B = n_{1/2}, I = 1).\end{aligned}$$

Thus, we conclude that the error rate is given by

$$\begin{aligned}&\inf_{\alpha\in[0,1]} \lim_{n\to\infty} -\frac{1}{n}\log \mathbb{P}(B = n_\alpha, \hat{\Theta} \neq \Theta)\\ &\qquad = \inf_{\alpha\in[0,1/2]} \left\{(\bar{\epsilon} + \epsilon\alpha)D(1/2||p) + \epsilon I_\alpha(1/2)\right\}.\end{aligned}$$

Combining the bounds for the three terms in equation (11), the overall learning rate for a fixed choice of ϵ is therefore equal to

$$\begin{aligned}&\min\Big(D(1/2||p), \epsilon D(1/2||q),\\ &\qquad \inf_{\alpha\in[0,1/2]} \left\{(\bar{\epsilon} + \epsilon\alpha)D(1/2||p) + \epsilon I_\alpha(1/2)\right\}\Big).\end{aligned}$$

Since the student is allowed to tune ϵ, the optimal learning rate is therefore

$$\begin{aligned}&\sup_{\epsilon\in[0,1]} \Big[\Big(D(1/2||p), \epsilon D(1/2||q),\\ &\qquad \inf_{\alpha\in[0,1]} \left\{(\bar{\epsilon} + \epsilon\alpha)D(1/2||p) + \epsilon I_\alpha(1/2)\right\}\Big)\Big].\end{aligned}$$

Finally, note that we may drop the term $D(1/2||p)$ from the inner minimization, since when $\epsilon = 0$ (and for any choice of α), we have

$$(\bar{\epsilon} + \epsilon\alpha)D(1/2\|p) + \epsilon I_\alpha(1/2) = D(1/2\|p).$$

Thus, the optimal learning rate is

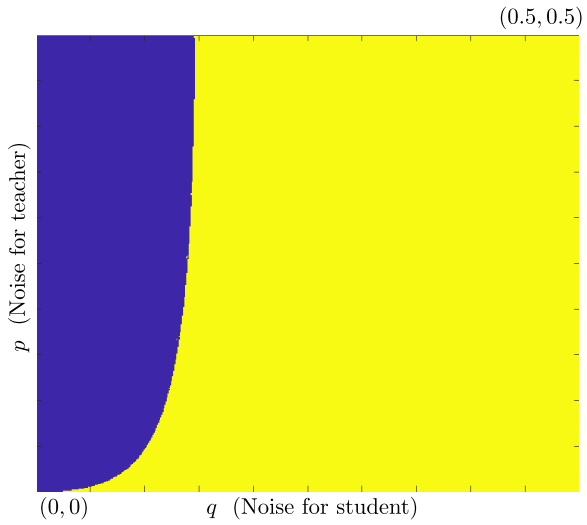
$$\begin{aligned}\mathcal{R} = \sup_{\epsilon\in[0,1]} \Big[&\min\Big(\epsilon D(1/2||q),\\ &\inf_{\alpha\in[0,1/2]} \left\{(\bar{\epsilon} + \epsilon\alpha)D(1/2||p) + \epsilon I_\alpha(1/2)\right\}\Big)\Big].\end{aligned}$$

□

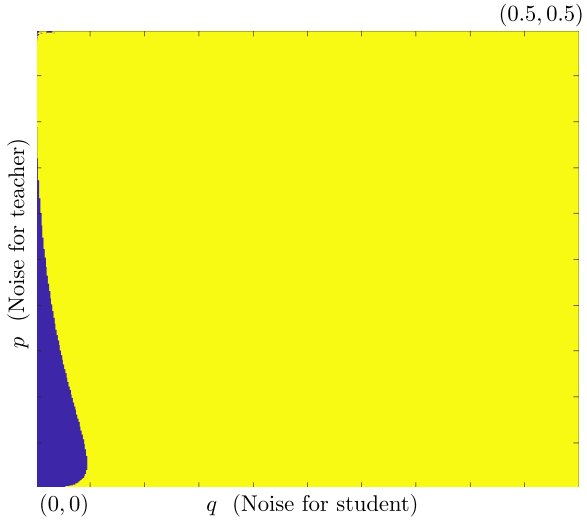
Note that when $q = 0$, i.e., the student is a perfect learner, the learning rate in Theorem 4 corresponds to the optimum of

$$\sup_{\epsilon\in[0,1]} \inf_{\alpha\in[0,1/2]} \left\{(\bar{\epsilon} + \epsilon\alpha)D(1/2||p) + \epsilon I_\alpha(1/2)\right\},$$

which occurs when $\epsilon = 0$ and produces a learning rate of $D(1/2\|p)$. In other words, the optimal strategy of the student is to estimate the state of the world based on the final transmission of the teacher, in which case her learning rate agrees with the teacher's learning rate. Although it is generally



(a) simple forwarding + majority learning (blue) vs. cumulative teaching + majority learning (yellow)



(b) simple forwarding + majority learning (blue) vs. cumulative teaching + ϵ-majority learning (yellow)

Fig. 3. Plots showing the benefits of ϵ-majority learning over majority learning, when the teacher employs a cumulative teaching strategy. The learning rate for the simple forwarding + majority learning strategy is $D(1/2||p\star q)$; the learning rate for the cumulative teaching + majority learning strategy is simulated using the formula in Theorem 2; and the learning rate for the cumulative teaching + ϵ-majority learning strategy is simulated using the formula derived in Theorem 4. In both figures, the blue region indicates the parameter values where simple forwarding + majority learning dominates the competing strategy. In (a), note that the simple forwarding + majority learning strategy is better for smaller values of q (i.e., sharper students). However, we see a curious threshold around $q = 0.15$, such that if the student's channel is more noisy that this value, it is better to use the cumulative teaching strategy *regardless* of the teacher's noise parameter. In (b), note that the (darker) blue region is significantly smaller, since the ϵ-majority learning strategy is uniformly better than the majority learning strategy. As before, we observe that the simple forwarding + majority learning strategy is better for smaller values of q. Furthermore, for all q larger than approximately 0.05, it is always preferable to use the ϵ-majority learning strategy over the majority learning strategy.

not possible to simplify the expression in Theorem 4 further for other choices of q, it is easy to calculate the learning rate to a high degree of accuracy using standard computing software. See Figure 3 for a comparison of the ϵ-learning strategy (and the majority learning strategy) with the simple forwarding + majority learning strategy.

VI. Gaussian Learning

We now turn to a continuous analog of the teacher-student learning problem. Suppose the state of the world corresponds to a parameter $\mu \in \mathbb{R}$. The teacher receives observations $\{y_i\}_{i=1}^n$, where $y_i \sim N(\mu, \sigma_1^2)$ are i.i.d. Furthermore, at time step i, the teacher transmits an estimate $x_i = f_i(y_1, \ldots, y_i)$ to the student, who receives $z_i = x_i + \epsilon_i$, where $\epsilon_i \overset{i.i.d.}{\sim} N(0, \sigma_2^2)$. In other words, the teacher observes the state of the world through a Gaussian channel with noise variance σ_1^2, and the student observes the teacher's transmissions through a Gaussian channel with noise variance σ_2^2. The final estimate of the student after n time steps is some function $\widehat{\theta}(z_1, \ldots, z_n)$.

We again seek to compare different teaching/learning strategies, where the learning rate of the student is now characterized by the parameter pairs (σ_1, σ_2) governing the two channels, rather than the pairs (p, q). In the continuous parameter setting, we replace the notion of a majority learner (which no longer makes sense) by a learner who simply takes an average over all observations. More broadly, we allow both the teacher and student to learn by taking a linear combination of observations; i.e., the functions $\{f_i\}_{i=1}^n$ and $\widehat{\theta}$ are linear. We compare various strategies in terms of the variance $\mathrm{Var}(\widehat{\theta})$ of the final estimate of the student.

Consequently, we may reparametrize the problem according to a matrix-vector pair (A, b). If we use $x, y, z, \epsilon \in \mathbb{R}^n$ to denote the vectorized versions of the corresponding random variables, we see that the teacher receives the observation vector $y \sim N(\mu\mathbf{1}, \sigma_1^2\ I_n)$ and transmits the vector Ay. The student receives $z = Ay + \epsilon$, where $\epsilon \sim N(0, \sigma_2^2\ I_n)$, and deduces $\widehat{\theta} = b^T z$.

The quality of the student's estimator may be calculated as

$$\begin{aligned} f(A, b) &:= \mathrm{Var}(b^T z) = \mathrm{Var}(b^T A y + b^T \epsilon) \\ &= \sigma_1^2 b^T A A^T b + \sigma_2^2 b^T b. \end{aligned} \tag{13}$$

We seek to minimize f with respect to A and b.

Note that the fact that the teacher must transmit messages depending only on his past observations constrains the matrix A to be lower-triangular. Furthermore, we impose the additional assumption that the teacher's transmissions on successive days are unbiased estimators of the state of the world, so

$$\mu\mathbf{1} = \mathbb{E}[z] = \mathbb{E}[Ay] = \mu A\mathbf{1}. \tag{14}$$

Equivalently (assuming we are in an arbitrary setting where $\mu \neq 0$), we have $A\mathbf{1} = \mathbf{1}$, so A is a row-stochastic matrix. Similarly, we require the student to output an unbiased estimator, so

$$\mu = \mathbb{E}[b^T z] = \mathbb{E}[b^T(Ay + \epsilon)] = b^T A \cdot \mu\mathbf{1},$$

implying that $b^T A\mathbf{1} = 1$, so $b^T\mathbf{1} = 1$.

Due to the relatively simple form of expression (13), we can analyze the optimal student strategy for a fixed teacher strategy with relative ease. We can also determine a jointly optimal strategy in terms of the pair (A, b).

Remark: We briefly remark on the condition (14) that the teacher must transmit unbiased estimators of the state of the world. Note that if this were not the case, the teacher and the student could agree a priori that the teacher would simply amplify each observation by a large constant M, and the student would rescale the received transmissions by $\frac{1}{M}$. Thus, the student would receive the vector $z = My + \epsilon$, which would be transformed to $z' = y + \frac{1}{M}\epsilon$. As $M \to \infty$, this would correspond to a noiseless channel from the teacher to the student. By the Cramer-Rao bound, the minimum variance unbiased estimate for the teacher based on n i.i.d. observations $\{y_i\}_{i=1}^n$ is the average, which has variance $\frac{\sigma_1^2}{n}$. Furthermore, it is possible to achieve this lower bound by taking $A = MI_n$ and $b = \frac{1}{Mn}\mathbf{1}$.

A. Optimal Student Strategy

We first consider the optimal strategy of the student when the strategy of the teacher (corresponding to the matrix A) is fixed. We have the following result:

Theorem 5: Let A be a fixed row-stochastic, lower-triangular matrix. The student strategy that minimizes the variance $f(A, b)$ of the estimator is given by

$$b^* = \frac{(\sigma_1^2\ AA^T + \sigma_2^2\ I)^{-1}\mathbf{1}}{\mathbf{1}^T(\sigma_1^2\ AA^T + \sigma_2^2\ I)^{-1}\mathbf{1}},$$

resulting in a variance of

$$f(A, b^*) = \frac{1}{\mathbf{1}^T(\sigma_1^2\ AA^T + \sigma_2^2 I)^{-1}\mathbf{1}}.$$

Proof: The optimal strategy of the student may be obtained by optimizing the expression (13):

$$\begin{aligned} \min_b \quad & b^T(\sigma_1^2\ AA^T + \sigma_2^2\ I)b \\ \text{s.t.} \quad & b^T\mathbf{1} = 1. \end{aligned}$$

We can optimize this using the method of Lagrange multipliers. Define the function

$$g(b, \lambda) = b^T(\sigma_1^2\ AA^T + \sigma_2^2\ I)b + \lambda(b^T\mathbf{1} - 1).$$

Then

$$\frac{\partial g}{\partial b} = 2(\sigma_1^2\ AA^T + \sigma_2^2\ I)b + \lambda\mathbf{1},$$

so setting $\frac{\partial g}{\partial b} = 0$ and solving for b gives

$$b = \frac{-\lambda}{2}(\sigma_1^2\ AA^T + \sigma_2^2\ I)^{-1}\mathbf{1}.$$

Hence, setting $b^T\mathbf{1} - 1 = 0$ implies that

$$\frac{-\lambda}{2}\mathbf{1}^T(\sigma_1^2\ AA^T + \sigma_2^2\ I)^{-1}\mathbf{1} - 1 = 0,$$

so

$$\lambda = \frac{-2}{\mathbf{1}^T(\sigma_1^2\ AA^T + \sigma_2^2\ I)^{-1}\mathbf{1}},$$

implying that

$$b = \frac{(\sigma_1^2 AA^T + \sigma_2^2 I)^{-1}\mathbf{1}}{\mathbf{1}^T(\sigma_1^2 AA^T + \sigma_2^2 I)^{-1}\mathbf{1}}. \tag{15}$$

The variance of the optimal strategy is then obtained by plugging the value of b into the variance formula (13), to obtain

$$\frac{1}{\mathbf{1}^T(\sigma_1^2 AA^T + \sigma_2^2 I)^{-1}\mathbf{1}}. \tag{16}$$

□

Note that although we focused on fixed (and relatively simplistic) learning strategies for the student in the binary case, the simpler expressions for the variance of the student's estimators allow us to derive the optimal strategy that a student should employ for a particular teaching mechanism. Indeed, different values of the teacher's matrix A determine the relative weights of the vector $(\sigma_1^2\ AA^T + \sigma_2^2\ I)^{-1}\mathbf{1}$, which govern how the student should weight her observations as time progresses. In the case $A = I$, the student's optimal strategy would be to place equal weight on all observations (i.e., simple averaging). Note that the student's optimal strategy will never correspond to a weighted average of her last ϵn observations, since the teacher's transmissions in the first $(1 - \epsilon)n$ time steps are assumed to be unbiased estimators of μ, so an optimal student learner would always prefer to compute a weighted average that takes into account these initial observations, as well.

B. Teacher Strategies

As analogs to the strategies studied in the binary teaching/learning setting, we now discuss the following classes of teacher strategies, and provide a brief comparison of the relative quality of the ensuing estimators computed by an optimal student.

1) **Simple forwarding:** This corresponds to $A = I_n$.
2) **ϵ-teaching:** The teacher first learns for $(1-\epsilon)n$ steps, and then transmits his best estimate based on the first $(1-\epsilon)n$ observations, for the remaining ϵn steps. The teacher's best estimate corresponds to $\frac{y_1+\cdots+y_{\lfloor(1-\epsilon)n\rfloor}}{\lfloor(1-\epsilon)n\rfloor}$, so A is a block matrix with all entries in the lower left $\epsilon n \times \lfloor(1-\epsilon)n\rfloor$ block equal to $\frac{1}{\lfloor(1-\epsilon)n\rfloor}$, and the remaining entries equal to 0.
3) **Cumulative learning:** The teacher transmits his best estimate at each step. This corresponds to A being a lower-triangular matrix with all entries in column i equal to $\frac{1}{i}$.

In the simple forwarding teaching strategy, the student receives i.i.d. observations $z_i \sim N(\mu, \sigma_1^2 + \sigma_2^2)$, and the minimum-variance strategy is clearly for the student to take a simple average. We can also see this by plugging $A = I$ into the formula (15):

$$b = \frac{\frac{1}{\sigma_1^2+\sigma_2^2}I\mathbf{1}}{\mathbf{1}^T\frac{1}{\sigma_1^2+\sigma_2^2}I\mathbf{1}} = \frac{1}{n}\mathbf{1}.$$

The variance of the overall estimator is then equal to $\frac{\sigma_1^2+\sigma_2^2}{n}$.

We now make two surprising observations: First, suppose the teacher employs the ϵ-teaching strategy, and the student simply averages the latter ϵn observations. As shown above, this may not agree with the optimal strategy for the student; however, this leads to closed-form expressions and a simple comparison. This is analogous to the strategy in the binary

setting where the student takes a simple majority over the latter ϵn observations. Plugging into the formula (13), it is not hard to see that the variance of this strategy is

$$\frac{\sigma_1^2}{(1-\epsilon)n} + \frac{\sigma_2^2}{\epsilon n},$$

since the matrix AA^T has all entries equal to 0 other than the lower right $\epsilon n \times \epsilon n$ block, which has all entries equal to $\frac{1}{(1-\epsilon)n}$. Clearly, this expression is strictly larger than $\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}$. Thus, contrary to the conclusions in the binary setting, the simple forwarding teacher strategy *always* dominates the ϵ-teaching + simple averaging strategy.

Second, suppose the teacher is clairvoyant, and is allowed to transmit estimates of the state of the world based on his entire observation vector y, rather than only the observations seen up to time i. (As an alternative interpretation, suppose the teacher first learned from n observations in a previous epoch, prior to teaching.) The "best" strategy is intuitively to set $A = \frac{1}{n}\mathbf{1}\mathbf{1}^T$, corresponding to repeatedly transmitting the maximum likelihood estimator for μ. Based on the formula (16), we can see that the optimal student strategy actually produces the same variance as the best estimator in the case of simple forwarding: Note that A is doubly stochastic, so $\mathbf{1}$ is clearly an eigenvector of $(\sigma_1^2\ AA^T + \sigma_2^2\ I)$, and we can easily check that

$$(\sigma_1^2\ AA^T + \sigma_2^2\ I)\mathbf{1} = (\sigma_1^2 + \sigma_2^2)\mathbf{1}.$$

But then

$$\frac{1}{\sigma_1^2 + \sigma_2^2}\mathbf{1} = (\sigma_1^2\ AA^T + \sigma_2^2\ I)^{-1}\mathbf{1},$$

so we have

$$\mathbf{1}^T(\sigma_1^2\ AA^T + \sigma_2^2\ I)^{-1}\mathbf{1} = \frac{n}{\sigma_1^2 + \sigma_2^2},$$

from which the claim follows. This is markedly different from the binary learning setting.

C. Jointly Optimal Strategies

The calculations from the previous subsection suggest that the simple forwarding teacher strategy may sometimes be dominant. Indeed, we now prove that this strategy is *always* jointly optimal:

Theorem 6: The joint optimization problem

$$\begin{aligned} \min_{A,b} & f(A,b) \\ \text{s.t.} \quad & A\mathbf{1} = \mathbf{1}, \\ & b^T\mathbf{1} = 1, \end{aligned}$$

is minimized when $A^* = I_n$ and $b^* = \frac{1}{n}\mathbf{1}$.

Proof: Denote the set of parameters

$$\Theta := \left\{(A,b) : A\mathbf{1} = \mathbf{1}, b^T\mathbf{1} = 1\right\},$$

and for any $\kappa > 0$, define the set

$$\Theta_\kappa := \left\{(A,b) : \|b\|_2^2 \geq \kappa, b^T A\mathbf{1} = 1\right\}.$$

Note that for $(A,b) \in \Theta$, we have $\|b\|_1 \geq b^T\mathbf{1} = 1$, so

$$\|b\|_2^2 \geq \frac{1}{n}\|b\|_1^2 \geq \frac{1}{n},$$

implying that $\Theta \subseteq \Theta_\kappa$ for $\kappa \geq \frac{1}{n}$.

Now note that

$$\min_{(A,b)\in\Theta_\kappa} f(A,b) \geq \min_{(A,b)\in\Theta_\kappa} \sigma_1^2\ b^T AA^T b + \sigma_2^2\kappa.$$

Furthermore, if we define $w = A^T b$, we see that the expression $\sigma_1^2\ b^T AA^T b$ is simply the variance of the estimator $w^T y$ of μ, where the condition that $b^T A\mathbf{1} = 1$ simply constrains $w^T y$ to be an unbiased estimator. By the Cramer-Rao bound, we therefore conclude that

$$\min_{(A,b)\in\Theta_\kappa} f(A,b) \geq \frac{\sigma_1^2}{n} + \sigma_2^2\kappa.$$

Finally, taking $\kappa = \frac{1}{n}$, we conclude that

$$\min_{(A,b)\in\Theta} f(A,b) \geq \min_{(A,b)\in\Theta_\kappa} f(A,b) \geq \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}. \tag{17}$$

Since this lower bound is achieved when $A = I_n$ and $b = \frac{1}{n}\mathbf{1}$, the simple forwarding + simple averaging strategy must always be a joint minimizer. □

Remark: In fact, the preceding argument only requires A to be row-stochastic, and not necessarily lower-triangular. Thus, we see that the simple forwarding teaching strategy + simple averaging learning strategy is in fact optimal even for clairvoyant teachers.

D. Generalizations

Note that the optimality argument in the previous subsection does not actually require the ϵ_i's to be Gaussian, as long as they are i.i.d. with variance σ_2^2. Some natural questions are whether the results also rely on Gaussianity of the teacher's observations and/or linearity of the teacher or student strategies.

Regarding the Gaussian assumption on the teacher's observations $\{y_i\}_{i=1}^n$, we note that if $\mathrm{Var}(z_i) = \sigma_1^2$, we similarly have the lower bound

$$\min_{(A,b)\in\Theta_\kappa} f(A,b) \geq \min_{(A,b)\in\Theta_\kappa} \sigma_1^2\ b^T AA^T b + \sigma_2^2\kappa,$$

if the teacher and student strategies are parametrized by A and b, respectively. Although it is no longer true that the Cramer-Rao lower bound is achieved for non-Gaussian data, the best linear unbiased estimator (BLUE) based on n i.i.d. samples is nonetheless still achieved by the empirical average. Hence, inequality (17) holds, with equality achieved in the case $A = I_n$, $b = \frac{1}{n}\mathbf{1}$, and $\kappa = \frac{1}{n}$, as before.

Moving to the question of whether linearity of strategies is required, note that we can generalize our optimality argument in the case when the y_i's are Gaussian to the case when the teacher is allowed to transmit *any* family of functions $\{f_i\}_{i=1}^n$ at each time step (even functions that depend on any future observations in the set $\{y_1, \ldots, y_n\}$). In such a setting, the Cramer-Rao lower bound (17) still applies to the wider class of estimators, showing that the linear/simple forwarding strategy is optimal over the entire class. The question of whether such a statement holds when the y_i's are not Gaussians remains open. We also do not currently have a characterization of the set of jointly optimal strategies when the student is allowed to employ a more sophisticated nonlinear strategy.

VII. Discussion and Open Problems

In Figure 3, we compared the learning rates for the student for various teacher strategies. Our results indicate that if the teacher to student channel has a low level of noise, it is better for the student to transmit "uncoded" information. Intuitively, the teacher might receive many incorrect observations initially by chance, in which case the cumulative teaching strategy would have a significant delay in correcting the teacher's opinion. However, if the teacher were following the simple forwarding strategy, the flipped observations in the beginning would have no effect on the teacher's future communications. Furthermore, since the student has a relatively clean channel, she does not need a cumulative teaching strategy to learn quickly. We also noted a surprising threshold of $q \approx 0.15$ that emerged from the figure: if the teacher to student channel is more noisy than this threshold, then it is *always* beneficial to use the cumulative teaching strategy, no matter how bad the teacher's channel.

Various alternative teaching and learning strategies exist that have not been analyzed here. In particular, we did not analyze Bayesian strategies for the student. Although the learning rate of such strategies could be simulated for small n, obtaining an accurate approximation of the (exponentially small) error probability from simulations for larger values of n is challenging. We also note that analyzing specific teaching and learning strategies provide lower bounds on the best possible learning rate.

An interesting line of inquiry is to characterize the optimal learning rate over all possible joint strategies between the teacher and student. Observe that the optimal learning rate for the teacher is $D(1/2\|p)$ and the optimal learning rate for the student, if she were to observe Θ directly through her channel, is $D(1/2\|q)$. Simple arguments using the data processing inequality show that the optimal learning rate is at most $\min(D(1/2\|p), D(1/2\|q))$. Can this bound be achieved? We think this would be very surprising, leading us to formulate the following conjecture:

Conjecture 1: Let $0 < p, q < 1/2$. We conjecture that the optimal learning rate of the student over all possible joint strategies between the teacher and the student is strictly less than $\min(D(1/2\|p), D(1/2\|q))$.

It is interesting to note that if the teacher is allowed to have *non-causal* strategies, the upper bound of $\min(D(1/2\|p), D(1/2\|q))$ may be achieved: The teacher can use all n time slots to learn Θ, and then transmit the learned value over n time slots to the student. The conjecture above essentially states that a price must be paid for using *causal* teaching and learning strategies (at least in the binary case, since our analysis shows that no such penalty exists in the Gaussian setting). A harder open problem is to determine optimal joint strategies that the teacher and student could employ to maximize the student's learning rate; note that since the student will always be Bayesian in the optimal strategy, this problem boils down to identifying an optimal teaching strategy.

We also state the fascinating conjecture from Huleihel *et al.* [29] alluded to in Section II, which concerns identifying the optimal learning rate:

Conjecture 2 (Huleihal et al. [29]): Let $p = q$. In the regime $p \to 1/2$, the optimal learning rate of the student is $D(1/2\|p)(1 + o(1))$; i.e., the optimal learning rate is the same as that of the teacher up to first-order error terms.

In the case of the Gaussian learner, we showed that the teacher-student problem is of an entirely different nature, and the simplest strategy where the teacher simply forwards information and the student constructs a simple average is always optimal over the class of linear, unbiased estimators. However, we also showed that if the teacher is allowed to transmit estimators that are not unbiased, which the student subsequently decodes, the overall estimator can have an even smaller variance. In general, the question of the best estimator over different classes of teaching/learning strategies (e.g., non-linear strategies, or biased strategies with appropriate power constraints) remains open.

Finally, we concede that the teacher/student model analyzed in this paper is a vast oversimplification of reality, and the dynamics of social learning may be substantially different in practice. The simple forwarding and ϵ-teaching strategies employed by the teacher have a "learn by rote" flavor, which any experienced teacher would realize is not the best way to convey information to a student: artful teaching involves presenting concepts from different angles, rather than simply repeating the same lesson from one day to the next. Furthermore, we have assumed that no feedback is available to the teacher from the student, whereas a teacher should be able to adapt his strategy based on how well the student is learning. A fascinating new framework in which the teacher feeds the student carefully crafted examples from a set of lessons is known as machine teaching [32]. It is also unreasonable to expect that the student (or teacher) would have knowledge of the noise parameters of the channels beforehand, and a more realistic setting might involve gradually estimating these parameters based on feedback and adapting strategies over time. After all, a student would rightfully choose to pay less attention to a teacher if she thinks he does not know what he is talking about!

Appendix A
Proof of Lemma 1

It is enough to show that $\mathbb{E}e^{\lambda_1 \tilde{T} + \lambda_2 |\tilde{T}|} \leq \frac{1}{2p}$. Using the generating function (2), we may calculate

$$\begin{aligned}
\mathbb{E}e^{\lambda_1 \tilde{T} + \lambda_2 |\tilde{T}|} &= \sum_{k=1}^{\infty} P(\tilde{T} = 2k)\left(e^{(\lambda_1+\lambda_2)2k} + e^{(\lambda_2-\lambda_1)2k}\right) \\
&= \sum_{k=1}^{\infty} \frac{1}{2p}(p\bar{p})^k C_{k-1}\left(e^{(\lambda_1+\lambda_2)2k} + e^{(\lambda_2-\lambda_1)2k}\right) \\
&= \frac{\bar{p}}{2}\left(e^{2(\lambda_1+\lambda_2)}\sum_{k=0}^{\infty} C_k\left(p\bar{p}e^{2(\lambda_1+\lambda_2)}\right)^k \right. \\
&\qquad \left. + e^{2(\lambda_2-\lambda_1)}\sum_{k=0}^{\infty} C_k\left(p\bar{p}e^{2(\lambda_2-\lambda_1)}\right)^k\right)
\end{aligned}$$

$$= \frac{\bar{p}}{2}\left(e^{2(\lambda_1+\lambda_2)} f\left(p\bar{p}e^{2(\lambda_1+\lambda_2)}\right) + e^{2(\lambda_2-\lambda_1)} f\left(p\bar{p}e^{2(\lambda_2-\lambda_1)}\right)\right)$$
$$= \frac{2-\sqrt{1-4p\bar{p}e^{2(\lambda_1+\lambda_2)}}-\sqrt{1-4p\bar{p}e^{2(\lambda_2-\lambda_1)}}}{4p}.$$

Suppose $\lambda_1+\lambda_2 > D(1/2\|p)$. Then

$$\lambda_1+\lambda_2 \geq \frac{1}{2}\log\left(\frac{1}{2p}\right)+\frac{1}{2}\log\left(\frac{1}{2\bar{p}}\right)=\frac{1}{2}\log\left(\frac{1}{4p\bar{p}}\right).$$

Thus,

$$4p\bar{p}e^{2(\lambda_1+\lambda_2)} > 1,$$

so the moment generating function is undefined, and $L(\lambda_1,\lambda_2)=+\infty$. We can argue similarly if $-\lambda_1+\lambda_2 > D(1/2\|p)$.

Now consider $(\lambda_1,\lambda_2)\in\mathcal{D}$. We have

$$\lambda_1+\lambda_2 \leq D(1/2||p),$$
$$-\lambda_1+\lambda_2 \leq D(1/2||p).$$

It is easy to see that for fixed p, the maximum value of the moment generating function expression is attained when $\lambda_1=0$ and $\lambda_2=D(1/2||p)$, and that this value is $1/2p$. This concludes the proof.

Appendix B
Proof of Lemma 2

We may rewrite the probability as $\mathbb{P}(B>n)\mathbb{P}(M_n \leq n(1-\delta)|B>n)$. Furthermore, conditioned on the event $\{B>n\}$, the random variable $\frac{M_n}{n}$ has a symmetric distribution around $1/2$, since the mirror image of every path up to time n has the exact same probability as the original path when conditioned on the event $\{B>n\}$. Thus,

$$\frac{1}{2} \leq \mathbb{P}\left(\frac{M_n}{n} \leq (1-\delta)\Big| B>n\right) \leq 1,$$

implying that

$$\mathbb{P}(M_n \leq n(1-\delta), B>n) = \Theta\Big(\mathbb{P}(B>n)\Big).$$

We now upper-bound

$$\begin{aligned}\mathbb{P}(B>n) &= \sum_{i=n/2}^{\infty}\mathbb{P}(B=2i)\\ &= \sum_{i=n/2}^{\infty}\binom{2i}{i}p^i(1-p)^i(1-2p)\\ &\leq \sum_{i=n/2}^{\infty}2^{2i}p^i(1-p)^i(1-2p)\\ &= (4p\bar{p})^{n/2}\sum_{i=0}^{\infty}(4p\bar{p})^i(1-2p)\\ &= (4p\bar{p})^{n/2}\cdot\frac{1-2p}{1-4p\bar{p}}\\ &= (4p\bar{p})^{n/2}\cdot\frac{1}{1-2p}\\ &= e^{-n(D(1/2||p))}\cdot\frac{1}{1-2p}.\end{aligned}$$

Furthermore, letting $m=\lceil\frac{n+1}{2}\rceil$, we have

$$\begin{aligned}\mathbb{P}(B>n) \geq \mathbb{P}(B=2m) &= \binom{2m}{m}p^m(1-p)^m(1-2p)\\ &\geq \frac{2^{2m}}{2m}p^m(1-p)^m(1-2p)\\ &\geq \frac{4^{n/2}}{n+1}(p\bar{p})^{n/2}p\bar{p}(1-2p)\\ &= e^{-n(D(1/2\|p))}\cdot\frac{p\bar{p}(1-2p)}{n+1}.\end{aligned}$$

Combining the upper and lower bounds, we clearly have $\mathbb{P}(B>n)=e^{-n(D(1/2||p)+o(1))}$, as claimed.

Appendix C
Proof of Lemma 3

Consider the event $\{G=1 \text{ and } \tilde{T}_1 < -\lceil n\delta\rceil\}$. This corresponds to the event that there is only one return to 0, but sojourn time is at least $n\delta$ and the sojourn is on the negative side of the integers. If this event occurs, then M_n can at most be $n(1-\delta)$, giving us the lower bound

$$\begin{aligned}\mathbb{P}(M_n \leq n(1-\delta)) &\geq \mathbb{P}(G=1,\tilde{T}_1 \leq \lfloor -n\delta\rfloor)\\ &\geq \mathbb{P}(G=1,\tilde{T}_1=-n\tilde{\delta}_n),\end{aligned}$$

where $\tilde{\delta}_n$ is such that $-n\tilde{\delta}_n$ is an even integer that is at most $\lfloor -n\delta\rfloor$. Clearly, we can take $\delta_n\to\delta$ as $n\to\infty$. Continuing, we have

$$\begin{aligned}&\mathbb{P}(G=1,\tilde{T}_1=-n\tilde{\delta}_n)\\ &\quad= (1-2p)\times\frac{1}{n\tilde{\delta}_n/2+1}\binom{n\tilde{\delta}_n-2}{n\tilde{\delta}_n/2-1}(p\bar{p})^{n\tilde{\delta}_n/2}.\end{aligned}$$

Taking logarithms, dividing by n, and taking the $\liminf$ as n tends to infinity, we obtain

$$\begin{aligned}&\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}(Z_n\in\mathcal{Q})\\ &\quad\geq \liminf_{n\to\infty}\frac{1}{n}\log\left((1-2p)\times\frac{1}{n\tilde{\delta}_n/2+1}\binom{n\tilde{\delta}_n-2}{n\tilde{\delta}_n/2-1}(p\bar{p})^{n\tilde{\delta}_n/2}\right)\\ &\quad= \liminf_{n\to\infty}\frac{1}{n}\log\binom{n\tilde{\delta}_n-2}{n\tilde{\delta}_n/2-1}(p\bar{p})^{n\tilde{\delta}_n/2}\\ &\quad= \liminf_{n\to\infty}\left(\tilde{\delta}_n\log 2+\frac{\tilde{\delta}_n}{2}\log(p\bar{p})\right)\\ &\quad= -\delta\cdot\frac{1}{2}\log\frac{1}{4p\bar{p}}\\ &\quad= -\delta D(1/2||p).\end{aligned}$$

Appendix D
Proof of Lemma 4

Using the independence of the U_i's and V_j's, we may compute the limit

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}e^{n\lambda W_n} &= \lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}e^{\lambda\left(\sum_{i=1}^{n-\lfloor n\theta\rfloor}U_i+\sum_{j=1}^{\lfloor n\theta\rfloor}V_j\right)}\\ &= \lim_{n\to\infty}\frac{1}{n}\Big((n-\lfloor n\theta\rfloor)\log(e^{\lambda}\bar{q}+q)\\ &\qquad+\lfloor n\theta\rfloor\log(e^{\lambda}q+\bar{q})\Big)\\ &= \bar{\theta}\log(e^{\lambda}\bar{q}+q)+\theta\log(e^{\lambda}q+\bar{q}).\end{aligned}$$

Thus, we may use the Gärtner-Ellis theorem to conclude that W_n satisfies the large deviation principle with rate function

$$\begin{aligned} I_\theta(w) &= \left(\bar{\theta}\log(e^\lambda \bar{q}+q)+\theta\log(e^\lambda q+\bar{q})\right)^*(w) \\ &= \sup_\lambda \left\{\lambda w - \bar{\theta}\log(e^\lambda\bar{q}+q)-\theta\log(e^\lambda q+\bar{q})\right\}. \end{aligned}$$

Differentiating with respect to λ, we see that the above supremum is attained when the following equality is satisfied:

$$w = \frac{\bar{\theta}e^\lambda\bar{q}}{e^\lambda\bar{q}+q}+\frac{\theta e^\lambda q}{e^\lambda q+\bar{q}}. \tag{18}$$

This is a quadratic equation in e^λ, which we may solve to obtain

$$e^\lambda = \frac{-\tau(\theta,w)+\sqrt{\tau(\theta,w)^2+4w\bar{w}}}{2\bar{w}},$$

where

$$\tau(\theta,w) := \frac{\bar{q}}{q}(\bar{\theta}-w)+\frac{q}{\bar{q}}(\theta-w).$$

The rate function is then given by

$$\begin{aligned} I_\theta(w) = {} & w\log\left(\frac{-\tau(\theta,w)+\sqrt{\tau(\theta,w)^2+4w\bar{w}}}{2\bar{w}}\right) \\ & -\bar{\theta}\log\left(\bar{q}\left(\frac{-\tau(\theta,w)+\sqrt{\tau(\theta,w)^2+4w\bar{w}}}{2\bar{w}}\right)+q\right) \\ & -\theta\log\left(q\left(\frac{-\tau(\theta,w)+\sqrt{\tau(\theta,w)^2+4w\bar{w}}}{2\bar{w}}\right)+\bar{q}\right). \end{aligned}$$

As a sanity check, when $\theta = 0$, we have $e^\lambda = \frac{qw}{\bar{q}\bar{w}}$, and the rate function is

$$\begin{aligned} I_\theta(w) &= w\log\frac{qw}{\bar{q}\bar{w}}-\log\left(\frac{qw}{\bar{w}}+q\right) \\ &= w\log\frac{qw}{\bar{q}\bar{w}}-\log\left(\frac{q}{\bar{w}}\right) \\ &= D(w||\bar{q}), \end{aligned}$$

which is what we expect. We also note that when $w = \bar{q}\bar{\theta}+q\theta$, the solution to equation (18) is $e^\lambda = 1$, which gives $I_\theta(w) = 0$.

ACKNOWLEDGMENT

The authors would like to thank the Associate Editor and anonymous reviewers for their comments and suggestions, which led to an improved manuscript. They would also like to thank the organizers of the Fourteenth Annual Workshop on Probability and Combinatorics in Barbados, where extensions to the ϵ-learning rate and Gaussian learning were derived.

REFERENCES

[1] C. P. Chamley, *Rational Herds: Economic Models of Social Learning*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[2] E. Mossel and O. Tamuz, "Opinion exchange dynamics," *Probab. Surv.*, vol. 14, pp. 155–204, 2017. [Online]. Available: https://www.i-journals.org/ps/viewissue.php?id=17
[3] X. Vives, "How fast do rational agents learn?" *Rev. Econ. Stud.*, vol. 60, no. 2, pp. 329–347, 1993.
[4] A. Jadbabaie, P. Molavi, and A. Tahbaz-Salehi, "Information heterogeneity and the speed of learning in social networks," Columbia Bus. School, Tech. Rep., 2013.
[5] P. Molavi, A. Tahbaz-Salehi, and A. Jadbabaie, "Foundations of non-Bayesian social learning," Columbia Bus. School, Tech. Rep., 2017.
[6] M. Harel, E. Mossel, P. Strack, and O. Tamuz, "Rational groupthink," *Quart. J. Econ.*, vol. 7, Jul. 2020, Art. no. qjaa026.
[7] M. Mobius and T. Rosenblat, "Social learning in economics," *Annu. Rev. Econ.*, vol. 6, no. 1, pp. 827–847, 2014.
[8] R. J. Aumann, "Agreeing to disagree," *Ann. Statist.*, vol. 4, no. 6, pp. 1236–1239, Nov. 1976.
[9] J. D. Geanakoplos and H. M. Polemarchakis, "We can't disagree forever," *J. Econ. Theory*, vol. 28, no. 1, pp. 192–200, Oct. 1982.
[10] A. V. Banerjee, "A simple model of herd behavior," *Quart. J. Econ.*, vol. 107, no. 3, pp. 797–817, Aug. 1992.
[11] S. Bikhchandani, D. Hirshleifer, and I. Welch, "A theory of fads, fashion, custom, and cultural change as informational cascades," *J. Political Economy*, vol. 100, no. 5, pp. 992–1026, Oct. 1992.
[12] L. Smith and P. Sorensen, "Pathological outcomes of observational learning," *Econometrica*, vol. 68, no. 2, pp. 371–398, Mar. 2000.
[13] D. Gale and S. Kariv, "Bayesian learning in social networks," *Games Econ. Behav.*, vol. 45, no. 2, pp. 329–346, Nov. 2003.
[14] E. Mossel, A. Sly, and O. Tamuz, "Asymptotic learning on Bayesian social networks," *Probab. Theory Rel. Fields*, vol. 158, nos. 1–2, pp. 127–157, Feb. 2014.
[15] M. H. DeGroot, "Reaching a consensus," *J. Amer. Statist. Assoc.*, vol. 69, no. 345, pp. 118–121, Mar. 1974.
[16] G. Ellison and D. Fudenberg, "Rules of thumb for social learning," *J. Political Economy*, vol. 101, no. 4, pp. 612–643, Aug. 1993.
[17] V. Bala and S. Goyal, "Learning from neighbours," *Rev. Econ. Stud.*, vol. 65, no. 3, pp. 595–621, Jul. 1998.
[18] M. Amin Rahimian and A. Jadbabaie, "Bayesian heuristics for group decisions," 2016, *arXiv:1611.01006*. [Online]. Available: http://arxiv.org/abs/1611.01006
[19] J. Hazła, A. Jadbabaie, E. Mossel, and M. A. Rahimian, "Bayesian decision making in groups is hard," *Oper. Res.*, to be published.
[20] Y. Kanoria and O. Tamuz, "Tractable Bayesian social learning on trees," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 4, pp. 756–765, Apr. 2013.
[21] Y.-C. Ho, "Team decision theory and information structures," *Proc. IEEE*, vol. 68, no. 6, pp. 644–654, Jun. 1980.
[22] H. S. Witsenhausen, "A counterexample in stochastic optimum control," *SIAM J. Control*, vol. 6, no. 1, pp. 131–147, Feb. 1968.
[23] E. S. Andersen, "Survey of fluctuation theory," *Adv. Appl. Probab.*, vol. 6, no. 2, pp. 213–214, Jun. 1974.
[24] L. Addario-Berry and B. A. Reed, "Ballot theorems, old and new," in *Horizons of Combinatorics*. Berlin, Germany: Springer, 2008, pp. 9–35.
[25] K. L. Chung and W. Feller, "On fluctuations in coin-tossing," *Proc. Nat. Acad. Sci. USA*, vol. 35, no. 10, pp. 605–608, 1949.
[26] E. S. Andersen, "On the fluctuations of sums of random variables II," *Math. Scandinavica*, vol. 2, no. 2, pp. 195–223, Feb. 1955.
[27] P. Lévy, "Sur certains processus stochastiques homogènes," *Compositio Math.*, vol. 7, pp. 283–339, 1940.
[28] R. Durrett, *Probability: Theory, and Examples*, vol. 49. Cambridge, U.K.: Cambridge Univ. Press, 2019.
[29] W. Huleihel, Y. Polyanskiy, and O. Shayevitz, "Relaying one bit across a tandem of binary-symmetric channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2019, pp. 2928–2932.
[30] R. P. Stanley, *Catalan Numbers*. Cambridge, U.K.: Cambridge Univ. Press, 2015.
[31] P. Dupuis and R. S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*, vol. 902. Hoboken, NJ, USA: Wiley, 2011.
[32] X. Zhu, "Machine teaching: An inverse problem to machine learning and an approach toward optimal education," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 4083–4087.

Varun Jog received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Bombay, in 2010, and the Ph.D. degree in electrical engineering and computer sciences from the University of California, Berkeley, in 2015.

From 2015 to 2016, he was a Post-Doctoral Scholar with the Department of Statistics, Wharton School of the University of Pennsylvania. Since 2016, he has been an Assistant Professor with the Department of Electrical and Computer Engineering, University of Wisconsin-Madison. He is a recipient of the 2015 Eli Jury Award from the EECS Department at UC Berkeley and a recipient of the 2015 IEEE Jack Keil Wolf ISIT Student Paper Award. He also received an NSF CAREER Award in 2020.

Po-Ling Loh received the B.S. degree in mathematics, with a minor in English, from the California Institute of Technology, Pasadena, CA, USA, in 2009, and the M.S. degree in computer science and the Ph.D. degree in statistics from the University of California, Berkeley, CA, USA, in 2013 and 2014, respectively.

From 2014 to 2016, she was an Assistant Professor with the Department of Statistics, Wharton School of the University of Pennsylvania. From 2016 to 2019, she was an Assistant Professor with the Department of Electrical and Computer Engineering, University of Wisconsin-Madison, where she has been an Associate Professor with the Department of Statistics since 2019. From 2019 to 2020, she was also a Visiting Associate Professor with the Department of Statistics, Columbia University. She is a recipient of a Student Paper Award at the NIPS Conference in 2012 and the 2014 Erich L. Lehmann Citation from the Statistics Department at UC Berkeley. She has also received an NSF CAREER Award (2018), Bernoulli Society New Researcher Award (2019), IMS Tweedie New Researcher Award (2019), and the Army Research Office Young Investigator Program Award (2019).