Crowd Learning without Herding: A Mechanism Design Approach

Jacob Glazer
Department of Economics
University of Warwick
and
Faculty of Management
Tel Aviv University
email: glazer@post.tau.ac.il

Ilan Kremer
Department of Economics
University of Warwick
and
Faculty of Management and Economics Department
The Hebrew University of Jerusalem
email: ikremer@huji.ac.il

Motty Perry
Department of Economics
University of Warwick
email: motty@huji.ac.il

October 2015

Abstract

We study a mechanism design version of the informational cascade model where acquiring a signal is costly. Agents do not observe the actions of previous agents. Instead, they obtain information from a welfare maximizing central planner who is informed about the history. Agents move sequentially, decide whether to purchase a signal, and then determine what action to take.

The planner faces a tradeoff: a policy of full disclosure may induce an inefficient cascade, while hiding information may result in wasteful information acquisition and wrong actions. We characterize the planner's optimal disclosure mechanism.

1

2 Introduction

We study a mechanism design version of the informational cascade model where acquiring a signal is costly. In contrast to the classical models, agents do not observe the actions of previous agents. Instead, they obtain information from a central planner who is informed about the history, and can choose how much of this information to reveal. Agents move sequentially, with each agent deciding whether to purchase a signal, and then - after receiving information from the planner - what action to take. The planner observes the agents' actions but not their payoffs. Furthermore, the planner does not observe the agents' decisions regarding whether to acquire a signal (nor their realizations).

The planner, whose objective is to maximize the present value of all agents' payoffs, faces the following tradeoff: if he applies a policy of full disclosure, an inefficient cascade may emerge, whereas if he chooses not to reveal all the information he possesses, some agents may end up taking the wrong action. Moreover, because signals are costly, if the planner chooses not to provide (sufficient) information "too many" agents may end up acquiring a signal. What makes the planner's task especially difficult is that when the information structure is asymmetric (i.e., in some states signals are more informative), remaining silent is, by itself, informative, and this affects agents' choices. In contrast, as we discuss later, in the case of a symmetric information structure, the planner's task is rather trivial, and he can easily achieve the first-best outcome.

Models of dynamic learning can get complicated very fast. In view of this, we chose to study a model with a simple information structure, yet rich enough to shed light on the essential characteristics of the planner's optimal disclosure policy. The main insights that we obtain are relevant to a number of real-life situations. Consider the case of crowdfunding platforms that observe investors' decisions, and choose how much of this information to reveal to future investors. As more information is revealed, incentives diminish for future investors to invest time (and money) in gathering information, and the quality of the information that can be revealed later also deteriorates. Internet recommendation websites that follow consumers' choices, and use this information to inform future consumers also face a similar tradeoff. A medical director of large healthcare organization who can record doctors' prescriptions, and use this information to provide recommendations to his doctors is another example of the situation we analyze. These planners (platforms) can act strategically in deciding how much information to reveal to each agent about other agents' decisions — and when to reveal such information. In all the cases mentioned above, other, sometimes context-specific elements play an important role. For example, in the case of crowdfunding, investors' decisions about when and how much to invest are usually endogenous, and in cases involving health care, ethical issues may reduce the ability of medical directors to hide information. Nonetheless, our paper provides key insights about the optimal policy of these planners.

To study these interactions we consider a model with two states of nature, A and B, and two possible actions, a and b. Action a is optimal in state A, while b is optimal in state B. There is a sequence of agents, each of whom needs to choose an action without knowing the state of the world. Agents cannot observe the choices that their predecessors have made. However, they can learn about the previous actions from a social planner, who observes all agents' actions, and chooses how much of the information to share with each agent. After receiving the information provided by the planner, and before deciding between the two actions, each agent can purchase a binary signal about the state of the world. We focus on the asymmetric case where one realization of the signal fully reveals that the state is B, while the other realization increases the probability that the state is A. The agent's decision on whether to purchase a signal and the realization of such a signal are unobservable by the planner. Thus, the only information available to the planner is the agent's action a or b. The social welfare-maximizing planner commits ex ante to a mechanism that specifies how much information to reveal to each agent as a function of the history of actions the planner has observed.

In the asymmetric setup, the fact that agents are rational and understand the planner's strategy makes the planner's task difficult. Suppose, that the planner's policy is to remain silent, and, by his silence, to induce agents to purchase a signal, and act according to it until he (the planner) becomes sufficiently confident about the state of nature. If the setup were symmetric, such a policy could be easily implemented because agents do not update their beliefs when the planner remains silent. However, our information structure implies that the probability that the planner becomes confident that the state is B, after a fixed number of periods, is significantly higher than the probability that he becomes confident that the state is A. In such a case, when the planner is silent, the agents update their beliefs in favor of state A. This may lead agents to take the action a, without acquiring a signal —∈ contrary to the action the planner would have preferred they take.

Our main results are:

- Without loss of generality the planner can make do with a "recommendation mechanism." In such a mechanism, at the beginning of every period, the planner recommends that the agent either acquire a signal, and follow it, or that the agent take one of the two actions without acquiring a signal[1].
- The optimal mechanism consists of three phases. In the first phase, a policy of full disclosure is adopted. In the second phase, the planner adopts a policy wherein with positive probability he recommends that some agents acquire signals - not only in cases where acquiring a signal is socially desirable, but also in cases with histories for which this is not the case. In the third stage, the planner discloses all the information and the agents do not acquire a signal, and choose the same action.

[1] Note that this result does not follow from Myerson (1986) because in our model, once the agent receives a recommendation, he can acquire an informative signal that might justify another round of exchange of information. Nevertheless, we show that this is never optimal in our model.

- The planner's recommendation as well as the length of each phase involve the planner undertaking some delicate randomization. This depends on, among other things, the agents' actions history, which the planner has observed.
- Because the optimal mechanism may require that the planner withhold valuable information, such a mechanism can not be implemented without the planner being able to commit to it ex ante.
- In contrast to many other cases of optimal information revelation, the optimal mechanism can be implemented as a fully public mechanism; that is, none of the messages exchanged between the planner and agents are conducted privately.

The intuition behind the structure of the optimal mechanism is as follows. At the beginning there is no conflict of interest between the planner and the agents; that is, in a world of complete information, the agents will find it optimal to acquire a signal for precisely the same histories (of previous agents' actions) that the planner would like them to acquire. The situation changes dramatically in the second phase. In this phase, had the agents have been able to observe the planner's information, there are histories in which the agents would have preferred to take one of the two actions without acquiring a signal. However, the planner may not be confident enough, and, thus, will want agents to acquire more signals. The planner's policy in this phase is to keep the agents sufficiently in the dark so that acquiring a signal becomes incentive compatible. Such a strategy is obviously costly because it involves wasteful exploration by some agents, as well as some of them choosing the wrong action, given the planner's information. The third stage starts when the planner finds that hiding information from the agents is too costly relative to the potential benefits of inducing them to explore.

A useful feature of the optimal mechanism is that it can be cast as a first-best solution to a modified problem where the cost of exploration also incorporates the additional costs that the planner must incur in order to induce agents to acquire a signal against their will. This feature enables us to derive a clear characterization of the optimal mechanism and, in particular, to solve for the length and the exact recommended policy in each phase. We show that the amount of information accumulated is less than the amount of information that would have been accumulated in a first-best case where the planner could simply decide whether each agent will acquire a signal.

The paper is organized as follows: Section 3 provides a short literature review. Section 4 presents the model. Section 5 presents the first-best solution. Section 6 formally defines a recommendation mechanism. Section 7 provides the main theorem. The optimal mechanism described in this section is a private mechanism in the sense that messages from the planner are known only to the agent they are sent to. In Section 9 we show how to modify the optimal mechanism so that the same outcome is achieved when messages are public. Thus, the optimal policy can be implemented when guaranteeing privacy is infeasible.

We conclude in Section 10 by proving that the restriction to a recommendation mechanism is without loss of generality.

3 Related Literature

Our paper is related to the literature on informational cascades that originated with the work of Bikhchandani, Hirshleifer, and Welch (1992) and Banerjee (1992). An informational cascade occurs when it is optimal for an individual who has observed the actions of those ahead of him to follow their behavior without regard to his own information. In our model, such a cascade will take place if, for example, the planner adopts a policy of full disclosure. In a similar setup, Sgroi (2002) evaluates, from the perspective of a social planner, the strategy of forcing a subset of agents to make decisions early. Along the same line, Smith, Sorensen, and Tian (2014) conduct a welfare analysis of the herding model, showing that the efficient outcome can be decentralized by rewarding individuals if their successor mimics their action.

This paper also belongs to the literature on mechanism design without monetary transfers. One such model that shares the sequential feature of our model is that of Gershkov and Szentes (2009), who analyze a voting model in which there is no conflict of interest between voters, and information acquisition is costly. In the optimal mechanism, the social planner asks voters randomly and one at a time to invest in information, and to report the resulting signal. As in our model, the planner does not observe whether the agent invests in acquiring information, but in contrast to our model the planner makes only one decision, and all agents share the same payoff. Martimort and Aggey (2006) consider the problem of communication between a principal and a privately informed agent when monetary incentives are not available.

Other related papers are Kremer, Mansour, and Perry (2014) and Che and Horner (2014). These papers also consider a mechanism-design problem in which the planner wishes to aggregate information. However, our paper relates to a very different scenario, and the resulting optimal policies are also very different. One difference is that in Kremer, Mansour, and Perry (2014) and Che and Horner (2014) the payoffs are realized and observed by the planner immediately after an action is taken — thus making it a multi-arm-bandit problem with selfish agents. A second important difference is that the present paper focuses on the acquisition of costly information. These two differences require very different analyses and yield different results. For example, the key idea behind the mechanism in Kremer, Mansour, and Perry (2014) and Che and Horner (2014) is that the recommendations to agents are private. An agent is uncertain how many agents previously received advice to take a certain action. Without this uncertainty, there is little that the planner can achieve. However, in our paper, the optimal mechanism is based on public recommendation – a key feature that reflects the different challenge that the planner in this paper is facing.

Three recent papers that examine disclosure of information in a dynamic setup that is very different from ours are Ely, Frankel, and Kamenica (2013), Horner and Skrzypacz (2012) and Ely (2015). Also relevant are the papers by Kamenica and Gentzkow (2011) and Rayo and Segal (2010). These last two papers consider optimal-disclosure policies in a static environment in which a principal wishes to influence an agent's choice by sending the right message.

4 The Model

4.1 The Agents

There are infinitely many periods, denoted by $t = 1, 2, 3, ...$ In every period a new agent has to choose between two actions a or b. There are two possible states of the world, A and B. In state A, the optimal action is a and in state B, it is action b. The optimal action yields a payoff of 1 and the other action yields a payoff of zero. The state of the world is unknown and is determined once and for all before the beginning of period 1. Agents know their place in line but cannot observe the actions taken by (nor the utility of) the agents who moved before them. Agents' (common) prior beliefs that the state is A is p_1. Each agent is interested only in his own payoff from choosing the right action. Hereafter, we will refer to the agent who moves in period t as "agent t."

Before making his choice, agent t can obtain, at a cost $c > 0$, an informative signal about the state of the world (in Section 5.1 we discuss why in the case of $c = 0$ the problem analyzed in this paper becomes uninteresting). We denote the choice of obtaining a signal by e (for "explore"). The signal can receive one of two values, s^a or s^b. If the information structure is fully symmetric in the sense that both states are equally likely and the signals carry the same amount of information, then the planner's problem becomes rather trivial (see more on this in Section 5.1). To make the problem interesting we assume an asymmetric structure. That is, if the state is A, the signal receives the value s^a with probability 1 and if the state is B, the signal receives the value s^b with probability q, and the value s^a with probability $1 - q$. Thus the signals are informative in the sense that, absent any other information, upon observing the signal s^i, $i \in \{a, b\}$, an agent's posterior that the optimal action is i increases. Notice that signal s^b fully reveals that the state is B. We say that an agent *follows the signal* if he chooses a after observing the signal s^a and if he chooses b after observing the signal s^b.

4.2 The Planner

We assume the existence of a planner who observes only the choices, a or b, made by each agent. The planner cannot observe whether an agent has acquired a signal (and, hence, the realization of such a signal) nor the agent's payoff. The planner's objective is to maximize the discounted present value of all agents' payoffs, taking into account also the costs of acquiring a signal. We let $\delta \in (0, 1)$ denote the planner's discount rate. The planner can commit to a mechanism that specifies the messages to be conveyed to each agent as a function of the history (to be defined formally shortly). The mechanism chosen by the planner is known to all agents.

Let p_t denote the planner's belief, at the beginning of period t, that the state is A. Before the arrival of agent 1, the planner is as informed as the agents about the state of the world and, hence, his prior belief that the state is A is p_1. At later stages, p_t will depend on the information the planner possesses at the beginning of the period. If, at the beginning of period t, on the basis of the information he possesses, the planner believes that l_a agents observe the signal s^a and l_b observe the signal s^b, then his belief will be

$$p_t = \left\{ \begin{array}{ccc} 0 & \text{if} & l_b \geq 1 \\ \frac{p_1}{p_1+(1-p_1)(1-q)^{l_a}} & \text{otherwise} & \end{array} \right\}. \qquad (1)$$

Note that for all t, either $p_t = 0$ or $p_t \geq p_1$.

4.3 The Agents' Strategy

Recall that agent t does not observe the actions taken by the agents who moved before him and the only information available to him, before taking any action, is (i) his place in line, (ii) his prior belief, (iii) the mechanism in place, and (iv) the information transmitted to him by the planner. Let μ_t denote agent t's belief that the state is A, on the basis of the information he has received from the planner. The first agent and the planner share the same belief $\mu_1 = p_1$; however, this is not necessarily true for the other agents who know less than the planner (unless the planner has revealed to them all the information he possesses). Note, however, that for all t, either $\mu_t = 0$ or $\mu_t \geq p_1$.

Consider some agent t whose belief is μ_t. Agent t's expected payoff from taking action a without acquiring a signal is μ_t and his expected payoff from the action b is $1-\mu_t$. Assume that no further communication between the agent and the planner takes place after an agent has observed a signal. Then the following claim (the proof of which is omitted) holds:

Claim 1 *If no further information is revealed to the agent after acquiring a signal, the agent will find it optimal to incur the cost of acquiring a signal only if he plans to follow it.*

Assume hereafter that the premises of the above claim hold. (This is formally proved in Section 10.) Thus, if an agent acquires a signal (i.e., when $\mu_t \geq p_1$) he will surely take the optimal action if the state is A (the probability of observing the signal s^a when the state is A is 1) and the probability that he will take the optimal action is q if the state is B (the probability of observing the signal s^b). Hence, if agent t's belief is μ_t, his expected payoff from acquiring a signal is:

$$\mu_t + (1-\mu_t)q - c = \mu_t + q - \mu_t q - c.$$

Let $u_t(\mu_t, d)$ denote the expected utility of agent t from taking action $d \in \{a, b, e\}$. Then,

$$u_t(\mu_t, d) = \left\{ \begin{array}{c} \mu_t \text{ if } d = a \\ 1 - \mu_t \text{ if } d = b \\ \mu_t + q - \mu_t q - c \text{ if } d = e \end{array} \right\}$$

where $u_t(\mu_t, e)$ follows from Claim 1. Based on this we can conclude that

Claim 2 *Agent t's expected utility-maximizing decision is given by:*

$$d_t(\mu_t) = \left\{ \begin{array}{c} a \textit{ if } \mu_t > \mu_a \\ b \textit{ if } \mu_t < \mu_b \\ e \textit{ if } \mu_b \leq \mu_t \leq \mu_a \end{array} \right\} \tag{2}$$

where

$$\mu_a = \frac{q - c}{q} \quad \textit{and} \quad \mu_b = \frac{1 + c - q}{2 - q}.$$

To make the model non-trivial we will make the following assumption.

Assumption 1: $c < q/2$ and $p_1 \in [\mu_b, \mu_a]$.

Remark 3 *(i) If the cost of acquiring a signal is above $q/2$, then there are no beliefs for which the agent is willing to invest in obtaining a signal (i.e., $\mu_a < \mu_b$) and all agents will choose the action a if $p_1 > 1/2$ and the action b otherwise. (ii) Our assumption that $p_1 \in [\mu_b, \mu_a]$ guarantees that an agent with no information other than the prior belief p_1 (e.g., agent 1) will find it optimal to explore.*

5 First-Best Solution

The first-best mechanism refers, in our model, to the planner's optimal policy in the (hypothetical) case in which the planner can decide, at the beginning of every period, whether the agent will take the action a, b, or e. If the planner decides that the agent will acquire a signal, he (the planner) will not observe the realization of the signal but he will observe the final action a or b that the agent has chosen. Recall from (1) that p_t is either 0 or above $p_1 \in [\mu_b, \mu_a]$. Thus, in the first-best mechanism, e is chosen only when $p_t \geq p_1$ and the agent will always follow the signal. Hence, under the first-best scheme, the planner can infer the realization of the signal from the agent's choice. It also follows that a solution to the first-best mechanism need only specify the decision to be taken for $p_t \geq p_1$. As will be shown later, analyzing the first-best mechanism is interesting not only because it enables us to better understand the social costs created by the fact that the agents are selfish, but also because it will turn out to be very useful in characterizing the planner's optimal (second-best) mechanism in the case in which it is the agent who decides whether or not to acquire a signal.

The first-best solution is based on a stopping rule. The planner dictates that the agents to acquire a signal until one of the following two events takes place: (i) one agent takes the action b, in which case the planners tells to all agents thereafter to take the action b without acquiring a signal, or (ii) all agents have taken the action a and the planner's posterior, p_t, reaches some prespecified level p_a (to be defined formally below), at which point the planner tells all agents thereafter to take the action a without acquiring a signal. The probability p_a, referred to as the first-best cutoff, has the property of being the lowest posterior at which the likelihood that the true state is A is high enough that the planner does not find acquiring a signal worthwhile. The proposition below (the proof of which is provided in the Appendix) is based on the observation that if, at some prior, it is not worthwhile for the planner to acquire a signal, then it is not worthwhile for him to acquire a signal at higher priors.

Proposition 4 *The first-best solution is given by a cutoff $p_a = \mu_a + c\delta/q$ such that:*

(i) if $p_t \in [p_1, p_a]$, agent t acquires a signal and follows it;

(ii) if $p_t = 0$, agent t takes the action b (without acquiring a signal); and

(iii) if $p_t > p_a$, agent t takes the action a (without acquiring a signal).

Based on Propositions 4 and 2 we can characterize the potential conflict of interests between the planner and the agents. Assume that the planner shares all his information with the agents, i.e., $p_t = \mu_t$ for all t. In such a case, when $\mu_a < p_t < p_a$, there is a conflict of interests between the agents, who would like to choose the action a without acquiring a signal, and the planner, who would like the agents to explore. It should be stressed, however, that this conflict of interests between the agents and the planner does not by itself explain why the first-best mechanism cannot be implemented when the planner cannot dictate to the agents which action to take. As will be illustrated in the next section, in the case of symmetry it is easy to overcome this conflict of interest and achieve the first-best solution.

5.1 Benchmarks: Symmetric Setup and Free Signals

Before solving the model we discuss two benchmarks that provide useful background. The first one is where the setup is fully symmetric (vis-a-vis the two states) and the second benchmark occurs when signals are free. In both of these cases the first-best solution is implementable, and, it is therefore the combination of costly signals and an asymmetric information structure that makes the task of the planner challenging.

Consider first the case in which the information setup is fully symmetric. That is, the prior is that both states are equally likely, i.e., $p_1 = 0.5$, and conditional on the state being $A(B)$, the signal gets the value s^a $\left(s^b\right)$ with probability

$q > 0.5$. The first-best strategy is based on a threshold $\bar{p} > 0.5$. That is, the planner instructs agents to acquire signal as long as he is uncertain about the true state and his belief $p_t \in [1-\bar{p},\ \bar{p}]$. Once he is sufficiently certain about the true state ($p_t > \bar{p}$ or $p_t < 1-\bar{p}$) he instructs the agents to take an action based on the more likely state. Note, however, that this strategy will also be incentive compatible when the planner cannot instruct the agents to acquire a signal. The reason for this result is that, since the planner's policy is symmetric, the agents do not update their beliefs when they receive a recommendation to acquire a signal but rather simply follow the planner's recommendation. Similarly, since the agent benefits less than the planner from acquiring a signal, when the planner recommends taking an action without acquiring a signal, the agent will be happy to do so.

Consider now the other benchmark that occurs when the signal is costless $(c = 0)$. In this case, because the signal is free, each agent will obtain it. The planner can simply ask the agent for the realization of the signal, and then reveal to the agent the information gathered so far (or alternatively recommend to him the optimal action based on this information.) This mechanism is incentive compatible and implements the first-best solution.

6 The Recommendation Mechanism

Assume now that it is the agent's decision, in every period, whether to take the action a or b or to acquire a signal (the realization of which he will then follow). In this section we focus on a particular class of mechanisms, referred to as *recommendation mechanisms.* Section 10 shows that restricting attention to this class of mechanism is without loss of generality. In a recommendation mechanism, at the beginning of every period the planner mixes between three possible recommendations: choose a, choose b, and explore (i.e., choose e). Upon receiving the planner's recommendation, the agent decides whether to follow it; if he chooses to explore he must pay the cost c, after which he observes the realization of the signal and finally chooses a or b according to the signal. After the agent has made his final choice a or b, the planner observes this choice and the game proceeds to the next period.

Let $m_t \in \{a, b, e\}$ denote the planner's recommendation at the beginning of period t and denote by h_t an history, of length t, observed by the planner. The history h_t consists of the planner's recommendations $m_{t'}$ in all periods $t' \leq t$ and the actions a or b made by the t agents. Thus, h_t is the relevant history before the planner recommends an action to agent $t+1$. Given a recommendation mechanism, let H_t denote the set of all possible histories of length t induced by the mechanism. The planner's recommendation policy in period $t+1$ is history dependent, and we write $\alpha_{t+1} : H_t \to [0, 1]$ to denote a function mapping the possible histories into a probability of recommending the action a. Similarly $\beta_{t+1} : H_t \to [0, 1]$ and $\varepsilon_{t+1} : H_t \to [0, 1]$ are defined for the actions b and e, respectively. We are now ready to provide the formal definition of a recommendation mechanism.

Definition 5 *A recommendation mechanism M is a sequence of mappings $\{M_t\}_{t=1}^{\infty}$ that specifies, for every t, the probability with which action $i \in \{a, b, e\}$ is recommended in period t, as a function of the history $h_{t-1} \in H_{t-1}$. Specifically, we have*

$$M_t = (\alpha_t, \beta_t, \varepsilon_t) \mid \alpha_t : H_{t-1} \to [0,1], \quad \beta_t : H_{t-1} \to [0,1], \quad \varepsilon_t : H_{t-1} \to [0,1] \tag{3}$$

such that for all $h_{t-1} \in H_{t-1}$

$$\alpha_t(h_{t-1}) + \beta_t(h_{t-1}) + \varepsilon_t(h_{t-1}) = 1,$$

and where H_{t-1} is the set of all possible histories induced by $M_\tau,\ \tau < t$.

Every recommendation mechanism determines a set of possible paths along which the planner's beliefs, p_t, can evolve over time. For $i \in \{a, b\}$ let (h_{t-1}, m_t, i) denote the history h_t that is generated from h_{t-1} and followed, at time t, by the recommendation m_t and the (observable) choice $i \in \{a, b\}$ made by the agent. Fix an incentive compatible recommendation mechanism M and some history h_{t-1}, induced by this mechanism. There are only four possible continuing histories of length t along this mechanism: (h_{t-1}, a, a), (h_{t-1}, b, b), (h_{t-1}, e, a), and (h_{t-1}, e, b). The history h_{t-1}, observed by the planner, determines a posterior probability for the planner, $p_t(h_{t-1})$, that the state is A. The law of motion of the planner's posterior beliefs can now be fully described:

$$\begin{array}{clcl} (i) & p_t(h_{t-1}, a, a) & = & p_{t-1}(h_{t-1}), \\ (ii) & p_t(h_{t-1}, b, b) & = & p_{t-1}(h_{t-1}), \\ (iii) & p_t(h_{t-1}, e, b) & = & 0, \\ (iv) & p_t(h_{t-1}, e, a) & = & \dfrac{p_{t-1}(h_{t-1})}{p_{t-1}(h_{t-1}) + (1 - p_{t-1}(h_{t-1}))(1-q)}. \end{array}$$

From agent t's perspective, the planner's beliefs at the beginning of period t, are random variables, which we denote by $\tilde{p}_t$. Given a mechanism, we let $\pi_t(p_t)$ denote the probability that $\tilde{p}_t = p_t$. In what follows we use h_{t-1} and p_t interchangeably whenever there is no risk of confusion. In particular, we sometimes use the notation $\alpha_t(p_t)$ to mean the probability that the planner recommends action a conditional on $\tilde{p}_t = p_t$. Similarly for $\beta_t(p_t)$ and $\varepsilon_t(p_t)$.

Knowing the mechanism and understanding the law of motion of the planner's beliefs, the agent, after hearing the planner's recommendation, can form beliefs about the distribution of p_t. Let $\mu_t(a)$ denote the belief of agent t that the state is A, after hearing the recommendation a and, similarly, let $\mu_t(b)$ and $\mu_t(e)$ denote the belief of agent t that the state is A,after hearing the recommendation b and e, respectively. Then,

$$\mu_t(a) = E(\tilde{p}_t | m_t = a), \quad \mu_t(b) = E(\tilde{p}_t | m_t = b) \text{ and } \mu_t(e) = E(\tilde{p}_t | m_t = e).$$

Recall that $\pi_t(p_t)$ denotes the probability that agent t assigns to p_t being the planner's belief. It follows that $\sum_{p_t} \alpha_t(p_t)\pi_t(p_t)$ is the probability of recommending α in period t.

Using Bayes' rule we can write

$$\mu_t(a) = \frac{\sum_{p_t} \alpha_t(p_t)\pi_t(p_t)p_t}{\sum_{p_t} \alpha_t(p_t)\pi_t(p_t)};\ \mu_t(b) = \frac{\sum_{p_t} \beta_t(p_t)\pi_t(p_t)p_t}{\sum_{p_t} \beta_t(p_t)\pi_t(p_t)};\ \mu_t(e) = \frac{\sum_{p_t} \varepsilon_t(p_t)\pi_t(p_t)p_t}{\sum_{p_t} \varepsilon_t(p_t)\pi_t(p_t)}.$$

A recommendation mechanism is incentive compatible if:

$$\mu_t(a) \geq \mu_a, \quad \mu_t(b) \leq \mu_b, \quad \text{and} \quad \mu_t(e) \in [\mu_a, \mu_b].$$

7 The Optimal Recommendation Mechanism

Let us start with a few definitions. Consider some history of length t, along which all agents explore, obtain the signal s^a, and choose the action a. Let $\hat{t}$ denote the first period along such a history for which the planner's posterior is strictly above μ_a. That is,

$$\hat{t} = \max\left[t \mid \frac{p_1}{p_1 + (1-p_1)(1-q)^{t-1}} \leq \mu_a\right]. \tag{4}$$

Thus, $\hat{t}-1$ is the maximal number of (consecutive) s^a signals after which the planner's posterior is weakly below μ_a. That is, if agent $\hat{t}$ knows that all the agents before him received the signal s^a he will still want to explore, but if he also receives the signal s^a and the history is known to agent $\hat{t}+1$, then agent $\hat{t}+1$ will want to take the action a without any exploration.

Similarly, let

$$\bar{t} = \max\left[t \mid \frac{p_1}{p_1 + (1-p_1)(1-q)^{t-1}} \leq p_a\right] \tag{5}$$

As above, $\bar{t}-1$ is the maximal number of (consecutive) s^a signals after which the planner's posterior is weakly below p_a. If all the agents who moved before agent $\bar{t}$ received the signal s^a, the planner will still want agent $\bar{t}$ to explore; but if agent $\bar{t}$ also receives the signal s^a, then the planner will want agent $\bar{t}+1$ and all the agents who move after him to take the action a without any exploration. Since $\mu_a \leq p_a$, we know that $\hat{t} \leq \bar{t}$. Hereafter we will assume that $\hat{t} < \bar{t}$[2].

Before we characterize the optimal recommendation mechanism, we find it useful to discuss in further details exactly why the first-best mechanism fails as a recommendation mechanism. This discussion illustrates the important role of the asymmetry between the two actions, a and b, in the failure of the first-best mechanism as a recommendation mechanism.

[2] Note that if $\hat{t} = \bar{t}$ then a policy of full revelation is optimal and the first-best is trivially achieved. This might be the case if the signals are very precise.

8 The Infeasibility of the First-Best Mechanism

Suppose that the planner employs the first-best mechanism as a recommendation mechanism. That is, the mechanism recommends that the first agent and all the agents who move after him acquire a signal, as long as one of the following events does not take place: (i) one of the agents chooses b, in which case the mechanism recommends that all agents thereafter take the action b, or (ii) $\bar{t}$ periods have elapsed without any of the agents taking the action b (i.e., the planner's posterior has essentially reached p_a), in which case the planner recommends that all agents thereafter take the action a.

It is easy to see that in the early stages of this mechanism and, in particular, as long as $t \leq \hat{t}$, there is no conflict of interest between the planner and the agents, and the first-best mechanism is incentive compatible. At these early periods, if the mechanism's recommendation that an agent is to take the action b, the agent will follow the planner's recommendation, and if the mechanism's recommendation to an agent is to explore, the agent will find it optimal to explore since his prior belief will still be below μ_a (recall that $t \leq \hat{t}$).

However, as soon as more than $\hat{t}$ periods have elapsed, the first-best mechanism is no longer incentive compatible. If, at some period $t > \hat{t}$, the mechanism's recommendation to the agent is to acquire a signal (as the first-best mechanism suggests), the agent will conclude that all proceeding agents observed the signal s^a. (Otherwise, the mechanism's recommendation would have been to take the action b). Hence, the agent's posterior will be higher than μ_a, in which case, instead of following the planner's recommendation, the agent will take the action a. Thus, the main reason why the first-best mechanism cannot be implemented, as a recommendation mechanism, is that, under the first-best mechanism, as soon as the planner learns that the true state is B, he will want to reveal this information to all agents thereafter. However, such a policy implies that if the planner does not announce that the true state is B, the agents must infer that all the $t \geq \hat{t}$ agents who moved before them got the signal s^a and they will update their prior beliefs accordingly, namely, to be above μ_a.

Since the first-best outcome cannot be achieved, the planner uses a different (second-best) mechanism to maximize social surplus. In what follows, and mainly in order to simplify the exposition, we first design the optimal mechanism as a fully private mechanism in which the planner's recommendation to each agent t are observed only by this agent and are not revealed to any other agent. We then show in Section 9 how the optimal mechanism can be cast as a fully public mechanism in which all messages are public.

8.1 The Main Theorem

Additional notations and discussion are warranted prior to the presentation of our main theorem. Let us define a "modified" cost of exploration function $c^*(p)$ as follows:

$$c^*(p) = c + \frac{\max\{0, (p - \mu_a)\}}{\mu_a}(1 + c - q) \,. \tag{6}$$

Observe that for $p \leq \mu_a$, where, as discussed above, there is no conflict of interest between the planner and the agents, $c^*(p) = c$ and, otherwise, $c^*(p)$ is monotone in p. As will be discussed below and will be shown formally in the proof of our main theorem, the solution to the second-best problem is, in fact, the solution to a modified first-best problem in so far as the cost of exploration is $c^*(p)$ rather than c. Indeed, as we will show shortly, the modified cost of exploration function, $c^*(p)$, captures not only the real cost of exploration but also the "implied" cost resulting from the need to make the agent's exploration incentive compatible when $\mu_a < p_t$

Using this modified cost function we can define a modified cutoff, p_a^*, much as we defined p_a for the original first-best problem. That is, p_a^* is the point at which the planner's posterior probability is high enough so that exploration becomes too costly under the modified first-best mechanism when the cost is $c^*(p_a^*)$. Specifically,

$$p_a^* = \mu_a + c^*(p_a)\delta/q$$

Similarly we define t^* by substituting p_a with p_a^* in the definition of $\bar{t}$, namely,

$$t^* = \max\left[t \mid \frac{p_1}{p_1 + (1 - p_1)(1 - q)^{t-1}} \leq p_a^*\right] . \tag{7}$$

That is, $t^* - 1$ is the maximal number of (consecutive) s^a signals after which the posterior is weakly below p_a^*.

We next provide an informal description of the optimal mechanism, which set the stage for the proof of the formal theorem.

8.2 The Optimal Mechanism: An Informal Description

The optimal private mechanism consists of the following three phases.

Phase One: This phase starts in period 1 and lasts for $\hat{t}$ periods. During this phase, all agents receive recommendation to explore unless one of them takes the action b. If, during this phase, an agent t' takes the action b, the mechanism recommends that all agents t, $t' < t \leq \hat{t}$, take the action b without acquiring a signal. Thus, at this phase, the recommendation policy coincides with that of the first-best mechanism. (Notice, however, that now the recommendations are private). During this phase, if an agent receives a recommendation to explore, he must conclude that all the agents who moved before him acquired a signal and received the signal s^a; and if he receives a recommendation to take the action b, he must conclude that one of the agents who moved before him observed the signal s^b.

It is important to note that for all t in this phase (and, as will be shown later, in all other phases as well), the planner's posterior, at the end of period

t, is either 0 or $p_1/(p_1 + (1-p_1)(1-q)^t) > p_1$. Thus, the distribution of $\tilde{p}_t$ consists of two points only.

Phase Two: This phase starts at the beginning of period $\hat{t}+1$ and ends at the end of period t^*, where, generically, $\hat{t} < t^* < \bar{t}$. (See the definitions in 4, 5 and 7). Since $t > \hat{t}$, the planner, in order to induce the agents to explore, must hide some information from the agents, for example, by committing to the following (randomizing) policy: if $p_t > 0$ (i.e., if none of the agents have yet observed the signal s^b), then, with probability 1, the planner recommends that agent t explore, namely, $\varepsilon_t^*(p_t) = 1$. If $p_t = 0$, the planner randomizes between recommending that agent t take the action b and recommending that he explore, namely, $\varepsilon_t^*(0) = 1 - \beta_t^*(0) > 0$. The probability with which action e is recommended when $p_t = 0$ is chosen in such a way that agent t's posterior after receiving the recommendation to explore (i.e., $\mu_t(e)$) will be exactly μ_a, and he will follow the recommendation.

Phase Three: This phase starts in period t^*+1 and goes on forever. In this phase none of the agents acquires a signal. The planner recommends that all agents take the action a, if none of the agents who moved in Phase One or Phase Two took the action b; otherwise, he recommends that all agents take the action b.

In sum, the second-best mechanism is different from the first-best mechanism in two respects: (i) exploration is conducted in the second phase and the action a may be taken by some of the agents, even after the planner has already learned that the true state is B, and (ii) exploration may be terminated because of its ex ante (wasteful) costs in the second phase and the action a is adopted earlier than in the first-best case, namely, when $p_t = p_a^* < p_a$.

8.3 The Optimal Mechanism: A Formal Statement and Proof

Recall that we write $\alpha_t(p_t)$ to denote the probability that the planner recommends that agent t take the action a conditional on $\tilde{p}_t = p_t$ and similarly for $\beta_t(p_t)$ and $\varepsilon_t(p_t)$. Let $p_t^* = p_1/[p_1 + (1-p_1)(1-q)^{t-1}]$.

Theorem 6 *The optimal mechanism M^* consists of three phases:*

Phase 1: *For all $t \leq \hat{t}$:*

$$\beta_t^*(0) = 1, \quad \textit{otherwise} \quad \varepsilon_t^*(p_t) = 1.$$

Phase 2: *For all t such that $\hat{t} < t \leq t^*$,*

$$\beta_t^*(0) = 1 - \varepsilon_t^*(0) = \frac{p_t^*}{p_t^* - p_1}\frac{(\mu_a - p_1)}{\mu_a}; \ \textit{otherwise} \ \varepsilon_t^*(p_t) = 1.$$

Phase 3: *For all t such that $t > t^*$,*

$$\beta_t^*(0) = 1; \ \textit{otherwise} \ \alpha^*(p_t) = 1.$$

While the formal proof is relegated to the Appendix, we now provide the main idea and intuition behind the proof.

8.4 Informal Proof: Main Idea and Intuition

The logic behind the first phase is based on the fact that in the first few periods of the mechanism no conflict of interests exists between the planner and the agents. They all prefer action b if they know that one of the agents has received the signal s^b. They also find it worthwhile to acquire a signal if all agents have realized a signal s^a because their belief is in $[\mu_b, \mu_a]$. Thus, the optimal policy is based on full transparency as long as $t \leq \hat{t}$. It is important to note that because messages are private there is no need to worry about the effect of this transparency on future agents. This is formally established in a series of simple claims in the Appendix.

The second phase is more interesting because it raises a potential conflict of interest between the agents and the planner. Following the planner's policy in the first phase, the agents know that the belief of the planner, during the second phase, is either $p_t = 0$ or $p_t > \mu_a$. When $p_t > \mu_a$ but small enough, the planner would like the agents to acquire a signal whereas they would like to take action a. To overcome this obstacle and make exploration incentive compatible the planner must, in some histories, recommend exploration even when he knows that the state is B. This leaves agents with sufficient uncertainty about the exact history so that their posterior is back in $[\mu_b, \mu_a]$. This strategy of the planner is however costly; as it means that agents acquire a costly signal when the state is already known and, even worse, they may take the (wrong) action a, when the planner already knows that the true state is B. To minimize these costs, the mechanism randomizes in such a way that when the agent receives a recommendation to acquire a signal, his posterior is exactly μ_a. To be precise, because agent t knows that the planner's prior, p_t, is either 0 or $p_1/(p_1 + (1-p_1)(1-q)^{t-1} \equiv p_t^*$, the planner must assign enough weight to exploration when $p_t = 0$ to make the posterior of the agent, conditional on hearing the recommendation to explore, equal μ_a. This weight is exactly $\frac{p_1}{p_t^* - p_1} \frac{(p_t^* - \mu_a,)}{\mu_a}$ (see the appendix for details). Once we establish that exploration beyond period $\hat{t}$ implies setting the agent's posterior $\mu_t(e)$ exactly at μ_a, it is left to determine which agents, beyond agent $\hat{t}$, will receive recommendation to explore and with what probability.

A key step in the proof is the relationship between the optimal mechanism and the first-best mechanism in a hypothetical environment in which the cost of exploration is a function of the planner's belief, given by $c^*(p)$, as defined in 6. The cost $c^*(p)$ internalizes the extra cost involved in equating the agent's posterior, conditional on hearing the recommendation to explore, to μ_a. In this hypothetical environment, the more likely the planner is to believe that the state is A, the more costly it is for him to acquire the signal. In the formal proof of the theorem above, it is shown that the optimal incentive-compatible mechanism of our problem can be obtained by solving the first-best problem in the hypothetical environment described above. The similarity between the two problems comes from the fact that the "indirect" cost that the planner has to incur, in our problem, if he wishes agents to explore when $p_t > \mu_a$, appears as a direct cost in the hypothetical first-best problem. For every $p_t > \mu_a$, the cost of exploration in the hypothetical first-best environment, $c^*(p)$, equals exactly the direct cost of exploration plus the indirect cost of inducing agents to explore in our original problem.

The solution to the modified first-best problem determines the threshold p_a^* and the last agent to explore t^* in our original problem. Furthermore, this solution also shows that for $p_t^* > 0$ the optimal policy does not involve any randomization. The mechanism recommends "explore," for $t \leq t^*$, and the action a, for $t > t^*$.

9 A Public Mechanism Is Optimal

The optimal mechanism presented above is fully private. Such a mechanism makes the planner's task a lot easier because he does not need to worry about the effect of the information revealed to one agent on the behavior of future agents. In reality, however, making the recommendations fully private is often not feasible. Privacy requires, among other things, that agents not reveal to other agents either the recommendation they received from the planner or the signal they observed. Nevertheless, the optimal private mechanism presented here features a useful property: it can easily be transformed into a fully public mechanism that will also be optimal.

To see how the optimal mechanism can be modified into a public one, assume first that agent $\hat{t}$ is just indifferent between acquiring a signal and taking the action a without any exploration, namely,

$$\frac{p_1}{p_1 + (1-p_1)(1-q)^{\hat{t}-1}} = \mu_a.$$

Consider now the following slightly modified optimal mechanism in which, besides the fact that messages are public, the only change is that in Phase Two, with probability 1, all agents receive recommendation to explore, regardless of the information that the planner accumulates during this phase. In particular, even if the planner learns that the state is B, he will recommend in Phase Two that the agents explore. Clearly, all the agents in Phase One (i.e., $t \leq \hat{t}$) are unaffected by the public nature of the messages because the private mechanism is essentially fully revealing for them. Consider now an agent in Phase Two ($\hat{t} < t \leq t^*$) under the public mechanism. This agent is fully informed about the history up to (but not including) the action of agent $\hat{t}$ but completely in the dark thereafter. Hence, if some agents receive a recommendation of b at some agent $t' \leq \hat{t}$, then agent t is informed that the state is B and he will choose b. Otherwise, all he knows is that the first $\hat{t}-1$ agents obtained a signal s^a and, hence, his posterior is exactly μ_a and his optimal choice is to explore. Thus, all the agents in this phase behave exactly as they would have behaved under a private mechanism. As for Phase Three, the recommendation to all the agents in this phase is exactly the same as under the private mechanism.

Consider now the case where

$$\frac{p_1}{p_1 + (1-p_1)(1-q)^{\hat{t}-1}} < \mu_a,$$

The first and third phases of the public mechanism are similar to those of the private mechanism above (except that messages are now public) and consequently all agents $t' \leq \hat{t}$ take the first-best action. To obtain the same outcome as in the private mechanism, it must be the case that every agent, in the second phase, is either informed that the state is B or possesses a posterior equal to μ_a. This outcome can be achieved by partially revealing the choice of agent $\hat{t}$. (Recall that in the public mechanism above only the choices of the first $\hat{t}-1$ agents were made public.) In particular, conditional on agent $\hat{t}$ choosing the action b, after receiving recommendation to acquire a signal (that is, agent $\hat{t}$ is the first one to observe the signal s^b), all agents starting from agents $\hat{t}+1$ onwards are (publicly) recommended, with some positive probability, to choose the action b and, with the remaining probability, all the agents in Phase Two are recommended to acquire a signal. If, however, agent $\hat{t}$ chooses the action a, all the agents who move in Phase Two receive (public) recommendation to acquire a signal. The probability with which exploration is recommended to all agents in Phase Two, after agent $\hat{t}$ was the first one to choose b, is chosen so that conditional on receiving the recommendation to explore, the posterior beliefs of all agents in Phase Two (starting from agent $\hat{t}+1$) are exactly μ_a. As in the previous case, all other agents in the second phase get the same recommendation as agent $\hat{t}+1$.

Proposition 7 *There exists an optimal mechanism that is public.*

10 Proof That Recommendation Is without Loss

So far we have restricted our attention to the class of recommendation mechanisms, and characterization of the optimal mechanism within that class. A recommendation mechanism is based on each agent receiving a single incentive-compatible recommendation. In general, of course, the set of messages could be larger (indeed, infinitely large). Furthermore, the planner need not restrict himself to move only once at the beginning of each period; he could also send another message after the agent has had a chance to explore and report his signal. The theorem below establishes that the restriction to the recommendation mechanism is indeed without loss of generality.

We preface the theorem with the following important facts: First observe that after an agent has taken an action a or b, the planner has nothing to gain from further communication with the agent. The agent's choice of a or b reveals all the private information he might have obtained during the period. Hence, we can restrict communication to three rounds of messages at most: 1) a message from the planner to the agent that might convey some information about the past; 2) a message from the agent to the planner that might be sent to convey the realization of the signal; and 3) a message from the planner to the agent sent prior to the agent's final choice of action.

Next we recall some consequences of Myerson (1986) that are relevant to our model. We can restrict our attention to incentive compatible mechanisms in which in the first round the planner randomizes between the following four recommendations: $m_t^1 \in \{a, b, e, n\}$, in which the (new) action n means that the agent receives a recommendation to avoid taking an action (for now). The agent then follows the recommendation and if the agent's action is a or b, his choice is observed by the planner and the mechanism moves to the next period. If, however, the recommendation is to acquire a signal (i.e., e), then the agent reports the realization of the signal truthfully. Finally, the mechanism randomizes between the recommendations a and b and the agent follows the recommendation. In the case where the planner recommends to the agent to do nothing (n) in the first round, he will recommend that the agent take an action, a or b, in the third round. The theorem below establishes that the second and third rounds are redundant. It is enough to consider a recommendation mechanism of one round, in which $m_t \in \{a, b, e\}$ and the recommendation e is followed by the action that agrees with the signal.

Proposition 8 *The optimal recommendation mechanism is optimal within the class of all mechanisms.*

Proof. Suppose, by way of contradiction, that there exists a history $\tilde{h}_{t-1}$ after which the mechanism recommends (with some positive probability) that the agent acquire a signal; after the agent reports the observed signal, the mechanism randomizes between recommending a or b. Clearly, if the agent observes the signal s^b he will take the action b regardless of the planner's recommendation. Thus, the planner can only send a useful message to the agent after the agent observes and reports that he has observed the signal s^a.

If after reporting s^a the planner recommends the action a with probability 1, then his message is of no value to, either the agent or the planner, because this action will be taken anyway. Thus, for a second message from the planner to have an impact, some history $\hat{h}_{t-1}$ must exist after which the mechanism recommends acquiring a signal. Then, after the agent acquires the signal and reports s^a, the planner recommends the action b. For this strategy to be optimal the planner must have observes that at least one agent $t' < t$ took the action b.

Consider a modified mechanism in which after $\hat{h}_{t-1}$ the mechanism recommends b instead of e. The modified mechanism will still be incentive compatible. (In particular, the agent will still prefer to acquire a signal in those cases in which he is recommended to do so.) The modified mechanism will also generate a higher surplus, both to the agent and to the planner. This modified mechanism eliminates the redundant cost of exploration when the state is known to be B. Hence, there is no need for a second message by the planner and, thus, there is no need for a message from the agent to the planner after the agent has acquired a signal. We therefore conclude that it is without loss of generality to allow for only one message from the planner to the agent, at the beginning of the period, from the set of messages $\{a, b, e\}$. ■

References

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

[2] Bikhchandani, S., D. Hirshleifer, and I. Welch (1992): "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy*, 100, 992–1026.

[3] Che, Y.K. and J. Horner (2014): "Optimal Design for Social Learning," working paper.

[4] Ely, J., A. Frankel, and E. Kamenica (2013): "Suspense and Surprise," working paper.

[5] Ely, J.C. (2015): "Beeps," working paper.

[6] Gershkov, A. and B. Szentes (2009): "Optimal Voting Schemes with Costly Information Acquisition," *Journal of Economic Theory*, 144, 36–68.

[7] Kamenica, E. and M. Gentzkow (2011): "Bayesian Persuasion," *American Economic Review,* 101, 2590–2615.

[8] Kremer, I., Y. Mansour, and M. Perry (2014): "Implementing the Wisdom of the Crowd," *Journal of Political Economy*, Vol. 122, No. 5, 988-1012.

[9] Martimort, D. and S. Aggey (2006): "Continuity in Mechanism Design without Transfers," *Economics Letters*, 93, 182–189.

[10] Myerson, R. (1986): "Multistage Games with Communication," *Econometrica* 54, 323–358.

[11] Rayo, L. and I. Segal (2010): "Optimal Information Disclosure," *Journal of Political Economy,* 118, 949–987.

[12] Sgroi, D. (2002): "Optimizing Information in the Herd: Guinea Pigs, Profits and Welfare," *Games and Economic Behavior,* 39, 137–166.

[13] Skrzypacz A. and J. Horner (2012): "Selling Information," working paper.

[14] Smith, L., N. Sorensen, and J. Tian (2014): "Informational Herding, Optimal Experimentation, and Contrarianism," working paper.

11 Appendix

11.1 Proof of the First-Best Policy (Proposition 4)

When the planner does not face incentive constraints his expected payoff, in every period, depends only on his prior belief at that period, p_t. Hence, the first-best solution can be viewed as a dynamic optimization where the state variable is p_t. In principle, the planner could randomize, at least in some periods, over the three alternatives, a, b, and e, but such a randomization will never be necessary. If the planner randomizes between, say, two actions, then his expected payoff is simply the average of the payoffs obtained from following each of these two actions. Thus, we can conclude that the first-best mechanism is deterministic.

By assumption, it is optimal for the first agent to explore and, hence, it is also optimal for the planner to do so in the first period. Also by assumption, since the signal s^b reveals the state to be B, we know that if at some period t the planner observes agent t taking the action b (after obtaining a signal), it is optimal for all agents, thereafter, to take the action b. The only remaining question is what should the planner do in a given period, if in all previous periods in which agents chose to explore, the realization of the signal was s^a (i.e., all agents took the action a)? Since $c > 0$, we know that there exists a posterior probability p_t large enough, so that it will not be optimal for the agent to acquire a signal, even if the signal was fully informative (i.e., $q = 1$). Thus (recall the updating rule 1), there exists a period $\bar{t}$ large enough so that if all agents observe the signal s^a in all periods $t \leq \bar{t}$, it will be optimal for the planner to choose the action a in all periods $t > \bar{t}$. It remains to be shown that if for some $t < \bar{t}$ the first-best mechanism chooses the action a without exploration, then it will choose the action a without exploration in all subsequent periods. This last claim follows immediately from the fact that if at period t the action a is taken, then $p_{t+1} = p_t$ and, therefore, the optimal action in $t + 1$ is also a.

Notice that p_a is the solution to the following equation:

$$\frac{p}{(1-\delta)} = \frac{p}{(1-\delta)} + (1-p)\frac{q}{(1-\delta)} - c.$$

The LHS of the equation above is the planner's payoff from taking the action a in the current period and in all periods thereafter, whereas the RHS is the planner's payoff from exploring one more time and following the signal in all periods thereafter, when the prior is p. If $p < p_a$ the RHS is greater than the LHS and, hence, the planner is better off exploring at least one more time. It is left to show that if $p > p_a$, the optimal action is a. Assume, by contradiction, that there exists some $p' > p_a$ at which the optimal action is to acquire a signal. Since we know that there exists a prior large enough so that at this prior it is optimal for the mechanism to choose a, there must be $p'' \geq p'$ at which it is optimal for the planner to acquire only one more signal and to follow it in all periods thereafter. This, however, leads to a contradiction since the LHS of the equation above is increasing with p at a higher rate than the RHS and hence it must be the case that it is better to terminate exploration a period earlier.

11.2 Proof of Main Theorem (Theorem 6)

The proof is done by proving several claims that, taken together, characterize the optimal recommendation mechanism M^*. Hereafter we refer only to p_t for which, given the mechanism in place, there is a positive probability that the planner has this belief, i.e., $\pi_t^*(p_t) > 0$. Note that given Assumption 1, every incentive-compatible mechanism must recommend that agent 1 explore, simply because this is what he is going to do anyway. The first claim says that for t small enough, when it is common knowledge that there is no conflict of interests between agent t and the principal, a policy of full revelation is adopted and the agent is recommended to take the best action for himself.

Claim 9 *For all* $t \leq \hat{t}$*, if the planner's prior at* t *is strictly positive (i.e.,* $p_t > 0$*), then* $\varepsilon_t^*(p_t) = 1$*; otherwise* $\beta_t^*(0) = 1$.

Proof. By the definition of $\hat{t}$, for all $t \leq \hat{t}$, it is common knowledge that $\Pr(\tilde{p}_t \in [\mu_b, \mu_a] \cup \{0\}) = 1$ and, hence, the agent's optimal choice is e if he knows that $\tilde{p}_t \neq 0$, and b otherwise. This is also the first-best choice. Consequently, if the agent is fully informed, he will follow the first-best choice strategy. Assume, by a way of contradiction, that the optimal IC mechanism is such that there exists some agent $t' \leq \hat{t}$ who is not fully informed about the moves of all preceding agents and as a result of this he does not take the action that is optimal for him. Consider a modified mechanism under which agent t' is informed about the moves of his predecessors and in all periods thereafter the planner (ignores the extra information obtained in period t' and) follows the original mechanism. Clearly, this modified policy yields a higher social welfare. A contradiction. ■

A consequence of Claim 9 is that in the first $\hat{t}$ periods, the optimal mechanism essentially reveals p_t to agent t and unless s^b is observed by one of the agents, all agents $t \leq \hat{t}$ acquire a signal. The corollary below follows directly from Lemma 9 together with the posterior's law of motion and the assumption that $p_1 \in [\mu_b, \mu_a]$.

Corollary 10 *For* $t \leq \hat{t}$, $\Pr(\tilde{p}_t^* \in [\mu_b, \mu_a] \cup \{0\}) = 1$ *and for* $t > \hat{t}$, $\Pr(\tilde{p}_t = 0) > 0$ *and* $\Pr(\tilde{p}_t \in (0, \mu_a)) = 0$.

Claim 9 and Corollary 10 above allow us to restrict our search for the optimal recommendation mechanism to histories at which $p_t = 0$ and $p_t > \mu_a$.

Claim 11 *For all* t *the planner never recommends the action* a *if he knows that the state is* B*, i.e.,* $\alpha_t^*(0) = 0$*, and, for* $p_t > \mu_a$*, the planner never recommends the action* b*, i.e.,* $\beta_t^*(p_t) = 0$.

Proof. Suppose that at some t we have $\alpha_t^*(0) > 0$. Consider the modified mechanism such that $\alpha_t'(0) = 0$ and $\beta'(0) = \beta^*(0) + \alpha_t^*(0)$. Such a modification will certainly increase agent t's utility and will not affect the other agent's payoff as the posterior beliefs in all periods thereafter will not change. A similar proof can be carried out to show that $\beta_t^*(p_t) = 0$ for all $p_t > \mu_a$. ■

An immediate consequence of Corollary 10 is that for $t > \hat{t}$, agent t knows that some agents (and, in particular, agent 1), who moved before him, chose e and with some positive probability received the signal s^b. Thus, agent t knows that either the planner has learned that $\tilde{p}_t = 0$, i.e., the state is B, or the planner's posterior is $\tilde{p}_t > \mu_a$. While we know from Claim 11 that $\alpha_t^*(0) = 0$, this does not imply that $\beta_t^*(0) = 1$.

Claim 12 *Consider period $t > \hat{t}$. If $\varepsilon_t^*(p_t) > 0$ for $p_t > \mu_a$, then $\varepsilon_t^*(0) > 0$ and $\mu_t^*(e) = \mu_a$.*

Proof. If $\pi_t^*(p_t) > 0$, for some $p_t > \mu_a$, then from Claim 9 we know that $\Pr(\tilde{p}_t \in (0, \mu_a)) = 0$. The proof of the first part of the claim follows immediately from the fact that in order for the mechanism to be IC it must be the case that the agent's posterior conditional on the recommendation to explore, i.e., $\mu_t^*(e)$, is such that $\mu_t^*(e) \in \left[\mu_b, \mu_a\right]$. Now assume by way of contradiction that $\mu_t^*(e) < \mu_a$. It is then possible to decrease $\varepsilon_t^*(0)$ by a small amount and increase $\beta_t^*(0)$ by the same amount. This increases the utility of agent t without affecting the distribution of $\tilde{p}_t^*$ for all $\tau > t$. ■

The above claims, taken together, summarize the IC constraints within which the optimization is carried out. Starting with the prior $p_1 \in [\mu_b, \mu_a]$, the random variable $\tilde{p}_t$ is either zero or above p_1. When $\tilde{p}_t = 0$, the planner recommends either b or e; similarly, when $\tilde{p}_t > \mu_a$, the planner recommends either a or e; and when $\tilde{p}_t \in [\mu_b, \mu_a]$, the planner recommends e. Moreover, whenever the recommendation is e, it must be that $\mu_t(e) \in [\mu_b, \mu_a]$ (where $\mu_t(e)$ is the agent's posterior following the recommendation e). When $t > \hat{t}$, the agent knows that $\Pr((\tilde{p}_t > \mu_a) \cup (\tilde{p}_t = 0)) = 1$. The expected value $\mu_t(\cdot)$ is over all possible p_ts given the mechanism in place, and so to provide agent $t > \hat{t}$ with incentives to explore when $\tilde{p}_t > 0$, the mechanism must also recommend exploration when $\tilde{p}_t = 0$ in order to bring down $\mu_t(e)$ to the region where the agent is willing to explore, i.e., $[\mu_b, \mu_a]$. Recommending exploration when $\tilde{p}_t = 0$ is costly, firstly because of the redundant cost of exploration when the state is already known (to the planner), and secondly because the agent may obtain a signal of s^a and choose the wrong action. To minimize these "costs," when $\tilde{p}_t = 0$ exploration is recommended with the minimal probability needed, which implies bringing the agent's posterior down to no lower than μ_a (i.e., $\mu_t(e) = \mu_a$).

With this in mind, we can now solve for the optimal mechanism by solving a modified first-best problem in which the cost of acquiring a signal is not only c but also the implied cost involved in keeping the agent's posterior $\mu_t(e)$ at μ_a. As we next show, this cost is monotone in p, and, hence, we can employ the same technique used in the solution to the original first-best problem to establish that for $t > \hat{t}$ and $p_t > \mu_a$, the optimal solution is deterministic, namely, either $\varepsilon_t^*(p_t) = 1$ or $\alpha^*(p_t) = 1$. Before proving this result formally, the following discussion will be helpful.

Let $\overline{\epsilon}_t(p) = \pi_t(p)\varepsilon_t(p)$ denote the "ex- ante" probability of exploration at p in period t. As discussed above, the efficient way to satisfy the IC constraints implies that

$$\mu_t^*(e) = \frac{\sum_{\tilde{p}_t} p\overline{\epsilon}_t(p)}{\sum_{\tilde{p}} \overline{\epsilon}_t(p)} = \frac{\sum_{\tilde{p}_t>0} p\overline{\epsilon}_t(p)}{\overline{\epsilon}_t(0) + \sum_{\tilde{p}_t>0} \overline{\epsilon}_t(p)} = \mu_a,$$

which can be written as

$$\overline{\epsilon}_t(0) = \frac{\sum_{\tilde{p}_t>0} p\overline{\epsilon}_t(p) - \mu_a \sum_{\tilde{p}_t>0} \overline{\epsilon}_t(p)}{\mu_a}.$$

Thus, if for some $p > 0$, exploration at p ($\overline{\epsilon}_t(p)$) is increased by one unit, this has a direct cost of c and an indirect cost of

$$H(p) = \frac{p - \mu_a}{\mu_a}[c + (1 - q)],$$

which is the cost involved in increasing exploration when the state is already known to be B.,Observe that $H(p)$ is increasing in p.

Based on this we will show that the optimal mechanism can be viewed as the solution to a modified first-best problem where the cost of exploration is

$$c^*(p) = c + H(p).$$

Using the same line of reasoning as we applied when solving for the (original) first-best mechanism, we denote by p_a^* the solution to the following equation:

$$p = \mu_a + c^*(p)\delta/q,$$

We note that:
(i) since $c^*(p)$ is non-decreasing, p_a^* is uniquely defined by the above equation, and (ii) $\mu_a < p_a^* < p_a$. We can, therefore, conclude that the first-best solution to the modified cost function is given by

$$\begin{aligned} \alpha_t(p) &= 1 \text{ for } p > p_a^*, \\ \varepsilon_t(p) &= 1 \text{ for } 0 < p < p_a^* \\ \beta_t(0) &= 1 \text{ for } t \leq \hat{t} \text{ and for } t > t'. \end{aligned} \tag{8}$$

We will now prove this formally.

Claim 13 *For every $\tau > \widehat{t}$ the optimal mechanism must satisfy the following maximization problem (hereafter referred to as the SB problem):*

$$\underset{\{\alpha_t(p)\}_{t=\tau}^{\infty}, \{\beta_t\}_{t=\tau}^{\infty}}{Max} V_\tau =$$

$$\sum_{p>\mu_a} \pi_\tau(p) \left(\alpha_\tau(p)p + (1 - \alpha_\tau(p)) \left(p + (1 - p) q - c\right)\right) + \pi_\tau(0)(\beta_\tau + (1-\beta_\tau)(q-c)) + \delta V_{\tau+1}$$

subject to:

$(I)\ V_t = \sum_{p>\mu_a} \pi_t(p)\left(\alpha_t(p)p + (1-\alpha_t(p))\left(p + (1-p)\,q - c\right)\right) + \pi_t(0)(\beta_t + (1-\beta_t)(q-c)) + \delta V_{t+1},$

for $t = \tau+1,\ \tau+2,\ \tau+3,...$

$(II)\ \pi_t(p) = \pi_{t-1}(p)\alpha_{t-1}(p) + \pi_{t-1}(p_{t-1}^{-1}(p))(1-\alpha_{t-1}(p_{t-1}^{-1}(p)))(p_{t-1}^{-1}(p) + (1-p_{t-1}^{-1}(p))(1-q)),$

for $p > \mu$ *and* $t = \tau+1,\ \tau+2,\ \tau+3,...$

$(III)\ \sum_{p>\mu_a}[(\mu_a - p)\pi_t(p)(1-\alpha_t(p)] + \pi_t(0)(1-\beta_t)\mu_a = 0,$

$for\ t = \tau,\ \tau+1,\ \tau+2,\ \tau+3,...$ *and*

$(IV)\ \pi_t(0)\ = 1 - \sum_{p>\mu_a} \pi_t(p),\ \alpha_t(p) \in [0,1]\,,\ \beta_t \in [0,1]\,,$

for $t = \tau,\ \tau+1,\ \tau+2,\ \tau+3,...$

where:

(i) $\pi_\tau(p)$ *is the distribution of possible beliefs at period* t*, given the mechanism.*

(ii) $p_t^{-1}(p)$ *is the inverse of the function*

$$p(p_t) = \frac{p_t}{p_t + (1-p_t)(1-q)},$$

which is the probability that the posterior is p *given that the prior was* p_t *and the agent explored and received the signal* s^a*.*

(iii) The function V_t*, specifies the expected present value to the planner, given a mechanism* M *for all* $t > \tau$*.*

(iv) The second constraint specifies the evolution of the distribution of the random variable $\tilde{p}_t$*, given the mechanism* M*.*

(v) The third constraint specifies the incentive-compatible constraint, guaranteeing that agent t' s posterior is exactly μ_a *when the mechanism recommends* e*.*

Proof. From the IC constraint (III) we obtain that, for every $t \geq \widehat{t}$,

$$\pi_t(0)(1-\beta_t) = \sum_{p>\mu_a} \frac{(p-\mu_a)}{\mu_a} \pi_t(p)(1-\alpha_t(p)). \tag{9}$$

Plugging (9) into V_t, for all $t \geq \widehat{t}$, we obtain that

$$V_t = \sum_{p>\mu_a} \pi_t(p) \left(\alpha_t(p)p + ((1-\alpha_t(p)) \left(p + (1-p) q - c^*(p)\right) + \pi_t(0) + \delta V_{t+1},\right.$$

where

$$c^*(p) = c + \frac{(p-\mu_a)}{\mu_a}(1 + c - q).$$

Thus, the SB problem can be simplified as follows (thereafter referred to as the $SB\prime$ problem):

$$\underset{\{\alpha_t(p)\}_{t=\tau}^{\infty}}{Max} V_\tau =$$

$$\sum_{p>\mu_a} \pi_\tau(p) \left(\alpha_\tau(p)p + ((1-\alpha_\tau(p)) \left(p + (1-p) q - c^*(p)\right) + \pi_\tau(0) + \delta V_{\tau+1},\right.$$

subject to:

$$(I)\ V_t = \sum_{p>\mu_a} \pi_t(p) \left(\alpha_t(p)p + ((1-\alpha_t(p)) \left(p + (1-p) q - c^*(p)\right) + \pi_1(0) + \delta V_{t+1},\right.$$

for $t = \tau + 1,\ \tau + 2,\ \tau + 3, \ldots,$

$$\begin{aligned}(II)\ \pi_t(p) \quad &= \quad \pi_{t-1}(p)\alpha_{t-1}(p) + \\ &\pi_{t-1}(p_{t-1}^{-1}(p))(1-\alpha_{t-1}(p_{t-1}^{-1}(p)))(p_{t-1}^{-1}(p) + (1-p_{t-1}^{-1}(p))(1-q))\end{aligned}$$

for $p > \mu$ and $t = \tau + 1,\ \tau + 2,\ \tau + 3, \ldots,$ and

$$(III)\ \pi_t(0) \ = 1 - \sum_{p>\mu_a} \pi_t(p),\ \alpha_t(p) \in [0,1]\,,\ \beta_t \in [0,1]\,,$$

for $t = \tau,\ \tau + 1,\ \tau + 2,\ \tau + 3, \ldots.$

To complete the proof of the theorem we can now show that the $\{\alpha_t(p)\}_{t=\tau}^{\infty}$ that solves the $SB\prime$ problem above also solves a modified first-best problem where the cost of exploration is $c^*(p)$. Notice that for every period $\tau > \widehat{t}$, the solution to the modified first-best problem is given by the solution to our original SB problem with the following adjustments: c is replaced everywhere by $c^*(p)$, the IC constraint (III) is deleted, and $\beta_t(0) = 1$ in all periods. It follows that the solution to the modified first-best problem is identical to the solution of the $SB\prime$ problem. Thus, the optimal mechanism (i.e., the solution to the original SB problem) does not randomize at $p_t > \mu_a$ and either $\varepsilon_t(p) = 1$ (for $0 < p < p_a^*$) or $\alpha_t(p) = 1$ (for $p > p_a^*$). ■