Optimal Design for Social Learning

Yeon-Koo Che* Johannes Hörner†

November 14, 2013

Abstract

The goal of a recommender system is to facilitate social learning about a product based on the experimentation by early users of the product. Without appropriating their social contribution, however, early users may lack the incentives to experiment on a product. The associated "cold start" could then result in a demise of a potentially valuable product and a collapse of the social learning. This paper studies design of the optimal recommender system focusing on this incentive problem and the pattern of dynamic social learning that emerges from the recommender system. The optimal design trades off fully transparent social learning to improve incentives for early experimentation, by selectively over-recommending a product in the early phase of the product release. Under the optimal scheme, experimentation occurs faster than under full transparency but slower than under the first-best optimum, and the rate of experimentation increases over an initial phase and lasts until the posterior becomes sufficiently bad in which case the recommendation stops along with experimentation on the product. Fully transparent recommendation may become optimal if the (socially-benevolent) designer does not observe the agents' costs or the agents choose the timing of receiving a recommendation.

Keywords: experimentation, social learning, mechanism design.

JEL Codes: D82, D83, M52.

1 Introduction

Most of our choices rely on recommendations by others. Whether it is for picking movies or stocks, choosing hotels or buying online, the experiences of others can help us greatly.

*Department of Economics, Columbia University, 420 West 118th Street, 1029IAB, New York, NY 10027, USA. Email: yeonkooche@gmail.com.

†Department of Economics, Yale University, 30 Hillhouse Ave., New Haven, CT 06520, USA, Email: johannes.horner@yale.edu.

Increasingly, internet platforms play the role of recommenders, enabling users to learn from other users in a scale never seen before. Amazon and Netflix are two best known mediators of social learning on books and movies, but for virtually any "experience" goods, there are platforms that mediate social learning among users: Yelp (for restaurants), TripAdvisors (for hotels), RateMD (for doctors), and RateMyProfessors (for professors) are just a few of many examples. Search engines, such as Google, Bing and Yahoo, organize social discovery of relevant websites based on "search experiences" of users themselves. Social networking sites such as Facebook and LinkedIn do the same for the other quintessential "experience" good: friends; as they link us to new candidates for friends based on the experiences of other friends.

While the economics of social learning is now well appreciated, the prior literature has focused only on its positive aspect.[1] A normative perspective—namely, how best to supervise users both to experiment with new products and to inform about their experiences to others—has been lacking. In particular, providing incentives for experimentation by early users is challenging, precisely because they do not internalize the benefit late users will reap. The lack of sufficient initial experimentation — known as the "cold start" problem — leads to a collapse of social learning, and to the death (even before takeoff) of products that should have been worthy of mainstream discovery. The cold start problem is particularly relevant for new unestablished products. With proliferation of self-production, such products are no longer an exception but rather are a rule.[2] In fact, the availability of social learning may worsen the incentives for early experimentation, since individuals would rather wait and free ride on information that others may provide.[3] That is, a recommender system may in fact crowd out the information production, undermining its foundation.

The current paper studies the design of a recommender system that optimally achieves the dual purpose of social learning. In keeping with the realism, we focus on the non-monetary tools for achieving them. Indeed, monetary transfers are seldom used for motivating experimentation. The reason may be because it is difficult to tell whether a reviewer performs experimentation conscientiously or submits an unbiased review; monetary transfers cannot incentivize a genuine experimentation. Instead, our key insight is that incentives for information acquisition can be best provided by the judicious use of the recommender system

[1] For instance, the main question asked in the literature was on under what circumstances, observational learning leads to full revelation of the underlying state.

[2] For instance, once considered vanity publishing, self-publication has exploded in recent years with the availability of easy typesetting and e-books media. Bowker Market Research estimates that in 2011 more than 300,000 self-published titles were issued (New York Times, "The Best Book Review Money Can Buy," August 25, 2012). While still at infancy, 3D printing and other similar technologies anticipate a future with an increased self-manufacturing of products. Proliferation of self production points means a market place populated by too many products/titles to be easily recognized, further increasing the importance of the recommender system.

[3] See Chamley and Gale (1994), and Gul and Lundholm (1995) for models illustrating this.

itself.

To fix an idea, suppose the recommender system, say an online movie platform, recommends a user movies that she will truly enjoy based on the reviews by the past users — call this truthful recommendation — for the most part, but mixes it with recommendation of new movies that need experimenting — call this fake recommendation or "spamming." As long as the platform keeps the users informed about whether recommendation is truthful or fake and as long as it commits not to "spam" too much, users will happily follow the recommendation and in the process perform the necessary experimentation. Indeed, we show that the optimal design of the recommender system trades off fully transparent social learning to improve incentives for early experimentation, by selectively over-recommending a product in the early phase of its product life cycle.

Of course, the extent to which information acquisition can be motivated in this way depends on the agent's cost of acquiring information, and the frequency with which the platform provides truthful recommendation (as opposed to fake recommendation). Also important is the dynamics of how the platform mixes truthful recommendation with the fake recommendation over time after the initial release of the product (e..g, movie release date). For instance, for an ex ante unappealing product, it is unlikely for many users even with low cost of experimentation to have experienced it immediately after its release, so recommending such a product in the early stage is likely to be met with skepticism. To be credible, therefore, the platform must commit to truthful recommendation with sufficiently high probability in the early stage of the product life, meaning that the designer will spam very little on such a product and the learning will be slow in the early stage; but over time, recommendation becomes credible so learning will speed up. This suggests that there will be a nontrivial dynamics in the optimal recommendation strategy as well as social learning.

The current paper seeks to explore how a recommender system optimally balances the tradeoff between experimentation and learning, and what kind of learning dynamics such a mechanism would entail and what implications they will have on the welfare, particularly when compared with the (i) No Social Learning benchmark (where there is no platform supervision of learning) and the (ii) Full Transparency (where the platform commits to always recommend truthfully). In the baseline analysis, we consider a platform that commits to maximize the welfare of its users. Social welfare maximization is an important normative benchmark that should be of any interest in any market design or public policy inquiry. From a more positive perspective, social welfare maximization could result from a Bertrand type competition.[4] We later relax the main assumptions, allowing for both profit maximization

[4]One can imagine that the each platform provides recommendation for a large set of products with varying vintages. The recommendation on a number of products with different vintages means that the social welfare gain from optimal experimentation will be spread evenly across users arriving in different times. For instance, a user who "sacrifices" for future users on some products will be over-compensated by the benefits from past learning on other products. Hence, firms competing in a Bertrand fashion by offering

and lack of commitment on the part of the designer.

While the optimal recommender system we identify rests on normative justification, its main feature is consistent with some aspects of the practice. For instance, search engines such as Google is known to periodically shuffle its rankings of search items to give exposure to relatively new unknown sites, very much consistent with the idea of distorting recommendation for the less exposed and lesser known items. More generally, our finding is consistent with the casual observations that ratings of many products appear to be inflated.[5] Ratings inflations are often committed by sellers (as opposed to the platforms) who have every interest to push their products even against the interests of the consumers.[6] But importantly, platforms have at their disposal control the degrees of ratings inflation; for instance, they can rely on filters detecting fake reviews and on the review of reviewers through purchase verification and the voting of "most helpful" reviews.[7] Our theory suggests that some toleration of reviews inflation can be a result of optimal supervision of social learning organized by a benevolent designer.

Our starting point is the standard "workhorse" model of experimentation, borrowed from Keller, Rady and Cripps (2005). The designer provides a good to agents whose binary value is unknown. By consuming this good, a possibly costly choice, short-run agents might find out whether the value is high or not. Here, we are not interested in the incentives of agents to report truthfully their experience to the designer: because they consume this good only once, they are willing to do so. But while agents do not mind reporting their experience, their decision to consume the good or not does not account for the benefits of experimentation. Importantly, agents do not communicate to each other directly. The designer mediates the information transmission. This gives rise to a difficult problem for the designer: How should she control social learning as to yield the right amount of experimentation?

Our model can also be viewed as introducing design into the standard model of social learning (hence the title). In standard models (for instance, Bikhchandani, Hirshleifer and Welsch, 1992; Banerjee, 1993), the sequence of agents take decisions myopically, ignoring the impact of their action on learning and future decisions and welfare. Here instead, the interaction between consecutive agents is mediated by the designer, who controls the flow of information. Such dynamic control is present in Gershkov and Szentes (2009), but that paper considers a very different environment, as there are direct payoff externalities (voting). Much closer is a recent working paper of Kremer, Mansour and Perry (2012). There, how-

a maximal (contemporaneous) social welfare will be forced to offer a maximal intertemporal social welfare on each product.

[5] Jindal and Liu (2008) find that, on Amazon, 60% of the reviews have a rating of 5.0, and roughly 45% products and 59% members have an average ranting of 5.

[6] Luca and Zervas (2013) suggest that as much as 16% of the Yelp reviews are suspect fraudulent.

[7] Mayzlin et al (2012) find that verification of stay required for a reviewer to leave a review on a hotel at Expedia resulted in fewer fake reviews at Expedia than at TripAdvisor which has no such requirement.

ever, learning is trivial: the quality of the good is ascertained as soon as a single consumer buys it. Finally, our paper joins a growing literature on Bayesian persuasion. This literature asks how a principal can credibly manipulate agents' beliefs to influence their behavior. See Aumann, Maschler and Stearns (1995) for a general analysis in the case of repeated games with incomplete information, and Kamenica and Gentzkow (2011), Rayo and Segal (2010), and Ostrovsky and Schwarz (2010) for applications into a variety of settings. The current paper pursues the same question in a dynamic setting. In our model, the designer manipulates the beliefs of agents so as to influence their experimentation behavior, and the agents' experimentation in turn determines the designer's belief and the optimal manipulation she chooses.[8]

2 Model

A product, say a "movie," is released at time $t = 0$, and, for each continuous time $t \geq 0$, a constant flow of unit mass of consumers arrive, having the chance to consume the product, i.e., watch the movie. In the baseline model, the consumers are short-lived, so they make one time decisions, and leave the market for good. (We later extend the model to allow the agents to choose the time after arrival at which they make an experimentation decision.) A consumer incurs the cost $c \in (0, 1)$ for watching the movie. The cost can be the opportunity cost of the time spent, or the price charged, for the movie. The movie is either "good," in which case a consumer derives the surplus of 1, or "bad," in which case the consumer derives surplus of 0. The quality of the movie is a priori uncertain but may be revealed over time. At time $t = 0$, the probability of the movie being good, or simply "the prior," is p_0. We shall consider all values of the prior, although the most interesting case will be $p_0 \in (0, c)$, so consumers would not consume given the prior.

Consumers do not observe the decisions and experiences by previous consumers. There is a **designer** who can mediate social learning by collecting information from previous consumers and disclosing that information to the current consumers. We can think of the designer as an Internet platform, such as Netflix, Google or Microsoft, who have access to users' activities and reviews, and based on this information, provide search guide and product recommendation to future users. As is natural with these examples, the designer may obtain information from its own marketing research or other sources, but importantly from the consumers' experiences themselves. For instance, there may be some flow of "fans" who try out the good at zero cost. We thus assume that some information arrives at a constant

[8]Ely, Frankel and Kamenica (2013) study design of optimal signal structure in a dynamic setting, but the information in their model does not have any consequence on behavior and thus entails no incentive issues. By contrast in our model, the information is of instrumental value, affecting both consumption and future information acquisition.

base rate $\rho > 0$ plus the rate at which consumers experience the good. Specifically, if a flow of size μ consumes the good over some time interval $[t, t + \mathrm{d}t)$, then the designer learns during this time interval that the movie is "good" with probability $\lambda_g(\rho + \mu)\mathrm{d}t$, that it is "bad" with probability $\lambda_b(\rho+\mu)\mathrm{d}t$, where $\lambda_g, \lambda_b \geq 0$, and ρ is the rate at which the designer obtains the information regardless of the consumers' behavior. The designer starts with the same prior p_0, and the consumers do not have access to the "free" learning.

The designer provides feedback on the movie to the consumers at each time, based on the information she has learnt so far. Since the decision for the consumers are binary, without loss, the designer simply decides whether to recommend the movie or not. The designer commits to the following policy: At time t, she recommends the movie to a fraction $\gamma_t \in [0, 1]$ of consumers if she learns the movie to be good, a fraction $\beta_t \in [0, 1]$ if she learns it to be bad, and she recommends or **spam** to fraction $\alpha_t \in [0, 1]$ if no news has arrived by t. We assume that the designer maximizes the intertemporal net surplus of the consumers, discounted at rate $r > 0$, over (measurable) functions $(\gamma_t, \beta_t, \alpha_t)$.

The information possessed by the designer at time $t \geq 0$ is succinctly summarized by the **designer's belief**, which is either 1 in case the good news has arrived, 0 in case the bad news has arrived by that time, or some $p_t \in [0, 1]$ in the event of no news having arrived by that time. The "no news" posterior, or simply **posterior** p_t must evolve according to Bayes rule. Specifically, suppose for time interval $[t, t + \mathrm{d}t)$, there is a a flow of learning by the designer at the rate μ_t, (including both the "free" learning ρ and the flow α_t of agents experimenting) during the period. Suppose no news has arrived by $t+\mathrm{d}t$, then the designer's updated posterior at time $t + \mathrm{d}t$ must be

$$p_t + \mathrm{d}p_t = \frac{p_t(1 - \lambda_g\mu_t\mathrm{d}t)}{p_t(1 - \lambda_g\mu_t\mathrm{d}t) + (1 - p_t)(1 - \lambda_b\mu_t\mathrm{d}t)}.$$

Rearranging and simplifying, the posterior must follow the law of motion:[9]

$$\dot{p}_t = -(\lambda_g - \lambda_b)\mu_t p_t(1 - p_t), \tag{1}$$

with the initial value at $t = 0$ given by the prior p_0. It is worth noting that the evolution of the posterior depends on the relative speed of the good news arrival versus the bad news arrival. If $\lambda_g > \lambda_b$ (so the good news arrive faster than the bad news), then "no news" leads the designer to form a pessimistic inference on the quality of the movie, with the posterior falling. By contrast, if $\lambda_g < \lambda_b$, then "no news" leads to on optimistic inference, with the

[9] Subtracting p_t from both sides and rearranging, we get

$$\mathrm{d}p_t = -\frac{(\lambda_g - \lambda_b)\mu_t p_t(1 - p_t)\mathrm{d}t}{p_t(1 - \lambda_g\mu_t\mathrm{d}t) + (1 - p_t)(1 - \lambda_b\mu_t\mathrm{d}t)} = -(\lambda_g - \lambda_b)\mu_t p_t(1 - p_t)\mathrm{d}t + o(\mathrm{d}t),$$

where $o(\mathrm{d}t)$ is a term such that $o(\mathrm{d}t)/\mathrm{d}t \to 0$ as $\mathrm{d}t \to 0$.

posterior rising. We label the former case **good news** case and the latter **bad news** case. (Note that the former case includes the special case of $\lambda_b = 0$, a pure good news case, and the latter includes $\lambda_g = 0$, a pure bad news case.)

In our model, the consumers do not directly observe the designer's information, or her belief. They can form a rational belief, however, on the designer's belief. Let g_t and b_t denote the probability that the designer's belief is 1 and 0, respectively. Just as the designer's belief evolves, the consumers' belief on the designer's belief evolves as well, depending on the rate at which the agents (are induced to) experiment. Specifically, given the experimentation rate μ_t,

$$\dot{g}_t = (1 - g_t - b_t)\lambda_g \mu_t p_t, \tag{2}$$

with the initial value $g_0 = 0$, and

$$\dot{b}_t = (1 - g_t - b_t)\lambda_b \mu_t (1 - p_t), \tag{3}$$

with the initial value $b_0 = 0$.[10] Further, these beliefs must form a martingale:

$$p_0 = g_t \cdot 1 + b_t \cdot 0 + (1 - g_t - b_t)p_t. \tag{4}$$

The designer chooses the policy (α, β, γ), measurable, to maximize social welfare, namely

$$\mathcal{W}(\alpha, \beta, \chi) := \int_{t \geq 0} e^{-rt} g_t \gamma_t (1 - c)\mathrm{d}t + \int_{t \geq 0} e^{-rt} b_t \beta_t (-c)\mathrm{d}t + \int_{t \geq 0} e^{-rt}(1 - g_t - b_t)\alpha_t(p_t - c)\mathrm{d}t,$$

where (p_t, g_t, b_t) must follow the required laws of motion: (1), (2), (3), and (4), where $\mu_t = \rho + \alpha_t$ is the total experimentation rate and r is the discount rate of the designer.[11]

In addition, for the policy (α, β, γ) to be implementable, there must be an incentive on the part of the agents to follow the recommendation. Given policy (α, β, γ), conditional on being recommended to watch the movie, the consumer will have the incentive to watch the movie, if and only if the expected quality of the movie—the posterior that it is good—is no

[10] These formulae are derived as follows. Suppose the probability that the designer has seen the good news by time t and the probability that she has seen the bad news by t are respectively g_t and b_t. Then, the probability of the good news arriving by time $t + \mathrm{d}t$ and the probability of the bad news arriving by time $t + \mathrm{d}t$ are respectively

$$g_{t+\mathrm{d}t} = g_t + \lambda_g \mu_t p_t \mathrm{d}t(1 - g_t - b_t) \text{ and } b_{t+\mathrm{d}t} = b_t + \lambda_b \mu_t (1 - p_t)\mathrm{d}t(1 - g_t - b_t).$$

Dividing these equations by $\mathrm{d}t$ and taking the limit as $\mathrm{d}t \to 0$ yields (2) and (3).

[11] More precisely, the designer is allowed to randomize over the choice of policy (α, β) (using a relaxed control, as such randomization is defined in optimal control). A corollary of our results is that there is no gain for him from doing so.

less than the cost:

$$\frac{g_t\gamma_t + (1-g_t-b_t)\alpha_t p_t}{g_t\gamma_t + b_t\beta_t + (1-g_t-b_t)\alpha_t} \geq c. \tag{5}$$

Since the agents do not directly access the news arriving to the designer, so the exact circumstances of the recommendation—whether the agents are recommended because of good news or despite no news—is kept hidden, which is why the incentives for following the recommendation is based on the posterior formed by the agents on the information of the designer. (There is also an incentive constraint for the agents not to consume the good when not recommended by the designer. Since this constraint will not bind throughout, as the designer typically desires more experimentation than the agents, we shall ignore it.)

Our goal is to characterize the optimal policy of the designer and the pattern of social learning it induces. To facilitate this characterization, it is useful to consider three benchmarks.

- **No Social Learning (NSL)**: In this regime, the consumers receive no information from the designer, so they decide based on the prior p_0. Since $p_0 < c$, no consumer ever consumes.
- **Full Transparency (FT)**: In this regime, the designer discloses her information, or her beliefs, truthfully to the consumers. In our framework, the full disclosure can be equivalently implemented by the policy of $\gamma_t \equiv 1, \beta_t \equiv 0$ and $\alpha_t = \mathbf{1}_{\{p_t \geq c\}}$.
- **First-Best (FB)**: In this regime, the designer optimizes on her policy, without having to satisfy the incentive compatibility constraint (5).

To distinguish the current problem relative to the first-best, we call the optimal policy maximizing $\mathcal{W}$ subject to (1), (2), (4) and (5), the **second-best** policy.

Before proceeding, we observe that it never pays the designer to lie about the news if they arrive.

Lemma 1. *It is optimal for the designer to disclose the breakthrough (both good and bad) news immediately. That is, an optimal policy has $\gamma_t \equiv 1, \beta_t \equiv 0$.*

Proof. If one raises γ_t and lowers β_t, it can only raise the value of objective $\mathcal{W}$ and relax (5) (and do not affect other constraints). □

Lemma 1 reduces the scope of optimal intervention by the designer to choosing α, the recommendation policy following "no news." In the sequel, we shall thus fix $\gamma_t \equiv 1, \beta_t \equiv 0$ and focus on α as the sole policy instrument.

3 Optimal Recommendation Policy

We begin by characterizing further the process by which the designer's posterior, and the agents' beliefs over designer's posterior, evolve under arbitrary policy α. To understand how the designer's posterior evolves, it is convenient to work with the likelihood ratio $\ell_t = \frac{p_t}{1-p_t}$ of the posterior p_t. Given the one-to-one correspondence between the two variables, we shall refer to ℓ simply as a "posterior" when there is little cause for confusion. It then follows that (1) can be restated as:

$$\dot{\ell}_t = -\ell_t \Delta \lambda_g \mu_t, \quad \ell_0 := \frac{p_0}{1-p_0}, \tag{6}$$

where $\Delta := \frac{\lambda_g - \lambda_b}{\lambda_g}$, assuming for now $\lambda_g > 0$.

The two other state variables, namely the posteriors g_t and b_t on the designer's belief, are pinned down by ℓ_t (and thus by p_t) at least when $\lambda_g \neq \lambda_b$ (i.e., when no news is not informationally neutral.) (We shall remark on the case of the neutrality case $\Delta = 0$.)

Lemma 2. *If $\Delta \neq 0$, then*

$$g_t = p_0 \left(1 - \left(\frac{\ell_t}{\ell_0}\right)^{\frac{1}{\Delta}}\right) \text{ and } b_t = (1-p_0)\left(1 - \left(\frac{\ell_t}{\ell_0}\right)^{\frac{1}{\Delta}-1}\right).$$

This result is remarkable. A priori, there is no reason to expect that the designer's belief p_t serves as a "sufficient statistic" for the posteriors that the agents attach to the arrival of news, since different histories for instance involving even different experimentation over time could in principle lead to the same p.

It is instructive to observe how the posterior on the designer's belief evolves. At time zero, there is no possibility of any news arriving, so the posterior on the good and bad news are both zero. As time progresses without any news arriving, the likelihood ratio either falls or rises depending on the sign of Δ. Either way, both posteriors rise. This enables the designer to ask credibly the agents to engage in costly experimentation with an increased probability. To see this specifically, substitute g_t and b_t into (5) to obtain:

$$\alpha_t \leq \bar{\alpha}(\ell_t) := \min\left\{1, \frac{\left(\frac{\ell_t}{\ell_0}\right)^{-\frac{1}{\Delta}} - 1}{k - \ell_t}\ell_t\right\}, \tag{7}$$

if the normalized cost $k := c/(1-c)$ exceeds ℓ_t and $\bar{\alpha}(\ell_t) := 1$ otherwise.

The next lemma will figure prominently in our characterization of the second-best policy later.

LEMMA 3. *If $\ell_0 < k$ and $\Delta \neq 0$, then $\bar{\alpha}(\ell_t)$ is zero at $t = 0$, and increasing in t, strictly so whenever $\bar{\alpha}(\ell_t) \in [0, 1)$.*[12]

At time zero, the agents have no incentive for watching the movie since the good news could never have arrived instantaneously, and their prior is unfavorable. Interestingly, the agents can be asked to experiment more over time, even when $\Delta > 0$, in which case the posterior ℓ_t falls over time! If $\lambda_g > 0$, the agents attach increasingly higher posterior on the arrival of good news as time progresses. The "slack incentives" from the increasingly probable good news can then be shifted to motivate the agents' experimentation when there is no news.

Substituting the posteriors from Lemma 2 into the objective function and using $\mu_t = \rho + \alpha_t$, and with normalization of the objective function, the second-best program is restated as follows:

$$[SB] \qquad \sup_{\alpha} \int_{t \geq 0} e^{-rt} \ell_t^{\frac{1}{\Delta}} \left(\alpha_t \left(1 - \frac{k}{\ell_t} \right) - 1 \right) \mathrm{d}t$$

subject to

$$\dot{\ell}_t = -\Delta \lambda_g (\rho + \alpha_t) \ell_t, \tag{8}$$

$$0 \leq \alpha_t \leq \bar{\alpha}(\ell_t). \tag{9}$$

Obviously, the first-best program, labeled $[FB]$, is the same as $[SB]$, except that the upper bound for $\bar{\alpha}(\ell_t)$ is replaced by 1. We next characterize the optimal recommendation policy. The precise characterization depends on the sign of Δ, i.e., whether the environment is that of predominantly good news or bad news.

3.1 "Good news" environment: $\Delta > 0$

In this case, as time progresses with no news, the designer becomes pessimistic about the quality of the good. Nevertheless, as observed earlier, the agents' posterior on the arrival of good news improves. This enables the designer to incentivize the agents to experiment increasingly over time. The designer can accomplish this through "spamming" — by recommending agents to watch even when no good news has arrived. Such a policy "pools" recommendation across two very different circumstances; one where the good news has arrived and one where no news has arrived. Although the agents in the latter circumstance will

[12] The case $\Delta = 0$ is similar: the same conclusion holds but $\bar{\alpha}$ need to be defined separately.

never follow the recommendation wittingly, pooling the two circumstances for recommendation enables the designer to siphon the slack incentives from the former circumstance to the latter, and thus can incentivize the agents to "experiment" for the future generation, so long as the recommendation in the latter circumstance is kept sufficiently infrequent/improbable. Since the agents do not internalize the social benefit of experimentation, spamming becomes a useful tool for the designer's second-best policy.

Whether and to what extent the designer will induce such an experimentation depends on the tradeoff between the cost of experimentation for the current generation and the benefit of social learning the experimentation will yield for the future generation of the agents. And this tradeoff depends on the posterior p. To fix the idea, consider the first-best problem and assume $\rho = 0$, so there is no free learning. Suppose given posterior p, the designer contemplates whether to experiment a little longer or to stop experimentation altogether for good. This is indeed the decision facing the designer given the cutoff decision rule: intuitively once experimentation becomes undesirable, it can only remain undesirable with a worsened posterior. If she induces experimentation a little longer, say by $\mathrm{d}t$, the additional experimentation will cost $(c-p)\mathrm{d}t$. But, the additional experimentation may bring a good news in which case the future generation of agents will enjoy the benefit of $v := \int e^{-rt}(1-c)\mathrm{d}t = (1-c)/r$. Since the good news will arrive at the rate $\lambda_g \mathrm{d}t$, but only if the movie is good, the probability of which is the posterior p, the expected benefit from the additional experimentation is $v\lambda_g p\mathrm{d}t$. Hence, the additional experimentation is desirable if and only if $v\lambda_g p \geq c - p$, or $p \geq c\left(1 - \frac{v}{v+\frac{1}{\lambda_g}}\right)$. The same tradeoff would apply to the second-best scenario, except that absent free learning, the designer can never motivate the agents to experiment; that is, $\bar{\alpha}(p) = 0$, so the threshold posterior is irrelevant.

For the general case with free learning $\rho > 0$, the optimal policy is described as follows.

PROPOSITION 1. *The second-best policy prescribes, absent any news, the maximal experimentation at $\alpha(p) = \bar{\alpha}(p)$ until the posterior falls to p_g^*, and no experimentation $\alpha(p) = 0$ thereafter for $p < p_g^*$, where*

$$p_g^* := c\left(1 - \frac{rv}{\rho + r(v + \frac{1}{\lambda_g})}\right),$$

where $v := \frac{1-c}{r}$ is the continuation payoff upon the arrival of good news. The first-best policy has the same structure with the same threshold posterior, except that $\bar{\alpha}(p)$ is replaced by 1. If $p_0 \geq c$, then the second-best policy implements the first-best, where neither NSL nor FT can. If $p_0 < c$, then the second-best induces a slower experimentation/learning than the first-best.

The detailed analysis generating the current characterization (as well as the subsequent ones) is provided in the Appendix. Here we provide some intuition behind the result. With a free learning at rate $\rho > 0$, stopping experimentation does not mean abandoning all learning. The good news that may be learned from additional experimentation may be also learned in the future for free. One must thus discount the value of learning good news achieved by additional experimentation by the rate at which the same news will be learned in the future for free: $\frac{\lambda_g \rho}{r}$. The threshold posterior can be seen to equate this adjusted value of experimentation with its flow cost:[13]

$$\underbrace{\lambda_g p v \left(\frac{1}{(\lambda_g \rho / r) + 1} \right)}_{\text{value of experimentation}} = \underbrace{c - p}_{\substack{\text{cost of} \\ \text{experimentation}}} .$$

Note the opportunity cost of experimenting is the same for first-best and second-best scenario, since the consequence of stopping the experimentation is the same in both cases. Hence, the first-best policy has the same threshold p_g^*, but the rate of experimentation is $\alpha = 1$ when the posterior is above the threshold level. If $p_0 \geq c$, then since $\bar{\alpha}(p) = 1$ for all p, the second-best implements the first-best. Note that spamming is necessary to attain the first-best. Since $p_g^* < c$, the policy calls for experimentation even when it is myopically suboptimal. This means that FT cannot implement the first-best even in this case since under FT agents will stop experimenting before p reaches p_g^*.

Suppose next $p_0 \in (p_g^*, c)$. Then, the second-best cannot implement the first-best. The rate of experimentation is larger under first-best than under second-best, so the threshold posterior is reached sooner under FB than under SB, and of course the news arrives sooner, thus enabling the social learning benefit to materialize sooner (both in the sense of stochastic dominance), under FB than under SB. Another difference is that the rate of experimentation at a given posterior p depends on the prior p_0 in the SB (but not in FB). The reason that, for the same posterior $p > p_g^*$, the designer would have built more credibility of having received the good news so she can ask the agents to experiment more, the higher the prior is.

Figure 1 plots the time it takes to reach the threshold posterior under FB and SB. Clearly, the experimentation induced under FB is front-loaded and thus monotonic. Interestingly, the experimentation induced by SB is front-loaded but hump-shaped. Closer to the release date, the arrival of good news is very unlikely, so the unfavorable prior means that the agents can hardly be asked to experiment. Consequently, experimentation takes off very slowly under SB. As time progresses, the agents attach increasingly higher probability on the arrival of good news, increasing the margin for the designer to leverage the slack incentives from the good news event to encourage agents to experiment even when no news has arrived. The experimentation rate accordingly picks up and increases, until the posterior falls below p_g^*,

[13] The validity of this interpretation rests on the necessary condition of the optimal control problem analyzed in the appendix. It also can be found from the dynamic programming heuristic, which is available upon request.

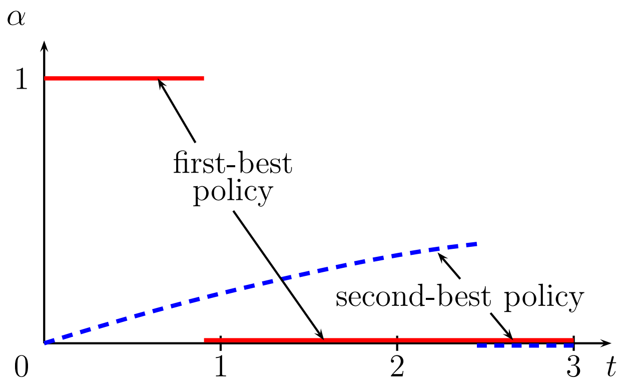

Figure 1: Path of α for $\Delta > 0$ and $(c, \rho, p_0, r, \lambda_g, \lambda_b) = (2/3, 1/4, 1/2, 1/10, 4/5, 0)$.

at which point all experimentation stops.

Interestingly, the arrival rate of bad news λ_b does not affect the threshold posterior p_g^*. This is because the tradeoff does not depend on the arrival rate of bad news. But the arrival rate of the bad news does affect both the duration and the rate of incentive-compatible experimentation. As can be seen from (1), as λ_b rises (toward λ_g), it slows down the decline of the posterior. Hence, it takes a longer time for the posterior to reach the threshold level, meaning that the agents are induced to experiment longer (until the news arrive or the threshold p^* is reached), holding constant the per-period experimentation rate. Meanwhile, the incentive compatible rate of experimentation $\bar{\alpha}(p)$ increases with λ_b, since the commitment never to recommend in the event of bad news means that a recommendation is more likely to have been a result of a good news. Hence, experimentation rises in two different senses when λ_b rises.

Next, Proposition 1 also makes the comparison with the other benchmarks clear. First, recall that both NSL and FT involve no experimentation by the agents (i.e., they only differ in that the FT enables social learning once news arrives whereas the NSL does not). By contrast, as long as $p_0 \in (p_g^*, c)$, SB involves nontrivial experimentation, so SB produces strictly more information than either regime, and it enables full social learning when good news arrives. In fact, since full transparency is a feasible policy option under SB, this suggests that SB strictly dominates FT (which in turn dominates NSL).

3.2 "Bad news" environment: $\Delta < 0$

In this case, the designer grows optimistic over time about the quality of the good absent no news. Likewise, the agents also grow optimistic over time from no breakthrough news under full transparency. In this sense, unlike the good news case, the incentive conflict between the designer and the agents is lessened in this case. The conflict does not disappear altogether, however, since the agents still do not internalize the social benefit from their experimentation. For this reason, "spamming" proves to be valuable to the designer even in this case.

The logic of the optimal recommendation policy is similar to the good news case. Namely, it is optimal for the agents to experiment if and only if the (true) posterior is higher than some threshold (as is shown in the Appendix). This policy entails a different experimentation pattern in terms of time, however; now the experimentation is back-loaded rather than front-loaded (which was the case with good news). This also changes the nature of the learning benefit at the margin, and this difference matters for design of the recommendation policy.

To appreciate this difference, consider the first-best problem in a simplified environment in which there is no free learning (i.e., $\rho = 0$) and no arrival of good news (i.e., $\lambda_g = 0$). Suppose as before the designer contemplates whether to engage the agents in experimentation or not, at a given posterior p. If she does not trigger experimentation, there will be no more experimentation in the future (given the structure of the optimal policy), so the payoff is zero. If she does trigger experimentation, likewise, there will continue to be an experimentation unless bad news arrives (in which case the experimentation will be stopped for good). The payoff from such an experimentation consists of two terms:

$$\frac{p-c}{r} + (1-p)\left(\frac{c}{r}\right)\left(\frac{\lambda_b}{r+\lambda_b}\right).$$

The first term represents the payoff from consuming the good irrespective of the bad news. The second term captures the saving of the cost by stopping consumption whenever the bad news arrives.[14] In this simple case, therefore, the first-best policy prescribes full experimentation if and only if this payoff is nonnegative, or $p \geq c\left(1 - \frac{\lambda_b(1-rc)}{r+\lambda_b(1-rc)}\right)$.

This reasoning reveals that the nature of learning benefit is crucially different here. In the good news case, at the margin the default is to stop watching the movie, so the benefit of learning was to trigger (permanent) "consumption of the good movie." By contrast, the benefit of learning here is to trigger (permanent) "avoidance of the bad movie," since at the margin the default is to watch the movie. With free learning $\rho > 0$ and good news $\lambda_g \in (0, \lambda_b)$, the same learning benefit underpins the tradeoff both in the first-best and second-best policy, but the tradeoff is moderated by two additional effects: (1) free learning introduces opportunity cost of experimentation, which reduces its appeal and thus increases the threshold posterior and time for triggering experimentation;[15] and (2) good news may

[14]The bad news arrives only if the good is bad (whose probability is $1-p$) with (the time discounted) probability $\frac{\lambda_b}{r+\lambda_b}$, and once it arrives, there is a permanent cost saving of c/r, hence the second term.

[15]Due to free learning, the "no experimentation" phase is never "absorbing." That is, the posterior will continue to rise and eventually pass the critical value, thus triggering the experimentation phase. This feature contrasts with the "breakdowns" case of Keller and Rady (2012).

trigger permanent conversion even during no experimentation phase. The result is presented as follows.

PROPOSITION 2. *The first-best policy (absent any news) prescribes no experimentation until the posterior p rises to p_b^{**}, and then full experimentation at the rate of $\alpha(p) = 1$ thereafter, for $p > p_b^{**}$, where*

$$p_b^{**} := c\left(1 - \frac{rv}{\rho + r(v + \frac{1}{\lambda_b})}\right).$$

The second-best policy implements the first-best if $p_0 \geq c$ or if $p_0 \leq \hat{p}_0$ for some $\hat{p}_0 < p_b^{**}$. If $p_0 \in (\hat{p}_0, c)$, then the second-best policy prescribes no experimentation until the posterior p rises to p_b^*, and then maximal experimentation at the rate of $\bar{\alpha}(p)$ thereafter for any $p > p_b^*$, where $p_b^* > p_b^{**}$. In other words, the second-best policy triggers experimentation at a later date and at a lower rate than does the first-best.

The proof of the proposition as well as the precise formula for $\hat{p}_0$ is in the appendix. Here we provide some explanation for the statement.

The first-best policy calls for the agents to start experimenting *strictly before* the true posterior rises to c, namely when $p_b^{**} < c$ is reached. Obviously, if $p_0 \geq c$, then since the posterior can only rise, absent any news, the posterior under full transparency is always above the cost, so FT is enough to induce the first-best outcome. Recall that FT can never implement first-best even for $p_0 \geq c$ in the good news case.

If $p_0 < c$, then the optimal policy cannot be implemented by full transparency. In order to implement such a policy, the designer must again rely on spamming, recommending the good even when the good news has not arrived. Unlike the case of $\rho > 0$, the first-best may be implementable as long as the initial prior p_0 is sufficiently low. In that case, it takes a relatively long time for the posterior to reach p_b^{**}, so by the time the critical posterior is reached, the agents attach a sufficiently high posterior on the arrival of good news that the designer's recommendation becomes credible.

If $p_0 \in (\hat{p}_0, c)$, then the first-best is not attainable even with spamming. In this case, the designer triggers experimentation at a higher posterior and thus at a later date than she does under the first-best policy. This is because the scale of subsequent experimentation is limited by incentive compatibility, which lowers the benefit from triggering experimentation. Since there is less benefit to be had from triggering experimentation, the longer will the designer hang on to free learning than in first best.[16] Figure 2 depicts the second-best recommendation policy for this case.

[16]This feature contrasts with the result of $\rho > 0$. The posterior at which to stop experimentation was the same in the case between second-best and first-best regimes, since the consequence of stopping experimentation was the same. This result changes when there are heterogeneous observable costs, as will be seen later.

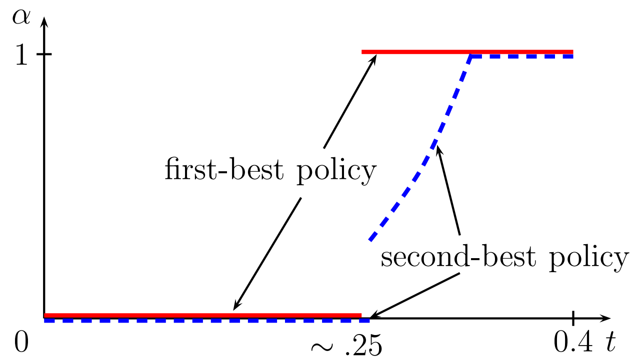

Figure 2: Path of α for $\Delta < 0$ and $(c, \rho, p_0, r, \lambda_g, \lambda_b) = (1/2, 1, 2/7, 1/10, 1, 2))$.

The comparative statics with respect to parameters such as λ_b, r and ρ is immediate from the inspection of the threshold posterior. Its intuition follows also naturally from the reasoning provided before the proposition. The higher λ_b and r and the lower ρ are, the higher is the "net" benefit from experimentation, so the designer triggers experimentation sooner.

3.3 "Neutral news" environment: $\Delta = 0$

In this case, the designer's posterior on the quality of the good remains unchanged in the absence of breakthrough news. Experimentation could be still desirable for the designer. If $p_0 \geq c$, then the agents will voluntarily consume the good, so experimentation is clearly self-enforcing. If $p_0 < c$, then the agents will not voluntarily consume, so spamming is needed to incentivize experimentation. As before the optimal policy has the familiar cutoff structure.

Proposition 3. *The second-best policy prescribes, absent any news, the maximal experimentation at $\alpha(p_0) = \bar{\alpha}(p_0)$ if $p_0 < p_0^*$, and no experimentation $\alpha(\ell) = 0$ if $p < p_0^*$, where $p_0^* = p_g^* = p_b^*$. The first-best policy has the same structure with the same threshold posterior, except that $\bar{\alpha}(p_0)$ is replaced by 1. The first-best is implementable if and only if $p_0 \geq c$.*

3.4 Heterogenous observable costs

While the analysis of the bad and good news case brings out some common features of supervised social learning, such as increasing experimentation early on, as well as delay, one feature that apparently distinguishes the two cases is that the *belief* level at which experimentation stops in the good news case is the socially optimal one, while –except when first-best is achievable– experimentation starts too late in the bad news case (even in terms of beliefs). Here, we argue that the logic prevailing in the bad news case is the robust one.

In particular, a similar phenomenon arises with good news once we abandon the extreme assumption that all regular agents share the same cost level.

To make this point most clearly, consider the case of "pure" good news: $\lambda^b = 0$, $\lambda := \lambda^g > 0$. Suppose agents come in two varieties, or types $j = L, H$. Different types have different costs, with $0 < \ell_0 < k_L < k_H$, where, as before, $k_j = \frac{c_j}{1-c_j}$. Hence, we assume that at the start, neither type of agent is willing to buy. The flow mass of agents of type j is denoted q_j, with $q_L + q_H = 1$.

Furthermore, assume here that the cost is observable to the designer, so that she can condition her recommendation on this cost. This implies that, conditional on her posterior being $\ell_t < 1$, she can ask an agent of type j to buy with a probability up to

$$\bar{\alpha}_j(\ell_t) := \min\left\{1, \frac{\left(\frac{\ell_t}{\ell_0}\right)^{-\frac{1}{\Delta}} - 1}{k_j - \ell_t}\ell_t\right\}, \tag{10}$$

as per (7). The following proposition elucidates the structure of the optimal policy. As before, we index thresholds by either one or two asterisks, according to whether this threshold pertains to the second- or first-best policy.

PROPOSITION 4. *Both the first-best and second-best policies are characterized by a pair of thresholds* $0 \le \ell_L \le \ell_H \le \ell_0$, *such that (i) all agents are asked to experiment with maximum probability for* $\ell \ge \ell_H$; *(ii) only low cost agents experiment (with maximum probability) for* $\ell \in [\ell_L, \ell_H)$; *(iii) no agent experiments for* $\ell < \ell_L$. *Furthermore,* $\ell_L^{**} = \ell_L^*$, *and* $\ell_H^{**} \ge \ell_H^*$, *with a strict inequality whenever* $\ell_0 > \ell_H^{**}$.

Not surprisingly, the lower common threshold is the threshold that applies whenever there is only type of agent, namely the low-cost agent. There is no closed-form formula for the upper threshold (although there is one for its limit as $r \to 0$).

More surprising is the fact that, despite experimenting with a lower intensity in the second-best, the designer chooses to call upon high-cost agents to experiment at beliefs *below* the level at which she would do so in the first-best. She induces them to experiment "more" than they should do, absent the incentive constraint. This is because of the incentive constraint of the low-cost agents: as she cannot call upon them to experiment as much as she would like, she hangs on to the high-cost type longer (*i.e.*, for lower beliefs) than she would otherwise. This is precisely what happened in the bad news case: because agents are only willing to buy with some probability in the second-best, the designer hangs on the "free" experimentation provided by the flow ρ a little longer there as well.

Let us turn to the case of a continuum of costs, to see how the structure generalizes. Agents' costs are uniformly distributed on $[0, \bar{c}]$, $\bar{c} \le 1$. The principal chooses a threshold

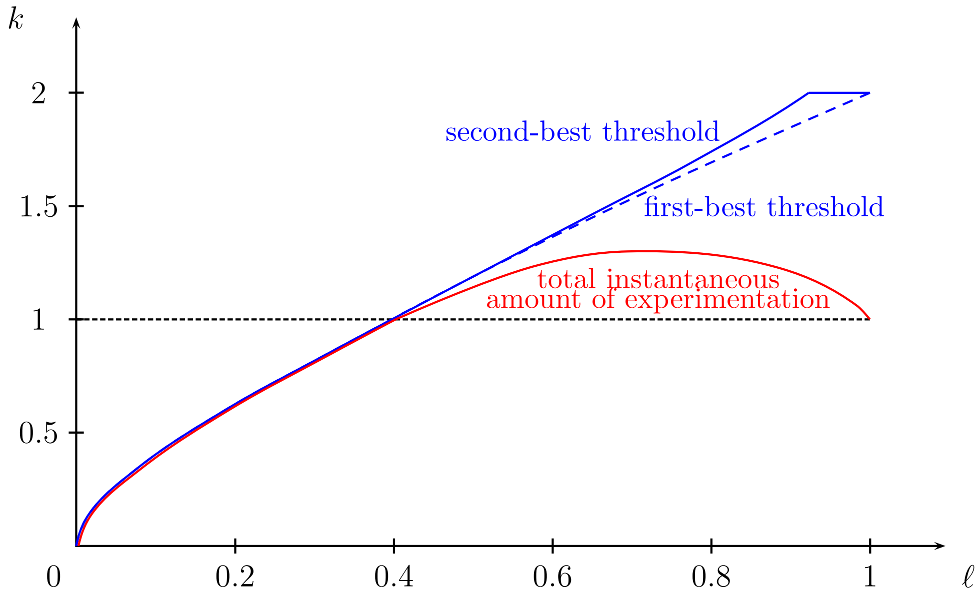

Figure 3: Optimal policy with a continuum of cost levels (here, $\bar{k} = 2, \ell_0 = 1$).

$k_t \in [0, \bar{k}]$ (where $\bar{k} := \bar{c}/(1-\bar{c})$) such that, when the principal's belief is p_t, an agent with $k \leq k_t$ is recommended to buy with probability $\alpha_t(k)$ or 1, depending upon whether $k \leq \ell_0$ or not, while agents with higher cost types are recommended not to buy. (When the principal's belief is 1, all types of agents are recommended to buy.) Clearly, k_t is decreasing in time. Let $t_1 := \inf\{t : k_t = \ell_0\}$. As some agents have arbitrarily low cost levels, we may set $\rho = 0$.

The optimal policy can be studied via optimal control. Appendix 6 provides the details for the second-best (the first-best being a simpler case in which $\bar{a} = 1$). Figure 3 illustrates the main conclusions for some choice of parameters. As time passes, ℓ and the upper threshold of the designer γ decrease. At some time (t_1), γ hits ℓ_0 (here, for $\ell \simeq .4$). For lower beliefs, the principal's threshold coincides with the first-best. Before that time, however, the threshold that he picks lies above the first-best threshold for this belief; for these parameters, the designer encourages all types to buy (possibly with some very small probability) when the initial belief is high enough (for ℓ above .9). Although all types are solicited, the resulting amount of experimentation is small early on, because types above l_0 are not willing to buy with high probability. As a result, the amount of experimentation is not monotonic: it first increases, and then decreases. The dotdashed line in Figure 3 represents the total amount of experimentation: it is below the first-best amount, as a function of the belief, until time t_1.

4 Endogenous entry

We now consider the case in which agents can decide *when* to get a recommendation. Agents arrive at a unit flow rate over time, and an agent arriving at time t can choose to get a recommendation at any date $\tau \geq t$ (possibly, $\tau = +\infty$). Of course, if agents could "continuously" get recommendations for free until they decide to purchase, if ever, it would be weakly dominant to do so. Here instead, we assume that it is sufficiently costly to get a recommendation that agents get only one, although we will ignore the cost from actually getting a recommendation. Given this, there is no benefit in delaying the decision to buy or not beyond that time. Hence, an agent born at time t chooses a stopping time $\tau \geq t$ at which to get a recommendation ("checking in"), as well as a decision to buy at that time, as a function of the recommendation he gets. Between the time the agent is born and the time he checks is, he receives no information. Agents share the designer's discount rate r.

We restrict attention to the case of "pure" good news: $\lambda = \lambda^g > 0 = \lambda^b$. All agents share the same cost $c > p_0$ of buying. Recommendations α are a function of time and the designer's belief, as before (they cannot be conditioned on a particular agent's age, assumed to be private information).

Hence, an agent maximizes the payoff

$$\max_{\tau \geq t} e^{-r\tau} \left(g_\tau(1-c) + (1-g_\tau)\alpha_\tau(p_\tau - c) \right).$$

Substituting k, ℓ_0 and ℓ, this is equivalent to maximizing

$$\mathcal{U}_\tau := e^{-r\tau} \left(\ell_0 - \ell_\tau - \alpha_\tau(k - \ell_\tau) \right).$$

Suppose first full transparency, that is, $\alpha \equiv 0$, and so $\mathcal{U}_t = e^{-rt}(\ell_0 - \ell_t)$. Note that the timing at which agents check in is irrelevant for belief updating (because those who check in never experiment), so that

$$\ell_t = \ell_0 e^{-\lambda\rho t}.$$

The function $\mathcal{U}_t$ is quasi concave in t, with a maximum achieved at the time

$$t^* = -\frac{1}{\rho\lambda} \ln \frac{\ell^*}{\ell_0}, \quad \ell^* := \frac{r\ell_0}{r + \lambda\rho}.$$

The optimal strategy of the agents is then intuitive: an agent born before t^* waits until time t^*, trading off the benefits from a more informed decision with his impatience; an agent arriving at a later time checks in immediately. Perhaps surprisingly, the cost c is irrelevant for this decision: as the agent only buys if he finds out that the posterior is high, the surplus $(1-c)$ only scales his utility, without affecting his preferred timing.

Note that

$$\ell_0 - \ell^* = \frac{\ell_0}{\frac{r}{\rho\lambda} + 1}, \quad e^{-rt^*} = \left(\frac{\ell^*}{\ell_0}\right)^{\frac{r}{\rho\lambda}},$$

so that in fact his utility is only a function of his prior and the ratio $\frac{r}{\rho\lambda}$. In fact, ρ plays two roles: by increasing the rate at which learning occurs, it is equivalent to lower discounting (hence the appearance of r/ρ); in addition, it provides an alternative and cheaper way of learning to the agents consuming (holding fixed the total "capacity" of experimentation). To disentangle these two effects, we hold fixed the ratio $\frac{r}{\lambda\rho}$ in what follows.

As a result of this waiting by agents that arrive early, a queue Q_t of agents forms, which grows over time, until t^* at which it gets resorbed. That is,

$$Q_t = \begin{cases} t & \text{if } t < t^*; \\ 0 & \text{if } t \geq t^*, \end{cases}$$

with the convention that Q is a right-continuous process.

To build our intuition about the designer's problem, consider first the first-best problem. Suppose that the designer can decide when agents check in, and whether they buy or not. However, we assume that the timing decision cannot be made contingent on the actual posterior belief, as this is information that the agents do not possess. Plainly, there is no point in asking an agent to wait if he were to be instructed to buy independently of the posterior once he checks in. Agents that experiment do not wait. Conversely, an agent that is asked to wait –and so does not buy if the posterior is low once he checks in– exerts no externality on other agents, and the benevolent designer might as well instruct him to check in at the agent's favorite time.

It follows that the optimal policy of the (unconstrained) designer must involve two times, $0 \leq t_1 \leq t_2$, such that agents that arrive before time t_1 are asked to experiment; agents arriving later only buy if the posterior is one, with agents arriving in the interval $(t_1, t_2]$ waiting until time t_2 to check in, while later agents check in immediately. Full transparency is the special case $t_1 = 0$, as then it is optimal to set $t_2 = t^*$.

It is then natural to ask: is full transparency ever optimal?

Proposition 5. *Holding fixed the ratio $\frac{r}{\lambda\rho}$, there exists $0 \leq \rho_1 \leq \rho_2$ (finite) such that it is optimal to set*

- *for $\rho \in [0, \rho_1]$, $0 < t_1 = t_2$;*

- *for $\rho \in [\rho_1, \rho_2]$, $0 < t_1 < t_2$;*

- *for $\rho > \rho_2$, $0 = t_1 < t_2$.*

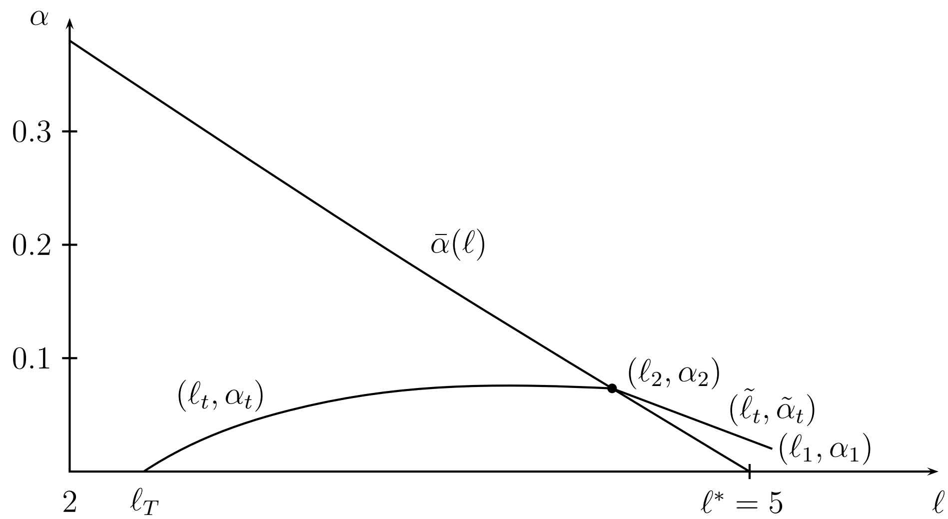

Figure 4: The probability of a recommendation as a function of ℓ (here, $(r, \lambda, \rho, k, \ell_0) = (1/2, 2, 1/5, 15, 9)$).

Hence, for ρ sufficiently large, it is optimal for the designer to use full transparency even in the absence of incentive constraints. Naturally, this implies that the same holds once such constraints are added.

Next, we ask, does it ever pay the incentive-constrained designer to use another policy?

PROPOSITION 6. *The second-best policy is different from full transparency if*

$$\rho \leq \frac{1}{\lambda}\left(\frac{r\ell_0 + \sqrt{r\ell_0}\sqrt{4k\lambda + \ell_0}}{2k} - r\right).$$

Note that the right-hand side is always positive if k is sufficiently close to ℓ_0. On the other hand, the right-hand side is negative for k large enough. While this condition is not tight, this comparative static is intuitive. If ρ or k is large enough, full transparency is very attractive: If ρ is large, "free" background learning occurs fast, so that there is no point in having agents experiment; if k is large, taking the chance of having agent make the wrong choice by recommending them to buy despite persistent uncertainty is very costly. To summarize: For some parameters, the designer's best choice consists in full transparency (when the cost is high, for instance, or when learning for "free" via ρ is scarce). Indeed, this would also be the first-best in some cases. For other parameters, she can do better than full transparency.

What is the structure of the optimal policy in such cases? This is illustrated in Figure 4. There is an initial phase in which the designer deters experimentation by recommending the good with a probability that is sufficiently high. In fact, given that agents would possibly be willing to wait initially even under full transparency, the designer might be able to do just that at the very beginning.

At some time t, all customers that have been waiting "check in" and are told to buy with some probability –unless of course a breakthrough has obtained by then. It is optimal for them to all check in at that very time, yet it is also optimal to experiment with them sequentially, increasing continuously the probability with which they are recommended to buy as the belief ℓ decreases, in a way that leaves them indifferent across this menu of (ℓ, α) offered sequentially at that instant.[17] The atom of customers is "split" at that instant, and the locus of (ℓ, α) that is visited is represented by the locus $(\tilde{\ell}, \tilde{\alpha})$, which starts at a belief ℓ_1 that is the result of the background experimentation only, and ends at ℓ_2 at which point all agents that have been waiting have checked in.

The locus $\bar{\alpha}(\ell)$ represents the locus of possible values (ℓ_2, α_2) that are candidate values for the beginning of the continuation path once no agents are left in the queue;[18] in particular, full transparency is the special case in which $\alpha_2 = 0$ (in which case $\alpha_1 = 0$ as well). From that point on, experimentation tapers off continuously, with agents buying with probability $\alpha(\ell)$ as time goes on (see locus $(\ell, \alpha(\ell))$ in Figure 4). Eventually, at some time T and belief ℓ_T, the probability $\alpha(\ell)$ that keeps them indifferent hits zero –full transparency prevails from this point on.

The condition of Proposition 5 is not tight, but an exact characterization of the parameters under which full transparency is optimal appears elusive.

5 Unobserved Costs

An important feature that has been missing so far from the analysis is private information regarding preferences, that is, in the opportunity cost of consuming the good. Section 3.4 made the strong assumption that the designer can condition on the agents' preferences. Perhaps this is an accurate description for some markets, such as Netflix, where the designer has already inferred substantial information about the agents' preferences from past choices. But it is also easy to think of cases where this is not so. Suppose that, conditional on the posterior being 1, the principal recommends to buy with probability γ_t (it is no longer obvious here that it is equal to 1). Then, conditional on hearing "buy," an agent buys if and

[17] Doing so sequentially is optimal because it increases the overall experimentation that can be performed with them; we can think of this outcome as the supremum over policies in which, at each time t, the designer can only provide one recommendation, although formally this is a well-defined maximum if the optimal control problem is written with ℓ as the "time" index.

[18] This locus of "candidate" values solves a simple optimization program, see the authors for detail. It is downward sloping and starts at the full transparency optimum, *i.e.*, $\bar{\alpha}(\ell^*) = 0$.

only if

$$c \leq \frac{\frac{p_0-p_t}{1-p_t}\gamma_t \cdot 1 + \frac{1-p_0}{1-p_t}p_t}{\frac{p_0-p_t}{1-p_t}\gamma_t + \frac{1-p_0}{1-p_t}},$$

or equivalently

$$\gamma_t \geq \frac{1-p_0}{1-c}\frac{c-p_t}{p_0-p_t}.$$

In particular, if $c < p_t$, he always buys, while if $c \geq p_0$, he never does. There is another type of possible randomization: when the posterior is 1, he recommends buying, while he recommends buying with probability α_t (as before) when the posterior is $p_t < p_0$. Then types for whom

$$\alpha_t \leq \frac{1-c}{1-p_0}\frac{p_0-p_t}{c-p_t}$$

buy, while others do not. Note that, defining $\gamma_t := 1/\alpha_t$ for such a strategy, the condition is the same as above, but of course now $\gamma_t \geq 1$.

Or the principal might do both types of randomizations at once: conditional on the posterior is 1, the principal recommends B ("buy") with probability γ, while he recommends B with probability α conditional on the posterior being p_t. We let N denote the set of cost types that buy even if recommended N ("not buy" which we take wlog to be the message that induces the lower posterior), and B the set of types that buy if recommended B.

A consumer with cost type γ_j buys after a B recommendation if and only if

$$\frac{\alpha_t}{\gamma_t} \leq \frac{\ell_0-\ell_t}{\gamma_j-\ell_t},$$

or equivalently, if

$$\gamma_j \leq \ell_t + \frac{\gamma_t}{\alpha_t}(\ell_0-\ell_t).$$

On the other hand, a N recommendation leads to buy if

$$\gamma_t \leq \alpha_t + \frac{(1-\alpha_t)(\ell_0-\gamma_j)}{\ell_0-\ell_t},$$

or equivalently

$$\gamma_j \leq \ell_0 - \frac{\gamma_t-\alpha_t}{1-\alpha_t}(\ell_0-\ell_t)$$

To gain some intuition, let us start with finitely many costs. Let us define Q^B, C^B the quantity and cost of experimentation for those who *only* buy if the recommendation is B, and we write Q^N, C^N for the corresponding variables for those who buy in any event.

We can no longer define $V(\ell)$ to be the payoff conditional on the posterior being ℓ: after all, what happens when the posterior is 1 matters for payoffs. So hereafter $V(\ell)$ refers to the expected payoff when the low posterior (*if* it is low) is ℓ. We have that

$$\begin{aligned} rV(\ell) &= \frac{1-p_0}{1-p}\big(\alpha(pQ^B - C^B) + (pQ^N - C^N)\big) \\ &+ \frac{p_0-p}{1-p}\big(\gamma(Q^B - C^B) + (Q^N - C^N)\big) - \ell(\alpha Q^B + Q^N)V'(\ell), \end{aligned}$$

where we have mixed p and ℓ. Rewriting, this gives

$$rV(\ell) = \frac{\ell_0}{1+\ell_0}(Q^N + \alpha Q^B) - (C^N + \alpha C^B) + \frac{\ell_0 - \ell}{1+\ell_0}(\gamma - \alpha)(Q^B - C^B) - \ell(Q^N + \alpha Q^B)V'(\ell).$$

Assuming –as always– that there is some free learning, we have that $rV(\ell) \to \ell(1-c)/(1+\ell)$, where c is the average cost.

It is more convenient to work with the thresholds. The principal chooses two thresholds: the lower one, γ_L is such that all cost types below buy independently of the recommendation. The higher one, γ_H, is such that types strictly above γ_L but no larger than γ_H buy only if there is a good recommendation. Solving, we get

$$\gamma_L = \ell + \frac{1-\gamma}{1-\alpha}(\ell_0 - \ell), \quad \gamma_H = \ell + \frac{\gamma}{\alpha}(\ell_0 - \ell),$$

and not surprisingly $\gamma_H \geq \gamma_L$ if and only if $\beta \geq \alpha$. In terms of γ_H, γ_L, the problem becomes

$$\begin{aligned} rV(\ell) &= \frac{\ell_0}{1+\ell_0}\left(\sum_{\gamma_j \leq \gamma_L} q_j + \frac{\ell_0 - \gamma_L}{\gamma_H - \gamma_L}\sum_{\gamma_L < \gamma_j \leq \gamma_H} q_j\right) - \left(\sum_{\gamma_j \leq \gamma_L} q_j c_j + \frac{\ell_0 - \gamma_L}{\gamma_H - \gamma_L}\sum_{\gamma_L < \gamma_j \leq \gamma_H} q_j c_j\right) \\ &+ \frac{\gamma_H - \ell_0}{\gamma_H - \gamma_L}\frac{\ell_0 - \gamma_L}{1+\ell_0}\sum_{\gamma_L < \gamma_j \leq \gamma_H} q_j(1-c_j) - \ell\left(\sum_{\gamma_j \leq \gamma_L} q_j + \frac{\ell_0 - \gamma_L}{\gamma_H - \gamma_L}\sum_{\gamma_L < \gamma_j \leq \gamma_H} q_j\right)V'(\ell), \end{aligned}$$

or rather, it is the maximum of the right-hand side subject to the constraints on γ_L, γ_H. While the interpretation of these formulas is straightforward, they are not so convenient, so we now turn to the case of uniformly distributed costs, with distribution $[0, \bar{c}]$. Note that our derivation so far applies just as well, replacing sums with integrals. We drop $\bar{c}$ from the equations that follows, reasoning per consumer. Computing these quantities and costs by integration, we obtain upon simplification

$$rV(\ell) = \max_{\gamma_L, \gamma_H}\left\{\frac{\ell_0}{2(1+\ell_0)} - \frac{\ell_0 + \gamma_L\gamma_H}{2(1+\gamma_L)(1+\gamma_H)(1+\ell_0)} - \ell\frac{\ell_0 + \gamma_L\gamma_H}{(1+\gamma_H)(1+\gamma_L)}V'(\ell)\right\},$$

or, defining

$$x := \frac{\ell_0 + \gamma_L \gamma_H}{(1+\gamma_H)(1+\gamma_L)},$$

we have that

$$rV(\ell) = \max_x \left\{ \frac{\ell_0 - x}{2(1+\ell_0)} - \ell x V'(\ell) \right\}.$$

Note that x is increasing in α and decreasing in γ,[19] and so it is maximum when $\alpha = \gamma$, in which case it is simply $\frac{\ell_0}{1+\ell_0}$, and minimum when $\gamma = 1, \alpha = 0$, in which case it is $\frac{\ell}{1+\ell}$. Note also that V must be decreasing in ℓ, as it is the expected value, not the conditional value, so that a higher ℓ means more uncertainty (formally, this follows from the principle of optimality, and the fact that a strategy available at a lower ℓ is also available at a higher ℓ). Because the right-hand side is linear in x, the solution is extremal, unless $rV(\ell) = \frac{\ell_0}{2(1+\ell_0)}$, but then $V' = 0$ and so x cannot be interior (over an interval of time) after all. We have two cases to consider, $x = \frac{\ell_0}{1+\ell_0}, \frac{\ell}{1+\ell}$.

If $x = \frac{\ell_0}{1+\ell_0}$ ("Pooling"), we immediately obtain

$$V(\ell) = \frac{\ell_0^2}{2r(1+\ell_0)^2} + C_1 \ell^{-r\left(1+\frac{1}{\ell_0}\right)}.$$

In that case, the derivative with respect to x of the right-hand side must be positive, which gives us as condition

$$\ell^{r\frac{1+\ell_0}{\ell_0}} \leq \frac{2r(1+\ell_0)^2 C_1}{\ell_0}.$$

The left-hand side being increasing, this condition is satisfied for a lower interval of values of ℓ: Therefore, if the policy $\alpha = \gamma$ is optimal at some t, it is optimal for all later times (lower values of ℓ). In that case, however, we must have $C_1 = 0$, as V must be bounded. The payoff would then be constant and equal to

$$V^P = \frac{\ell_0^2}{2r(1+\ell_0)^2}.$$

Such complete pooling is never optimal, as we shall see. Indeed, if no information is ever revealed, the payoff is $\frac{1}{r}\int_0^{p_0}(p_0 - x)\mathrm{d}x$, which is precisely equal to V^P.

If $x = \frac{\ell}{1+\ell}$, the solution is a little more unusual, and involves the exponential integral function $E_n(z) := \int_1^\infty \frac{e^{-zt}}{t^n}\mathrm{d}t$. Namely, we have

$$V^S(\ell; C_2) = \frac{\ell_0}{2r(1+\ell_0)} + e^{\frac{r}{\ell}} \frac{C_2 \ell^{-r} - E_{1+r}\left(\frac{r}{\ell}\right)}{2(1+\ell_0)}.$$

[19] Although it is intuitive, this is not meant to be immediate, but it follows upon differentiation of the formula for x.

In that case, the derivative with respect to x of the right-hand side must be negative, which gives us as condition

$$C_2 \leq \ell^{r-1} \left(\frac{e^{-\frac{r}{\ell}}}{\frac{r}{\ell}} - E_r\left(\frac{r}{\ell}\right) \right).$$

Note that the right hand side is also positive. Considering the value function, it holds that

$$\lim_{\ell \to 0} e^{\frac{r}{\ell}} \ell^{-r} = \infty, \quad \lim_{\ell \to \infty} e^{\frac{r}{\ell}} E_{1+r}\left(\frac{r}{\ell}\right) = 0.$$

The first term being the coefficient on C_2, it follows that *if* this solution is valid for values of ℓ that extend to 0, we must have $C_2 = 0$, in which case the condition is satisfied for all values of l. So one candidate is $x = \ell/(1+\ell)$ for all ℓ, with value

$$V^S(\ell) = \frac{\ell_0 - re^{\frac{r}{\ell}} E_{1+r}\left(\frac{r}{\ell}\right)}{2r(1+\ell_0)},$$

a decreasing function of ℓ, as expected. The limit makes sense: as $\ell \to 0$, $V(\ell) \to \frac{\ell_0}{2r(1+\ell_0)}$: by that time, the state will be revealed, and with this policy $(\alpha, \gamma) = (0, 1)$, the payoff will be either 0 (if the state is bad) or 1/2 (one minus the average cost) if the state is good, an event whose prior probability is $p_0 = \ell_0/(1+\ell_0)$.

We note that

$$V^S(\ell_0) - V^P = \frac{\ell_0 - re^{\frac{r}{\ell_0}} E_{1+r}\left(\frac{r}{\ell_0}\right)}{2r(1+\ell_0)} - \frac{\ell_0^2}{2r(1+\ell_0)^2} = \frac{\frac{\ell_0}{1+\ell_0} - re^{\frac{r}{\ell_0}} E_{1+r}\left(\frac{r}{\ell_0}\right)}{2r(1+\ell_0)}.$$

The function $r \mapsto re^{\frac{r}{\ell_0}} E_{1+r}\left(\frac{r}{\ell_0}\right)$ is increasing in r, with range $[0, \frac{\ell_0}{1+\ell_0}]$. Hence this difference is always positive.

To summarize: if the policy $x = \ell_0/(1+\ell_0)$ is ever taken, it is taken at all later times, but it cannot be taken throughout, because taking $x = \ell/(1+\ell)$ throughout yields a higher payoff. The last possibility is an initial phase of $x = \ell/(1+\ell)$, followed by a switch at some time $x = \ell_0/(1+\ell_0)$, with continuation value V^P. To put it differently, we would need $V = V^P$ on $[0, \bar{\ell}]$ for some $\bar{\ell}$, and $V(\ell) = V^S(\ell; C_2)$ on $[\bar{\ell}, \ell_0]$, for some C_2. But then it had better be the case that $V^S(\ell; C_2) \geq V^P$ on that higher interval, and so there would exist $0 < \ell < \ell' < \ell_0$ with $V^P = V(\ell) \leq V(\ell') = V^S(\ell'; C_2)$, so that V must be increasing over some interval of ℓ, which violates V being decreasing.

To conclude: if costs are not observed, and the principal uses a deterministic policy in terms of the amount of experimentation, he can do no better than to disclose the posterior honestly, at all times. This does not mean that he is useless, as he makes the information public, but he does not induce any more experimentation than the myopic quantity.

Proposition 7. *With uniformly distributed costs that are agents' private information, full transparency is optimal.*

To be clear, we allow the designer to randomize over *paths* of recommendation, "flipping a coin" at time 0, unbeknownst to the agents, and deciding on a particular function $(\alpha_t)_t$ as a function of the outcome of this randomization. Such "chattering" controls can sometimes be useful, and one might wonder whether introducing these would overturn this last proposition. In appendix, we show that this is not the case.

References

[1] Aumann, R. J., M. Maschler and R. E. Stearns, 1995. *Repeated Games With Incomplete Information*, Cambridge: MIT Press.

[2] Banerjee, A., 1993. "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, **107**, 797–817.

[3] 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.

[4] Cesari, L., 1983. *Optimization-Theory and Applications. Problems with ordinary differential equations, Applications of Mathematics* 17, Berlin-Heidelberg-New York: Springer-Verlag.

[5] Chamley, C. and D. Gale, 1994. "Information Revelation and Strategic Delay in a Model of Investment," *Econometrica*, **62**, 1065–1085.

[6] Ely, J., A. Frankel and E. Kamenica, 2013. "Suspense and Surprise," mimeo, University of Chicago.

[7] Fleming, W.H. and H.N. Soner, 2005. *Controlled Markov Processes and Viscosity Solutions*, Heidelberg: Springer-Verlag.

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

[9] Gul, F. and R. Lundholm, 1995. "Endogenous Timing and the Clustering of Agents' Decisions," *Journal of Political Economy*, **103**, 1039–1066.

[10] Jindal, N. and B. Liu, 2008. "Opinion Spam and Analysis," *Proceedings of the 2008 International Conference on Web Search and Data Mining, ACM,* 219-230.

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

[12] Keller, G. and S. Rady, 2012. "Breakdowns," *GESY Discussion Paper,* No. 396.

[13] Keller, G., S. Rady and M. Cripps, 2005. "Strategic Experimentation with Exponential Bandits," *Econometrica*, **73**, 39–68.

[14] Kremer, I., Y. Mansour, M. Perry, 2012. "Implementing the "Wisdom of the Crowd,"" mimeo, Hebrew University.

[15] Luca, M. and G. Zervas, 2013. "Fake It Till You Make It: Reputation, Competition and Yelp Review Fraud," mimeo, Harvard Business School.

[16] Ostrovsky, M. and M. Schwarz, 2010. "Information Disclosure and Unraveling in Matching Markets," *American Economic Journal: Microeconomics,* 2, 34-63.

[17] Rayo, L. and I. Segal, 2010. "Optimal Information Disclosure," *Journal of Political Economy* 118, 949-987.

[18] Seierstad, A. and K. Sydsæter, 1987. *Optimal Control Theory with Economic Applications.* North-Holland.

6 Appendix

Proof of Lemma 2. Let $\kappa_t := p_0/(p_0 - g_t)$. Note that $\kappa_t = 1$. Then, it follows from (2) and (4) that

$$\dot{\kappa}_t = \lambda_g \kappa_t \mu_t, \quad \kappa_0 = 1. \tag{11}$$

Dividing both sides of (11) by the respective sides of (6), we get,

$$\frac{\dot{\kappa}_t}{\dot{\ell}_t} = -\frac{\kappa_t}{\ell_t \Delta},$$

or

$$\frac{\dot{\kappa}_t}{\kappa_t} = -\frac{1}{\Delta}\frac{\dot{\ell}_t}{\ell_t}.$$

It follows that, given the initial condition,

$$\kappa_t = \left(\frac{\ell_t}{\ell_0}\right)^{-\frac{1}{\Delta}}.$$

We can then unpack κ_t to recover g_t, and from this we can obtain b_t via (4). $\square$

Proof of Lemma 3. We shall focus on

$$\hat{\alpha}(\ell) := \frac{\left(\frac{\ell}{\ell_0}\right)^{-\frac{1}{\Delta}} - 1}{k - \ell}\ell.$$

Recall $\bar{\alpha}(\ell) = \min\{1, \hat{\alpha}(\ell)\}$. Since ℓ_t falls over t when $\Delta > 0$ and rises over t when $\Delta < 0$. It suffices to show that $\hat{\alpha}(\cdot)$ is decreasing when $\Delta > 0$ and increasing when $\Delta < 0$.

We make several preliminary observations. First, $\hat{\alpha}(\ell) \in 0, 1)$ if and only if

$$1 - (\ell/\ell_0)^{\frac{1}{\Delta}} \geq 0 \text{ and } k\ell^{\frac{1}{\Delta}-1}\ell_0^{-\frac{1}{\Delta}} \geq 1. \tag{12}$$

Second,

$$\hat{\alpha}'(\ell) = \frac{(\ell_0/\ell)^{\frac{1}{\Delta}} h(\ell, k)}{\Delta(k - \ell)^2}, \tag{13}$$

where

$$h(\ell, k) := \ell - k(1 - \Delta) - k(\ell/\ell_0)^{\frac{1}{\Delta}}.$$

Third, (12) implies that

$$\frac{\mathrm{d}h(\ell, k)}{\mathrm{d}\ell} = 1 - k\ell^{\frac{1}{\Delta}-1}\ell_0^{-\frac{1}{\Delta}} \leq 0, \tag{14}$$

on any range of ℓ over which $\bar{\alpha} \leq 1$.

Consider first $\Delta < 0$. Given $k > \ell_0$, $\hat{\alpha}(\ell_0) = 0$. Then, (14) implies that, if $\hat{\alpha}'(\ell) \geq (>)0$, then $\hat{\alpha}'(\ell') \geq (>)0$ for all $\ell' \in (\ell, k)$. Since $h(\ell_0, k] < 0$, it follows that $\hat{\alpha}'(\ell) > 0$ for all $\ell \in [\ell_0, k]$. We thus conclude that $\bar{\alpha}(\ell)$ is strictly increasing on $\ell \in [\ell_0, \ell_2]$ and $\bar{\alpha}(\ell) = 1$ for all $\ell \in [\ell_2, k]$, for some $\ell_2 \in (\ell_0, k)$.

Consider next $\Delta > 0$. In this case, the relevant interval is $\ell \in [0, \ell_0]$. It follows from (14) that if $\hat{\alpha}'(\ell) \leq (<)0$, then $\hat{\alpha}'(\ell') \leq (<)0$ for all $\ell' \in (\ell, \ell_0]$. Since $h(0, k) < 0$,[20] $\hat{\alpha}'(0) < 0$, so $\hat{\alpha}(\ell)$ is decreasing in ℓ for all $\ell \in [0, \ell_0]$. It follows that $\bar{\alpha} = 1$ is on some interval $[0, \ell_1]$ and positive and decreasing (< 1) over $(\ell_1, \ell_0]$, for some $\ell_1 \in (0, \ell_0)$. $\square$

[20] Recall $\Delta \leq 1$. If $\Delta = 1$, then $h(0, k) = 0$, but $h(\epsilon, k)$, for sufficient small $\epsilon > 0$, so the argument follows.

Proof of Proposition 1. To analyze this tradeoff precisely, we reformulate the designer's problem to conform to the standard optimal control problem framework. First, we switch the roles of variables so that we treat ℓ as a "time" variable and $t(\ell) := \inf\{t|\ell_t \le \ell\}$ as the state variable, interpreted as the time it takes for a posterior ℓ to be reached. Up the constant (additive and multiplicative) terms, the designer's problem is written as: For problem $i = SB, FB$,

$$\sup_{\alpha(\ell)} \int_0^{\ell^0} e^{-rt(\ell)} \ell^{\frac{1}{\Delta}-1} \left(1 - \frac{k}{\ell} - \frac{\rho\left(1-\frac{k}{\ell}\right)+1}{\rho+\alpha(\ell)}\right) \mathrm{d}\ell.$$

s.t. $t(\ell^0) = 0$,

$$t'(\ell) = -\frac{1}{\Delta\lambda_g(\rho+\alpha(\ell))\ell},$$

$$\alpha(\ell) \in \mathcal{A}^i(\ell),$$

where $\mathcal{A}^{SB}(\ell) := [0, \bar{\alpha}(\ell)]$, and $\mathcal{A}^{FB} := [0,1]$.

This transformation enables us to focus on the optimal recommendation policy directly as a function of the posterior ℓ. Given the transformation, the admissible set no longer depends on the state variable (since ℓ is no longer a state variable), thus conforming to the standard specification of the optimal control problem.

Next, we focus on $u(\ell) := \frac{1}{\rho+\alpha(\ell)}$ as the control variable. With this change of variable, the designer's problem (both second-best and first-best) is restated, up to constant (additive and multiplicative) terms: For $i = SB, FB$,

$$\sup_{u(\ell)} \int_0^{\ell^0} e^{-rt(\ell)} \ell^{\frac{1}{\Delta}-1} \left(1 - \frac{k}{\ell} - \left(\rho\left(1-\frac{k}{\ell}\right)+1\right)u(\ell)\right) \mathrm{d}\ell, \tag{15}$$

s.t. $t(\ell^0) = 0$,

$$t'(\ell) = -\frac{u(\ell)}{\Delta\lambda_g \ell},$$

$$u(\ell) \in \mathcal{U}^i(\ell),$$

where the admissible set for the control is $\mathcal{U}^{SB}(\ell) := [\frac{1}{\rho+\alpha(\ell)}, \frac{1}{\rho}]$ for the second-best problem and $\mathcal{U}^{FB}(\ell) := [\frac{1}{\rho+1}, \frac{1}{\rho}]$. With this transformation, the problem becomes a standard linear optimal control problem (with state t and control α). A solution exists by the Filippov-Cesari theorem (Cesari, 1983).

We shall thus focus on the necessary condition for optimality to characterize the optimal recommendation policy. To this end, we write the Hamiltonian:

$$\mathcal{H}(t,u,\ell,\nu) = e^{-rt(\ell)}\ell^{\frac{1}{\Delta}-1}\left(1-\frac{k}{\ell}-\left(\rho\left(1-\frac{k}{\ell}\right)+1\right)u(\ell)\right)-\nu\frac{u(\ell)}{\Delta\lambda_g\ell}.$$

The necessary optimality conditions state that there exists an absolutely continuous function $\nu : [0,\ell^0]$ such that, for all ℓ, either

$$\phi(\ell) := \Delta\lambda_g e^{-rt(\ell)}\ell^{\frac{1}{\Delta}}\left(\rho\left(1-\frac{k}{\ell}\right)+1\right)+\nu(\ell)=0, \tag{16}$$

or else $u(\ell)=\frac{1}{\rho+\alpha(\ell)}$ if $\phi(\ell)>0$ and $u(\ell)=\frac{1}{\rho}$ if $\phi(\ell)<0$.

Furthermore,

$$\nu'(\ell) = -\frac{\partial\mathcal{H}(t,u,\ell,\nu)}{\partial t} = re^{-rt(\ell)}\ell^{\frac{1}{\Delta}-1}\left(\left(1-\frac{k}{\ell}\right)(1-\rho u(\ell))-u(\ell)\right)\ (\ell-\text{a.e.}). \tag{17}$$

Finally, transversality at $\ell=0$ ($t(\ell)$ is free) implies that $\nu(0)=0$.

Note that

$$\begin{aligned}\phi'(\ell) &= -rt'(\ell)\Delta\lambda_g e^{-rt(\ell)}\ell^{\frac{1}{\Delta}}\left(\rho\left(1-\frac{k}{\ell}\right)+1\right)\\ &+ \lambda_g e^{-rt(\ell)}\ell^{\frac{1}{\Delta}-1}\left(\rho\left(1-\frac{k}{\ell}\right)+1\right)+\rho k\Delta\lambda_g e^{-rt(\ell)}\ell^{\frac{1}{\Delta}-2}+\nu'(\ell),\end{aligned}$$

or using the formulas for t' and ν',

$$\phi'(\ell) = e^{-rt(\ell)}\ell^{\frac{1}{\Delta}-2}\left(r(\ell-k)+\rho\Delta\lambda_g k+\lambda_g(\rho(\ell-k)+\ell)\right), \tag{18}$$

so ϕ cannot be identically zero over some interval, as there is at most one value of ℓ for which $\phi'(\ell)=0$. Every solution must be "bang-bang." Specifically, $\phi'(\ell)>0$ is equivalent to

$$\phi'(\ell) \gtreqless 0 \Leftrightarrow \ell \gtreqless \tilde{\ell} := \left(1-\frac{\lambda_g(1+\rho\Delta)}{r+\lambda_g(1+\rho)}\right)k>0.$$

Also, $\phi(0)\le 0$ (specifically, $\phi(0)=0$ for $\Delta<1$ and $\phi(0)=-\Delta\lambda_g e^{-rt(\ell)}\rho k$ for $\Delta=1$). So $\phi(\ell)<0$ for all $0<\ell<\ell_g^*$, for some threshold $\ell_g^*>0$, and $\phi(\ell)>0$ for $\ell>\ell_g^*$. The constraint $u(\ell)\in\mathcal{U}^i(\ell)$ must bind for all $\ell\in[0,\ell^*)$ (a.e.), and every optimal policy must switch from $u(\ell)=1/\rho$ for $\ell<\ell_g^*$ to $1/(\rho+\alpha(\ell))$ in the second-best problem and to $1/(\rho+1)$ in the first-best problem for $\ell>\ell_g^*$. It remains to determine the switching point ℓ^* (and establish uniqueness in the process).

For $\ell < \ell^*$,

$$\nu'(\ell) = -\frac{r}{\rho} e^{-rt(\ell)} \ell^{\frac{1}{\Delta}-1}, \quad t'(\ell) = -\frac{1}{\rho \Delta \lambda_g \ell}$$

so that

$$t(\ell) = C_0 - \frac{1}{\rho \Delta \lambda_g} \ln \ell, \text{ or } e^{-rt(\ell)} = C_1 \ell^{\frac{r}{\rho \Delta \lambda_g}}$$

for some constants $C_1, C_0 = -\frac{1}{r} \ln C_1$. Note that $C_1 > 0$ as otherwise $t(\ell) = \infty$ for $\ell \in (0, \ell_g^*)$, which is inconsistent with $t(\ell_g^*) < \infty$. Hence,

$$\nu'(\ell) = -\frac{r}{\rho} C_1 \ell^{\frac{r}{\rho \Delta \lambda_g} + \frac{1}{\Delta} - 1},$$

and so (using $\nu(0) = 0$),

$$\nu(\ell) = -\frac{r \Delta \lambda_g}{r + \rho \lambda_g} C_1 \ell^{\frac{r}{\rho \Delta \lambda_g} + \frac{1}{\Delta}},$$

for $\ell < \ell_g^*$. We now substitute ν into ϕ, for $\ell < \ell_g^*$, to obtain

$$\phi(\ell) = \Delta \lambda_g C_1 \ell^{\frac{r}{\rho \Delta \lambda_g}} \ell^{\frac{1}{\Delta}} \left(\rho \left(1 - \frac{k}{\ell} \right) + 1 \right) - \frac{r \Delta \lambda_g}{r + \rho \lambda_g} C_1 \ell^{\frac{r}{\rho \Delta \lambda_g} + \frac{1}{\Delta}}.$$

We now see that the switching point is uniquely determined by $\phi(\ell) = 0$, as ϕ is continuous and C_1 cancels. Simplifying,

$$\frac{k}{\ell_g^*} = 1 + \frac{\lambda_g}{r + \rho \lambda_g},$$

which leads to the formula for p_g^* in the Proposition (via $\ell = p/(1-p)$ and $k = c/(1-c)$). We have identified the unique solution to the program for both first- and second-best, and shown in the process that the optimal threshold p^* applies to both problems.

The second-best implements the first-best if $p_0 \geq c$, since then $\bar{\alpha}(\ell) = 1$ for all $\ell \leq \ell_0$. If not, then $\bar{\alpha}(\ell) < 1$ for a positive measure of $\ell \leq \ell_0$. Hence, the second-best implements a lower and thus a slower experimentation than does the first-best.

As for sufficiency, we use Arrow sufficiency theorem (Seierstad and Sydsæter, 1987, Theorem 5, p.107). Note that, for all $u \in \mathcal{U}^i$, $i = FB, SB$,

$$1 - \frac{k}{\ell} - \left(\rho \left(1 - \frac{k}{\ell} \right) + 1 \right) u(\ell) \leq 1 - \frac{k}{\ell} - \left(\rho \left(1 - \frac{k}{\ell} \right) + 1 \right) \frac{1}{1+\rho} = -\frac{k}{(1+\rho)\ell} < 0.$$

Hence, given (15), the maximized Hamiltonian $\hat{\mathcal{H}}(t, \ell, \nu(\ell)) = \max_{u \in \mathcal{U}^i(\ell)} \mathcal{H}(t, u, \ell, \nu(\ell))$ is necessarily concave in t, for all ℓ, implying the result. $\square$

Proof of Proposition 2. The same steps must be applied to the case $\Delta < 0$. The same change of variable produces the following program for the designer: For problem $i = SB, FB$,

$$\sup_{u} \int_{\ell^0}^{\infty} e^{-rt(\ell)} \ell^{\frac{1}{\Delta}-1} \left(\left(1 - \frac{k}{\ell}\right) (1 - \rho u(\ell)) - u(\ell) \right) \mathrm{d}\ell,$$

s.t. $t(\ell^0) = 0$,

$$t'(\ell) = -\frac{u(\ell)}{\Delta \lambda_g \ell},$$

$$u(\ell) \in \mathcal{U}^i(\ell),$$

where as before $\mathcal{U}^{SB}(\ell) := [\frac{1}{\rho+\alpha(\ell)}, \frac{1}{\rho}]$ and $\mathcal{U}^{FB}(\ell) := [\frac{1}{\rho+1}, \frac{1}{\rho}]$. We pause and note that, for all $u(\ell) \in \mathcal{U}^i(\ell)$, $i = FB, SB$,

$$\left(1 - \frac{k}{\ell}\right) (1 - \rho u(\ell)) - u(\ell) \leq \left(1 - \frac{k}{\ell}\right) \left(1 - \frac{1}{\rho + 1}\right) - \frac{1}{\rho + 1} = -\frac{k}{\ell} \frac{1}{\rho + 1} < 0,$$

so that, as in the case $\Delta < 0$, the maximized Hamiltonian will necessarily be concave in t, which will imply optimality of our candidate solution, by Arrow's sufficiency theorem.

We now turn to the necessary conditions. As before, the necessary conditions for the second-best policy now state that there exists an absolutely continuous function $\nu : [0, \ell^0]$ such that, for all ℓ, either

$$\psi(\ell) := -\phi(\ell) = \Delta \lambda_g e^{-rt(\ell)} \ell^{\frac{1}{\Delta}} \left(\rho \left(1 - \frac{k}{\ell}\right) + 1 \right) - \nu(\ell) = 0, \tag{19}$$

or else $u(\ell) = \frac{1}{\rho+\alpha(\ell)}$ if $\psi(\ell) > 0$ and $u(\ell) = \frac{1}{\rho}$ if $\psi(\ell) < 0$. The formula for $\nu'(\ell)$ is the same as before, given by (17). Finally, transversality at $\ell = \infty$ ($t(\ell)$ is free) implies that $\lim_{\ell \to \infty} \nu(\ell) = 0$.

Since $\psi(\ell) = -\phi(\ell)$, we get from (19) that

$$\psi'(\ell) = -e^{-rt(\ell)} \ell^{\frac{1}{\Delta}-2} \left(r\left(\ell - k\right) + \rho \Delta \lambda_g k + \lambda_g \left(\rho \left(\ell - k\right) + \ell \right) \right).$$

Letting $\tilde{\ell} := \left(1 - \frac{\lambda_g(1+\rho\Delta)}{r+\lambda_g(1+\rho)}\right) k$, namely the solution to $\psi(\ell) = 0$. Then, ψ is maximized at $\tilde{\ell}$, and is strictly quasi-concave. Since $\lim_{\ell \to \infty} h(\ell) = 0$, this means that there must be a cutoff $\ell_b^* < \tilde{\ell}$ such that $\psi(\ell) < 0$ for $\ell < \ell_b^*$ and $\psi(\ell) > 0$ for $\ell > \ell_b^*$. Hence, the solution is bang-bang, with $u(\ell) = 1/\rho$ if $\ell < \ell_b^*$, and $u(\ell) = 1/(\rho + \alpha(\ell))$ if $\ell > \ell_b^*$.

The first-best policy has the same cutoff structure, except that the cutoff may be different from ℓ_b^*. Let ℓ_b^{**} denote the first-best cutoff.

First-best policy: We shall first consider the first best policy. In that case, for $\ell > \ell_b^{**}$,

$$t'(\ell) = -\frac{1}{\Delta\lambda_g(1+\rho)\ell}$$

gives

$$e^{-rt(\ell)} = C_2 \ell^{\frac{r}{(1+\rho)\Delta\lambda_g}},$$

for some non-zero constant C_2. Then

$$\nu'(\ell) = -\frac{rk}{1+\rho} C_2 \ell^{\frac{r}{(1+\rho)\Delta\lambda_g}+\frac{1}{\Delta}-2}$$

and $\lim_{\ell\to\infty} \nu(\ell) = 0$ give

$$\nu(\ell) = -\frac{rk\Delta\lambda_g}{r+(1+\rho)(1-\Delta)\lambda_g} C_2 \ell^{\frac{r}{(1+\rho)\Delta\lambda_g}+\frac{1}{\Delta}-1}.$$

So we get, for $\ell > \ell_b^{**}$,

$$\psi(\ell) = -\Delta\lambda_g C_2 \ell^{\frac{r}{(1+\rho)\Delta\lambda_g}} \ell^{\frac{1}{\Delta}-1} \left(\ell(1+\rho) - k\rho\right) + \frac{rk\Delta\lambda_g}{r+(1+\rho)(1-\Delta)\lambda_g} C_2 \ell^{\frac{r}{(1+\rho)\Delta\lambda_g}+\frac{1}{\Delta}-1}.$$

Setting $\psi(\ell_b^{**}) = 0$ gives

$$\frac{k}{\ell_b^{**}} = \frac{r+(1+\rho)(1-\Delta)\lambda_g}{r+\rho(1-\Delta)\lambda_g} = \frac{r+(1+\rho)\lambda_b}{r+\rho\lambda_b} = 1 + \frac{\lambda_b}{r+\rho\lambda_b},$$

or

$$p_b^{**} = c\left(1 - \frac{rv}{\rho + r(v + \frac{1}{(1-\Delta)\lambda_g})}\right) = c\left(1 - \frac{rv}{\rho + r(v+\frac{1}{\lambda_b})}\right).$$

Second-best policy. We now characterize the second-best cutoff. There are two cases, depending upon whether $\alpha(\ell) = 1$ is incentive-feasible at the threshold ℓ_b^{**} that characterizes the first-best policy. In other words, for the first-best to be implementable, we should have $\bar{\alpha}(\ell^{**}) = 1$, which requires

$$\ell_0 \geq k\left(\frac{r+\rho\lambda_b}{r+(1+\rho)\lambda_b}\right)^{1-\Delta} =: \hat{\ell}_0.$$

Observe that since $\Delta < 0$, $\hat{\ell}_0 < \ell^{**}$. If $\ell_0 \leq \hat{\ell}_0$, then the designer begins with no experimentation and waits until the posterior belief improves sufficiently to reach ℓ^{**}, at which point the agents will be asked to experiment with full force, i.e., with $\bar{\alpha}(\ell) = 1$, that is, given that no news has arrived by that time. This first-best policy is implementable since, given the sufficiently favorable prior, the designer will have built sufficient "credibility" by that time. Hence, unlike the case of $\Delta > 0$, the first best can be implementable even when $\ell_0 < k$.

Suppose $\ell_0 < \hat{\ell}_0$. Then, the first-best is not implementable. That is, $\bar{\alpha}(\ell_b^{**}) < 1$. Let ℓ_b^* denote the threshold at which the constrained designer switches to $\bar{\alpha}(\ell)$. We now prove that $\ell_b^* > \ell_b^{**}$.

For the sake of contradiction, suppose that $\ell_b^* \leq \ell_b^{**}$. Note that $\psi(x) = \lim_{\ell\to\infty} \phi(\ell) = 0$. This means that

$$\int_{\ell_b^*}^{\infty} \psi'(\ell)\mathrm{d}\ell = \int_{\ell_b^*}^{\infty} e^{-rt(\ell)}\ell^{\frac{1}{\Delta}-2}\left((r+\lambda_b\rho)k - (r+\lambda_g(\rho+1))\ell\right)\mathrm{d}\ell = 0,$$

where $\psi'(\ell) = -\phi'(\ell)$ is derived using the formula in (19).

Let t^{**} denote the time at which ℓ_b^{**} is reached along the first-best path. Let

$$f(\ell) := \ell^{\frac{1}{\Delta}-2}\left((r+\lambda_b\rho)k - (r+\lambda_g(\rho+1))\ell\right).$$

We then have

$$\int_{\ell_b^*}^{\infty} e^{-rt^{**}(\ell)} f(\ell)\mathrm{d}\ell \geq 0, \tag{20}$$

(because $\ell_b^* \leq \ell_b^{**}$; note that $f(\ell) \leq 0$ if and only if $\ell > \tilde{\ell}$, so h must tend to 0 as $\ell \to \infty$ from above), yet

$$\int_{\ell_b^*}^{\infty} e^{-rt(\ell)} f(\ell)\mathrm{d}\ell = 0. \tag{21}$$

Multiplying $e^{rt^{**}(\tilde{\ell})}$ on both sides of (20) gives

$$\int_{\ell_b^*} e^{-r(t^{**}(\ell)-t^{**}(\tilde{\ell}))} f(\ell)\mathrm{d}\ell \geq 0. \tag{22}$$

Likewise, multiplying $e^{rt(\tilde{\ell})}$ on both sides of (21) gives

$$\int_{\ell_b^*}^{\infty} e^{-r(t(\ell)-t(\tilde{\ell}))} f(\ell)\mathrm{d}\ell = 0. \tag{23}$$

Subtracting (22) from (23) gives

$$\int_{\ell_b^*} \left(e^{-r(t(\ell)-t(\tilde{\ell}))} - e^{-r(t^{**}(\ell)-t^{**}(\tilde{\ell}))}\right) f(\ell)\mathrm{d}\ell \leq 0. \tag{24}$$

Note $t'(\ell) \geq (t^{**})'(\ell) > 0$ for all ℓ, with strict inequality for a positive measure of ℓ. This means that $e^{-r(t(\ell)-t(\tilde{\ell}))} \leq e^{-r(t^{**}(\ell)-t^{**}(\tilde{\ell}))}$ if $\ell > \tilde{\ell}$, and $e^{-r(t(\ell)-t(\tilde{\ell}))} \geq e^{-r(t^{**}(\ell)-t^{**}(\tilde{\ell}))}$ if $\ell < \tilde{\ell}$, again with strict inequality for a positive measure of ℓ for $\ell \geq \ell_b^{**}$ (due to the fact that the first best is not implementable; i.e., $\bar{\alpha}(\ell_b^{**}) < 1$). Since $f(\ell) < 0$ if $\ell > \tilde{\ell}$ and $f(\ell) > 0$ if $\ell < \tilde{\ell}$, we have a contradiction to (24). $\square$

Proof of Proposition 3. In that case, $\ell = \ell_0$. The objective rewrites

$$\begin{aligned}
\mathcal{W} &= \int_{t\geq 0} e^{-rt}\left(g_t(1-c) + \frac{p_0-c}{p_0}\alpha_t(p_0-g_t)\right)\mathrm{d}t \\
&= \int_{t\geq 0} e^{-rt}\left(g_t(1-c) + \frac{p_0-c}{p_0}\left(\frac{\dot{g}_t}{\lambda_g} - (p_0-g_t)\rho\right)\right)\mathrm{d}t \\
&= \int_{t\geq 0} e^{-rt}\left(g_t(1-c) + \frac{p_0-c}{p_0}\left(r\frac{g_t}{\lambda_g} - (p_0-g_t)\rho\right)\right)\mathrm{d}t + \text{Const.} \quad \text{(Integr. by parts)} \\
&= \int_{t\geq 0} e^{-rt} g_t \left(1-c + \frac{p_0-c}{p_0}\left(\frac{r}{\lambda_g} + \rho\right)\right)\mathrm{d}t + \text{Const.} \\
&= \text{Const.} \times \int_{t\geq 0} e^{-rt} g_t \left((\ell_0-k)(r+\lambda_g\rho) + \lambda_g\ell_0\right)\mathrm{d}t + \text{Const.},
\end{aligned}$$

and so we see that it is best to set g_t to its maximum or minimum value depending on the sign of $(\ell_0-k)(r+\lambda_g\rho)+\lambda_g\ell_0$, specifically, depending on

$$\frac{k}{\ell_0} \lessgtr 1 + \frac{\lambda_g}{r+\lambda_g\rho},$$

which is the relationship that defines $\ell_g^* = \ell_b^{**}$. Now, g_t is maximized by setting $\alpha_\tau = \bar{\alpha}_\tau$ and minimized by setting $\alpha_\tau = 0$ (for all $\tau < t$).

We can solve for α from the incentive compatibility constraint, plug back into the differential equation for g_t and get, by solving the ode,

$$g_t = \frac{\left(e^{\frac{\lambda_g(\ell_0-k\rho)t}{k-\ell_0}} - 1\right)\ell_0(k-\ell_0)\rho}{(1+\ell_0)(\ell_0-k\rho)},$$

and finally

$$\alpha = \frac{\ell_0}{\frac{\rho k-\ell_0}{\rho\left(1-e^{\frac{\lambda_g(\ell_0-\rho k)t}{k-\ell_0}}\right)} - (k-\ell_0)},$$

which is increasing in t and convex in t (for $\gamma > l^0$) and equal to 1 when

$$\lambda_g t^* = \frac{k-\ell_0}{\ell_0-k\rho}\ln\frac{\ell_0}{k\rho}.$$

The optimal policy in that case is fairly obvious: experiment at maximum rate until t^*, at rate 1 from that point on (conditional on no feedback).

□

Proof of Proposition 4. The objective function reads

$$\int_{t\geq 0} e^{-rt} \left(g_t(1-\bar{c}) + (1-g_t-b_t)(q_H\alpha_H(p_t-c_L)q_L\alpha_L(p_t-c_L)\right) \mathrm{d}t,$$

where $\bar{c} := q_Hc_H + q_Lc_L$. Substituting for g_t, b_t and re-arranging, this gives

$$\int_{t\geq 0} e^{-rt}\ell(t)\left(\alpha_H(t)q_H\left(1-c_H\left(1+\frac{1}{\ell(t)}\right)\right) + \alpha_L(t)q_L\left(1-c_L\left(1+\frac{1}{\ell(t)}\right)\right) - (1-\bar{c})\right)\mathrm{d}t.$$

As before, it is more convenient to work with $t(\ell)$ as the state variable, and doing the change of variables gives

$$\int_0^{\ell_0} e^{-rt(\ell)}\left(x_H(\ell)u_H(\ell) + x_L(\ell)u_L(\ell) - \frac{1-\bar{c}}{\rho}\right)\mathrm{d}\ell,$$

where for $j = L, H$, $x_j(\ell) := 1 - c_j\left(1+\frac{1}{\ell}\right) + \frac{1-\bar{c}}{\rho}$, and $u_j(\ell) := \frac{q_j\alpha_j(t(\ell))}{\rho + q_L\alpha_L(t(\ell)) + q_H\alpha_H(t(\ell))}$ are the control variables that take values in the sets $\mathcal{U}^j(\ell) = [\underline{u}_k, \bar{u}_k]$ (whose definition depends on first- vs. second-best). This is to be maximized subject to

$$t'(\ell) = \frac{u_H(\ell) + u_L(\ell) - 1}{\rho\lambda\ell}.$$

As before, we invoke Pontryagin's principle. There exists an absolutely continuous function $\eta : [0, \ell_0] \to \mathbb{R}$, such that, a.e.,

$$\eta'(\ell) = re^{-rt(\ell)}\left(x_H(\ell)u_H(\ell) + x_L(\ell)u_L(\ell) - \frac{1-\bar{c}}{\rho}\right),$$

and u_j is maximum or minimum, depending on the sign of

$$\phi_j(\ell) := \rho\lambda\ell e^{-rt(\ell)}x_j(\ell) + \eta(\ell).$$

This is because this expression cannot be zero except for a specific value of $\ell = \ell_j$. Namely, note first that, because $x_H(\ell) < x_L(\ell)$ for all ℓ, at least one of $u_L(\ell), u_H(\ell)$ must be extremal, for all ℓ. Second, upon differentiation,

$$\phi_H'(\ell) = e^{-rt(\ell)}\left(\left(\lambda - \frac{r}{\rho}\right)(1-\bar{c}) + \rho\lambda(1-c_H) + ru_L(\ell)(c_H-c_L)\left(1+\frac{1}{\ell}\right)\right)$$

implies that, if $\phi_H(\ell) = 0$ were identically zero over some interval, then $u_L(\ell)$ would be extremal over this range, yielding a contradiction, as the right-hand side cannot be zero identically, for $u_L(\ell) = \bar{u}_L(\ell)$. Similar reasoning applies to $u_L(\ell)$, considering $\phi_L'(\ell)$. Hence, the optimal policy is characterized by two thresholds, ℓ_H, ℓ_L, with $\ell_0 \geq \ell_H \geq \ell_L \geq 0$, such that both types of regular consumers are asked to experiment whenever $\ell \in [\ell_H, \ell_0]$, low-cost consumers are asked to do so whenever $\ell \in [\ell_L, \ell_0]$, and neither is asked to otherwise. By the principle of optimality, the threshold ℓ_L must coincide with $\ell_g^* = \ell_g^{**}$ in the case of only one type of regular consumers (with cost c_L). To compare ℓ_H^* and ℓ_H^{**}, we proceed as as in the bad news case, by noting that, in either case,

$$\phi_H(\ell_H) = 0,$$

and

$$\phi_H(\ell_L) = \phi_L(\ell_L) + \rho\lambda\ell_L e^{-rt(\ell_L)}(x_H(\ell_L) - x_L(\ell_L)) = -\rho\lambda e^{-rt(\ell_L)}(c_H - c_L)(1 + \ell_L).$$

Hence,

$$\int_{\ell_L}^{\ell_H} e^{rt(\ell_L)}\phi_H'(\ell)\mathrm{d}\ell = \rho\lambda(c_H - c_L)(1 + \ell_L)$$

holds both for the first- and second-best. Note now that, in the range $[\ell_L, \ell_H]$,

$$e^{rt(\ell_L)}\phi_H'(\ell) = e^{-r\int_{\ell_H}^{\ell} \frac{u_L(l)+u_H(l)-1}{\rho\lambda l}\mathrm{d}l}\left(\left(\lambda - \frac{r}{\rho}\right)(1-\bar{c}) + \rho\lambda(1-c_H) + ru_L(\ell)(c_H - c_L)\left(1 + \frac{1}{\ell}\right)\right).$$

Because $\bar{\alpha}_L(\ell) > \bar{\alpha}_H(\ell)$, $\bar{u}_L^*(\ell) > \bar{u}_L^{**}(\ell)$, and also $\bar{u}_L^{**}(\ell) + \bar{u}_H^{**}(\ell) \geq \bar{u}_L^*(\ell) + \bar{u}_H^*(\ell)$, so that, for all ℓ in the relevant range,

$$e^{rt(\ell_L)}\frac{\mathrm{d}\phi_H^{**}(\ell)}{\mathrm{d}\ell} < e^{rt(\ell_L)}\frac{\mathrm{d}\phi_H^*(\ell)}{\mathrm{d}\ell},$$

and it then follows that $\ell_H^* < \ell_H^{**}$. □

Second-best analysis with a continuum of observable costs. We characterize the recommendation policy as $r \to 0$. To derive this policy, let us first describe the designer's payoff. This is his payoff *in expectation*. Her objective is

$$\int_0^{t_1} e^{-rt}\left[\int_0^{\frac{\ell_0}{1+\ell_0}} \frac{\ell_0 - \ell_t}{1+\ell_0}(1-c)\,\mathrm{d}c + \int_{\frac{k_t}{1+k_t}}^{\bar{c}} \frac{\ell_0 - \ell_t}{1+\ell_0}(1-c)\,\mathrm{d}c + \int_0^{\frac{\ell_0}{1+\ell_0}} \frac{1+\ell_t}{1+\ell_0}\left(\frac{\ell_t}{1+\ell_t} - c\right)\mathrm{d}c\right]\mathrm{d}t$$
$$+ \int_{t_1}^{\infty} e^{-rt}\left[\int_0^{\bar{c}} \frac{\ell_0 - \ell_t}{1+\ell_0}(1-c)\,\mathrm{d}c + \int_0^{\frac{k_t}{1+k_t}} \frac{1+\ell_t}{1+\ell_0}\left(\frac{\ell_t}{1+\ell_t} - c\right)\mathrm{d}c\right]\mathrm{d}t.$$

To understand this expression, consider $t < t_1$. Types in $t \in (\ell_0, k_t)$ derive no surplus, because they are indifferent between buying or not (what they gain from being recommended to buy when the good has turned out to be good is exactly offset by the cost of doing so when this is myopically suboptimal). Hence, their contribution to the expected payoff cancels out (but it does not mean that they are disregarded, because their behavior affects the amount of experimentation.) Types above k_t get recommended to buy only if the good has turned out to be good, in which case they get a flow surplus of $\lambda \cdot 1 - c = 1 - c$. Types below ℓ_0 have to purchase for both possible posterior beliefs, and while the flow revenue is 1 in one case, it is only $p_t = \ell_t/(1+\ell_t)$ in the other case.

The payoff in case $t \geq t_1$ can be understood similarly. There are no longer indifferent types. In case of an earlier success, all types enjoy their flow payoff $1 - c$, while in case of no success, types below γ_t still get their flow $p_t - c$.

This expression can be simplified to

$$\begin{aligned} J(k) &= \int_0^\infty e^{-rt}\left[\int_0^{\frac{\ell_0}{1+\ell_0}\wedge\frac{k_t}{1+\gamma_t}}\left(\frac{\ell_0}{1+\ell_0}-c\right)\mathrm{d}c+\int_{\frac{\gamma_t}{1+\gamma_t}}^{\frac{\bar{k}}{1+\bar{k}}}\frac{\ell_0-\ell_t}{1+\ell_0}(1-c)\,\mathrm{d}c\right]\mathrm{d}t \\ &= \int_0^\infty e^{-rt}\left[\frac{\ell_0}{1+\ell_0}\left(\frac{\ell_0}{1+\ell_0}\wedge\frac{\gamma_t}{1+\gamma_t}\right)-\frac{1}{2}\left(\frac{\ell_0}{1+\ell_0}\wedge\frac{\bar{k}_t}{1+\gamma_t}\right)^2\right]\mathrm{d}t \\ &+ \int_0^\infty e^{-rt}\left[\frac{\ell_0-\ell_t}{1+\ell_0}\left(\frac{\bar{k}}{1+\bar{k}}-\frac{k_t}{1+k_t}-\frac{1}{2}\left(\left(\frac{\bar{k}}{1+\bar{k}}\right)^2-\left(\frac{\gamma_t}{1+\gamma_t}\right)^2\right)\right)\right]\mathrm{d}t, \end{aligned}$$

with the obvious interpretation. For $t \geq t_1$,

$$\dot{\ell}_t = -\ell_t\int_0^{\frac{k_t}{1+k_t}}\frac{\mathrm{d}c}{\bar{c}} = -\frac{\ell_t}{\bar{c}}\frac{k_t}{1+k_t},$$

while for $t \leq t_1$, it holds that

$$\begin{aligned} \dot{\ell}_t &= -\ell_t\left(\frac{p_0}{\bar{c}}+\int_{p_0}^{\frac{\gamma_t}{1+\gamma_t}}\alpha_t(k)\frac{\mathrm{d}c}{\bar{c}}\right) = -\frac{\ell_t}{\bar{c}}\left(p_0+\int_{p_0}^{\frac{\gamma_t}{1+\gamma_t}}\frac{\ell_0-\ell_t}{k(c)-\ell_t}\mathrm{d}c\right) \\ &= -\frac{\ell_t}{\bar{c}}\left[\frac{k_t\ell_t+\ell_0}{(1+\gamma_t)(1+\ell_0)}-\frac{\ell_0-\ell_t}{(1+\ell_t)^2}\ln\frac{(1+k_t)(\ell_0-\ell_t)}{(1+\ell_0)(k_t-\ell_t)}\right]. \end{aligned}$$

Finally, note that the value of k_0 is free.

To solve this problem, we apply Pontryagin's maximum principle. Consider first the case $t \geq t_1$. The Hamiltonian is then

$$\mathcal{H}(\ell,\gamma,\mu,t) = \frac{e^{-rt}}{2(1+k_t)^2}\left(2k_t(1+k_t)\frac{\ell_0}{1+\ell_0}-k_t^2+\frac{(\bar{k}-k_t)(2+\gamma_t+\bar{k})(\ell_0-\ell_t)}{(1+\bar{k})^2(1+\ell_0)}\right)-\mu_t\ell_t\frac{\gamma_t(1+\bar{k})}{(1+\gamma_t)\bar{k}},$$

where μ is the co-state variable. The maximum principle gives, taking derivatives with respect to the control γ_t,

$$\mu_t = -e^{-rt}\bar{k}\frac{k_t - \ell_t}{(1+\bar{k})(1+k_t)(1+\ell_0)\ell_t}.$$

The adjoint equation states that

$$\dot{\mu} = -\frac{\partial \mathcal{H}}{\partial \ell} = \frac{e^{-rt}}{2(1+\bar{k})^2(1+\ell_0)(1+k_t)^2\ell_t}\left(k_t^2(2(1+\bar{k})^2+\ell_t) - \bar{k}(2+\bar{k})(2k_t+1)\ell_t\right),$$

after inserting the value for μ_t. Differentiate the formula for μ, combine to get a differential equation for k_t. Letting $r \to 0$, and changing variables to $k(\ell)$, we finally obtain

$$2(1+\bar{k})^2\frac{(1+\ell)\gamma(\ell)}{1+\gamma(\ell)}\gamma'(\ell) = \bar{k}(2+\bar{k})(1+2\gamma(\ell)) - \gamma(\ell)^2.$$

Along with $k(0)=0$, $k>0$ we get

$$\gamma(\ell) = \frac{\bar{k}(2+\bar{k})\ell + (1+\bar{k})\sqrt{\bar{k}(2+\bar{k})\ell(1+\ell)}}{(1+\bar{k})^2+\ell}.$$

This gives us, in particular, $\gamma(l_0)$. Note that, in terms of cost c, this gives

$$c(\ell) = \frac{\sqrt{\bar{k}(2+\bar{k})\ell/(1+\ell)}}{1+\bar{k}},$$

We now turn to the Hamiltonian for the case $t \le t_1$, or $\gamma_t \ge \ell_0$. It might be that the solution is a "corner" solution, that is, all agents experiment ($\gamma_t = \bar{k}$). Hence, we abuse notation, and solve for the unconstrained solution γ: the actual solution should be set at $\min\{\bar{k}, \gamma_t\}$. Proceeding in the same fashion, we get again

$$\mu_t = -e^{-rt}\bar{k}\frac{\gamma_t - \ell_t}{(1+\bar{k})(1+\gamma_t)(1+\ell_0)\ell_t},$$

and continuity of μ (which follows from the maximum principle) is thus equivalent to the values of $\gamma(\ell)$ obtained from both cases matching at $\ell = \ell_0$. The resulting differential equation for $\gamma(\ell)$ admits no closed-form solution. It is given by

$$\begin{aligned}
& (4k_0(k_0+2) + \ell(\ell+2)+5)\,k(\ell)^2 - k_0(k_0+2)\left((\ell+1)^2+4\ell_0\right) - 4\ell_0 \\
- \;& 2\left(k_0(k_0+2)(\ell(\ell+2)+2\ell_0-1) + 2(\ell_0-1)\right)k(\ell) \\
= \;& 2\frac{(k_0+1)^2}{1+\ell}(k(\ell)+1)\left((\ell-2\ell_0-1)\,k(\ell) - \ell_0 + \ell(\ell_0+2)\right)\log\left(\frac{(\ell-\ell_0)(k(\ell)+1)}{(\ell_0+1)(\ell-k(\ell))}\right)
\end{aligned}$$

$$+ \quad 2\,(k_0+1)^{\,2}(\ell+1)k'(\ell)\left((\ell_0-\ell)\log\left(\frac{(\ell-\ell_0)\,(k(\ell)+1)}{(\ell_0+1)\,(l-k(\ell))}\right)-\frac{(\ell+1)\,(\ell k(\ell)+\ell_0)}{k(\ell)+1}\right).$$

□

Proof of Proposition 5. Let us start with the designer's payoff, as a function of (t_1, t_2):

$$\int_0^{t_1}\frac{p_0-p_t}{1-p_t}e^{-rt}(1-c)\mathrm{d}t+\int_0^{t_1}\frac{1-p_0}{1-p_t}e^{-rt}(p_t-c)\mathrm{d}t+$$
$$\frac{p_0-p_{t_1}}{1-p_{t_1}}e^{-rt_1}\left((t_2-t_1)e^{-r(t_2-t_1)}+\frac{e^{-r(t_2-t_1)}}{r}\right)(1-c)+$$
$$\frac{1-p_0}{1-p_{t_1}}e^{-rt_1}\left(e^{-r(t_2-t_1)}(t_2-t_1)\frac{p_{t_1}-p_{t_2}}{1-p_{t_2}}+\int_{t_2}^{\infty}e^{-r(t-t_1)}\frac{p_{t_1}-p_t}{1-p_t}\mathrm{d}t\right)(1-c).$$

The first line corresponds to the utility garnered by agents arriving up to t_1, who experiment immediately. The second line is the payoff beyond time t_1 in case the posterior is 1 by then; there are two terms, corresponding to those agents that wait until time t_2, and those that arrive afterwards. Finally, the third line gathers the corresponding terms for the case in which the posterior is $p_{t_1} < 1$ at time t_1. Recall that the belief follows

$$\dot{\ell}_t=\begin{cases}-\lambda(1+\rho)\ell_t & \text{for } t\in[0,t_1];\\ -\lambda\rho\ell_t & \text{if } t\ge t_1.\end{cases}$$

We let

$$\psi:=\frac{\ell_0}{k},\delta:=\frac{\frac{r}{\lambda\rho}}{1+\frac{r}{\lambda\rho}},$$

both in the unit interval. We will hold δ fixed (that is, varying ρ will be done for fixed δ).

A first possibility is to set $t_2 = t_1$. (Clearly, setting $t_2 = +\infty$ is suboptimal.) This cannot be optimal if $t_1 = 0$, however, as the designer could replicate the outcome from full transparency and do better. Inserting $t_2 = 0$ and solving for the optimal t_1, we get

$$t_1=\frac{1}{\lambda(1+\rho)}\ln\left(\psi\left(1+\frac{1}{\rho+\frac{r}{\lambda}}\right)\right),$$

(a global maximum), but then evaluating the first derivative of the payoff w.r.t. t_2 at $t_2 = t_1$ gives a first derivative of 0 and a second derivative whose sign is negative if and only if

$$\rho<\rho_L:=\frac{\delta(1-\delta)\psi}{1-\delta\psi}.$$

If $\rho > \rho_L$, it means that $t_2 > t_1$ would increase the payoff, given t_1, a contradiction. If $\rho < \rho_L$, we also need that $t_1 > 0$, that is, $\rho < \frac{\psi(1-\delta)}{1-\psi}$, but this is easily seen to be implied by

$\rho < \rho_L$. The payoff from this policy involving Learning is (k times)

$$\mathcal{W}_L := \frac{1+\rho}{\rho}\psi\left(\frac{\rho}{\psi(1-\delta+\rho)}\right)^{\frac{1-\delta+\rho}{(1-\delta)(1+\rho)}} - \frac{1-\psi}{1-\delta}.$$

Note that $\mathcal{W}_L$ is decreasing then increasing in ρ, with a minimum at $\rho^* = \frac{\psi}{1-\psi}(1-\delta)$.

The second alternative is transparency (or Delay), that is, $t_1 = 0$. The maximum payoff is then (k times)

$$\mathcal{W}_D := \delta^{\frac{\delta}{1-\delta}}\left(1+\delta - \frac{\delta}{1-\delta}\ln\delta\right)\psi.$$

Note that $\mathcal{W}_D$ is independent of ρ. Clearly, $\mathcal{W}_L > \mathcal{W}_D$ for ρ sufficiently close to 0, and it is readily checked that the inequality is reversed at $\rho = \rho^*$.

Finally, the designer might want to choose $0 < t_1 < t_2$. To solve this problem, we may equivalently maximize the payoff with respect to x_1, x_2, where $x_1 = \delta^{-\frac{\delta}{1-\delta}}e^{-\frac{\delta\lambda(1+\rho)t_1}{1-\delta}}$, and $x_2 = e^{r(t_2-t_1)}$. It holds that $t_1 > 0$ if and only if $x_1 < \delta^{-\frac{\delta}{1-\delta}}$. Computing the payoff explicitly, taking first-order conditions with respect to x_2, we may solve for

$$x_2 = x_1.$$

Plugging into the derivative of the payoff with respect to x_1 gives an expression proportional to

$$\psi(1-\delta)\ln x_1 - \rho(1-\psi)x_1 + \psi(\delta-\rho)(1-\delta).$$

By elementary algebra, this equation admits a root $x_1 < \delta^{-\frac{\delta}{1-\delta}}$ if and only if

$$\rho < \rho_{LD} := W\left(\frac{\psi}{1-\psi}(1-\delta)e^{-(1-\delta)}\right), \tag{25}$$

where $W = W_0$ is the main branch of the Lambert function. It is easy to check that $\rho_{LD} \leq \rho^*$ for all ψ, δ. In this case, the root is given by

$$x_1^\dagger = -\frac{(1-\delta)\psi}{(1-\psi)\rho}W\left(-\frac{\rho(1-\psi)}{(1-\delta)\psi}e^{\rho-\delta}\right),$$

In that case, the designer's strategy involves both learning and delay, and the payoff is given by (k times)

$$\mathcal{W}_{LD} = \frac{1+\rho}{\rho}\delta^{\frac{\delta\rho}{(1-\delta)(1+\rho)}}\left(\rho - W\left(-\frac{\rho(1-\psi)}{(1-\delta)\psi}e^{\rho-\delta}\right)\right)(x_1^\dagger)^{-\frac{1}{1+\rho}} - \frac{1-\psi}{\psi(1-\delta)}.$$

Note that (recall that δ is fixed)

$$\lim_{\rho \to 0} \mathcal{W}_{LD} = e^{\delta}, \quad \lim_{\rho \to 0} \mathcal{W}_{L} = \frac{\psi}{1-\delta}.$$

Hence if $\psi > (1-\delta)e^{\delta}$, the first policy is optimal for small enough ρ. Also, full transparency is necessarily optimal for $\rho > \rho^*$. It is then a matter of tedious algebra to show that $\mathcal{W}_{LD}$ is decreasing in ρ, and can only cross $\mathcal{W}_L$ from below. Furthermore, $\mathcal{W}_{LD} \geq \mathcal{W}_L$ when $\mathcal{W}_L = \mathcal{W}_D$.

To summarize, we have $0 \leq \rho_{LD} \leq \rho^*$. In addition, because $\mathcal{W}_{LD}$ is decreasing and can only cross $\mathcal{W}_L$ from below, while $\mathcal{W}_D$ is independent of ρ, the "ranking" can only be, for $\mathcal{W} := \max\{\mathcal{W}_D, \mathcal{W}_L, \mathcal{W}_{LD}\}$: $\mathcal{W} = \mathcal{W}_L$ for small enough ρ, $\mathcal{W} = \mathcal{W}_{LD}$ for intermediate values of ρ, and $\mathcal{W} = \mathcal{W}_D$ for ρ above some upper threshold. We also have that this upper threshold is at most ρ_{LD} –so that full transparency is indeed optimal for high ρ–, while the lower threshold is strictly positive if and only $\psi > (1-\delta)e^{\delta}$. □

Proof of Proposition 6. This is a perturbation argument around full transparency. Starting from this policy, consider the following modification. At some time t_2 (belief ℓ_2), the designer is fully transparent ($\alpha_2 = 0$). An instant $\Delta > 0$ before, however, he recommends to buy with probability α_1 to some fraction κ of the queue $Q_{t_1} = t_1$, so that the agent is indifferent between checking in and waiting until time $t_2 = t_1 + \Delta$:

$$\ell_0 - \ell_1 - \alpha_1(k - \ell_1) = e^{-r\Delta}(\ell_0 - \ell_2), \tag{26}$$

where

$$\ell_1 = \ell_0 e^{-\lambda \rho t_1},$$

and

$$\ell_2 = \ell_1 e^{-\lambda(\rho\Delta + \kappa\alpha_1 t_1)}.$$

We solve (26) for κ (given ℓ_2), and insert into the payoff from this policy:

$$\mathcal{W}_\kappa = e^{-rt_2}\left((\ell_0 - \ell_2)t_2 + \frac{\ell_0}{r} - \frac{\ell_2}{r+\lambda\rho}\right).$$

Transparency is the special case $\kappa = \Delta = 0$, $t_1 = t^*$, and we compute a Taylor expansion of the gain for small enough Δ with $\ell_1 = \ell^* + \tau\Delta$ and $\alpha_1 = a_1\Delta^2$, with τ, a_1 to be chosen. We pick a_1 so that $\kappa = 1$, which gives

$$a_1 = \frac{\rho(r+\lambda\rho)(\lambda\ell_0\rho r - 2\tau(r+\lambda\rho))}{2\rho(k(\lambda\rho + r) - \ell_0 r) - 2\ell_0 r \ln\left(\frac{r}{r+\lambda\rho}\right)},$$

and choose τ to maximize the first-order term from the expansion, namely, we set

$$\tau = \frac{\lambda \ell_0 \rho^2 r \left(k(\lambda\rho+r)^2 - \ell_0 r(\lambda\rho+r) - \ell_0 r\lambda\right)}{(\lambda\rho+r)^2 \left(\rho(k(\lambda\rho+r)-\ell_0 r) - \ell_0 r \ln\left(\frac{r}{\lambda\rho+r}\right)\right)}.$$

Plugging back into the expansion, we obtain

$$\mathcal{W}_\kappa - \mathcal{W}_0 = \frac{\lambda^2 \ell_0^3 \rho^3 r^3 \left((\lambda\rho+r)\ln\left(\frac{r}{\lambda\rho+r}\right) - \lambda\rho\right)\left(k(\lambda\rho+r)^2 - \ell_0 r(\lambda\rho+\lambda+r)\right)}{(\rho+1)(\lambda\rho+r)^5 \left(\rho(k(\lambda\rho+r)-\ell_0 r) - \ell_0 r \ln\left(\frac{r}{\lambda\rho+r}\right)\right)} \Delta + \mathcal{O}(\Delta^2),$$

and the first term is of the same sign as

$$\ell_0 r(\lambda\rho+r) + \ell_0 r\lambda - k(\lambda\rho+r)^2.$$

Note that this expression is quadratic and concave in $(\lambda\rho+r)$, and positive for $(\lambda\rho+r)=0$. Hence it is positive if and only if it is below the higher of the two roots of the polynomial, *i.e.* if and only if

$$\rho \leq \frac{1}{\lambda}\left(\frac{r\ell_0 + \sqrt{r\ell_0}\sqrt{4k\lambda+\ell_0}}{2k} - r\right).$$

□

Randomized policies with unobserved cost. Here, we prove the following proposition, which generalizes Proposition 7 to the case of randomized policies.

PROPOSITION 8. *With uniformly distributed unobserved cost, the optimal policy is deterministic. It requires full disclosure of the posterior belief.*

We allow the principal to randomize over finitely many paths of experimentation, so there are finitely many possible posterior beliefs, 1, p^j, $j = 1, \ldots, J$. We allow then for multiple (finitely many) recommendations R. So a policy is now a collection $(\alpha_j^R, \gamma_j^R)_j$, depending on the path j that is followed. Along the path j, conditional on the posterior being 1, a recommendation R is given by probability γ_j^R, and conditional on the posterior being p_j, the probabilities α_j^R are used. One last parameter is the probability with which each path j is being used, μ_j.

Correspondingly, there are as many thresholds γ^R as recommendations; namely, given recommendation R, a consumer buys if his cost is no larger than

$$c^R = \frac{\sum_j \mu_j \left(\frac{p_0-p_j}{1-p_j}\beta_j^R + \frac{1-p_0}{1-p_j} p_j \alpha_j^R\right)}{\sum_j \mu_j \left(\frac{p_0-p_j}{1-p_j}\beta_j^R + \frac{1-p_0}{1-p_j}\alpha_j^R\right)},$$

Hence we set

$$\gamma^R = \frac{\sum_j \mu_j \left(\alpha_j^R \ell_j + \beta_j^R(\ell_0 - \ell_j)\right)}{\sum_j \mu_j \alpha_j^R}.$$

We remark for future reference that

$$\begin{aligned}
\sum_R \gamma^R \sum_j \mu_j \alpha_j^R &= \sum_R \sum_j \mu_j \left(\alpha_j^R \ell_j + \beta_k^R(\ell_0 - \ell_j)\right) \\
&= \sum_j \mu_j \left(\left(\sum_R \alpha_j^R\right) \ell_j + \left(\sum_R \beta_j^R\right)(\ell_0 - \ell_j)\right) \\
&= \sum_j \mu_j \ell_0 = \ell_0.
\end{aligned}$$

We now turn to the value function. We have that

$$\begin{aligned}
rV(\ell_1, \ldots, \ell_J) &= \sum_j \mu_j \left(\frac{1+\ell_j}{1+\ell_0} \sum_R \alpha_j^R \int_0^{c^R} (p_j - x)\mathrm{d}x + \frac{\ell_0 - \ell_j}{1+\ell_0} \sum_R \beta_j^R \int_0^{c^R} (1-x)\mathrm{d}x\right) \\
&- \sum_j \ell_j \mu_j \left(\sum_R \alpha_j^R \int_0^{c^R} \mathrm{d}x\right) \frac{\partial V(\ell_1, \ldots, \ell_J)}{\partial \ell_j}.
\end{aligned}$$

We shall do a few manipulations. First, we work on the flow payoff. From the first to the second equation, we gather terms involving the revenue ("p_j" and 1) on one hand, and cost ("x") on the other. From the second to the third, we use the definition of γ^R (in particular, note that the term in the numerator of γ^R appears in the expressions). The last line uses the remark above.

$$\begin{aligned}
&\sum_j \mu_j \left(\frac{1+\ell_j}{1+\ell_0} \sum_R \alpha_j^R \int_0^{c^R} \left(\frac{\ell_j}{1+\ell_j} - x\right) \mathrm{d}x + \frac{\ell_0 - \ell_j}{1+l_0} \sum_R \beta_j^R \int_0^{c^R} (1-x)\mathrm{d}x\right) \\
= \; & \frac{1}{1+\ell_0} \sum_R c^R \sum_j \mu_k \left(\ell_j \alpha_j^R + (\ell_0 - \ell_j)\beta_j^R\right) - \frac{1}{2(1+\ell_0)} \sum_R (c^R)^2 \sum_j \mu_j \left((1+\ell_j)\alpha_j^R + (\ell_0 - \ell_j)\beta_j^R\right) \\
= \; & \frac{1}{1+\ell_0} \sum_R \frac{\gamma^R}{1+\gamma^R} \left(\gamma^R \sum_j \mu_j \alpha_j^R\right) - \frac{1}{2(1+\ell_0)} \sum_R \left(\frac{\gamma^R}{1+\gamma^R}\right)^2 \left((1+\gamma_j^R) \sum_j \mu_j \alpha_j^R\right) \\
= \; & \frac{1}{2(1+\ell_0)} \sum_R \frac{(\gamma^R)^2}{1+\gamma^R} \left(\sum_j \mu_j \alpha_j^R\right) \\
= \; & \frac{1}{2(1+\ell_0)} \sum_R \left(\gamma^R - \frac{\gamma^R}{1+\gamma^R}\right) \left(\sum_j \mu_j \alpha_j^R\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2(1+\ell_0)} \sum_R \gamma^R \sum_j \mu_j \alpha_j^R - \frac{1}{2(1+\ell_0)} \sum_R \frac{\gamma^R}{1+\gamma^R} \left(\sum_j \mu_j \alpha_j^R \right) \\
&= \frac{\ell_0 - \sum_j \mu_j x_j}{2(1+\ell_0)},
\end{aligned}
$$

where we define

$$
x_j := \sum_R \frac{\gamma^R}{1+\gamma^R} \alpha_j^R.
$$

Let us now simplify the coefficient of the partial derivative

$$
\mu_j \left(\sum_R \alpha_j^R \int_0^{c^R} \mathrm{d}x \right) = \mu_j \sum_R \alpha_j^R \frac{\gamma^R}{1+\gamma^R} = \mu_j x_j.
$$

To conclude, given (μ_j) (ultimately, a choice variable as well), the optimality equality simplifies to

$$
rV(\ell_1, \ldots, \ell_J) \quad = \quad \frac{\ell_0}{2(1+\ell_0)} - \sum_j \max_{x_j} \mu_j x_j \left\{ \frac{1}{2(1+\ell_0)} + \ell_j \frac{\partial V(\ell_1, \ldots, \ell_J)}{\partial \ell_j} \right\},
$$

or letting $W = 2(1+\ell_0)V - \frac{\ell_0}{r}$,

$$
rW(\ell_1, \ldots, \ell_J) + \sum_j \mu_j \max_{x_j} x_j \left\{ 1 + \ell_j \frac{\partial W(\ell_1, \ldots, \ell_J)}{\partial \ell_j} \right\} = 0.
$$

where $(x_j)_j$ must be feasible, i.e. appropriate values for (α, γ) must exist. This is a tricky restriction, and the resulting set of (x_j) is convex, but not necessarily a polytope. In particular, it is not the product of the possible quantities of experimentation that would obtain if the agents knew which path were followed, $\times_j \left[\frac{\ell_j}{1+\ell_j}, \frac{\ell_0}{1+\ell_0} \right]$. It is a strictly larger set: by blurring recommendation policies, he can obtain pairs of amounts of experimentation outside this set, although not more or less in all dimensions simultaneously.

Let us refer to this set as B_J. This set is of independent interest, as it is the relevant set of possible experimentation schemes independently of the designer's objective function. This set is difficult to compute, as for a given J, we must determine what values of x can be obtained for *some* number of recommendations. Even in the case $J = 2$, this requires substantial effort, and it is not an obvious result that assuming without loss that $\ell_1 \geq \ell_2$, B_2 is the convex hull of the three points

$$
x^P := \left(\frac{\sum_j \mu_j \ell_j}{1+\sum_j \mu_j \ell_j}, \frac{\sum_j \mu_j \ell_j}{1+\sum_j \mu_j \ell_j} \right), x^S := \left(\frac{\ell_1}{1+\ell_1}, \frac{\ell_2}{1+\ell_2} \right), x^A := \left(\frac{\ell_0 - \mu_2 \ell_2}{1+\ell_0 - \mu_2(1+\ell_2)}, \frac{\ell_2}{1+\ell_2} \right),
$$

and the two curves

$$S^U := \left(x_1, 1 + \frac{\mu_2(1-x_1)}{\mu_1 - (1+\ell_0)(1-x_1)}\right),$$

for $x_1 \in \left[\frac{\ell_1}{1+\ell_1}, \frac{\ell_0 - \mu_2\ell_2}{1+\ell_0-\mu_2(1+\ell_2)}\right]$, and

$$S^L := \left(x_1, x_1 + \frac{(x_1-(1-x_1)\ell_0)(x_1-(1-x_1)(\mu_1\ell_1+\mu_2\ell_2))}{\mu_2(\mu_1\ell_1+\mu_2\ell_2+\ell_0\ell_2-(1+\ell_0)(1+\ell_2)x_1)}\right),$$

for $x_1 \in \left[\frac{\sum_j \mu_j\ell_j}{1+\sum_j \mu_j\ell_j}, \frac{\ell_1}{1+\ell_1}\right]$, that intersect at the point

$$\left(\frac{\ell_1}{1+\ell_1}, \frac{\ell_0-\mu_1\ell_1}{1+\ell_0-\mu_1(1+\ell_1)}\right).$$

It is worth noting that the point $\left(\frac{\ell_0}{1+\ell_0}, \frac{\ell_0}{1+\ell_0}\right)$ lies on the first (upper) curve, and that the slope of the boundary at this point is $-\mu_1/\mu_2$: hence, this is the point within B_2 that maximizes $\sum_j \mu_j x_j$. See Figure 5 below. To achieve all extreme points, more than two messages are necessary (for instance, achieving x^S requires three messages, corresponding to the three possible posterior beliefs at time t), but it turns out that three suffice.

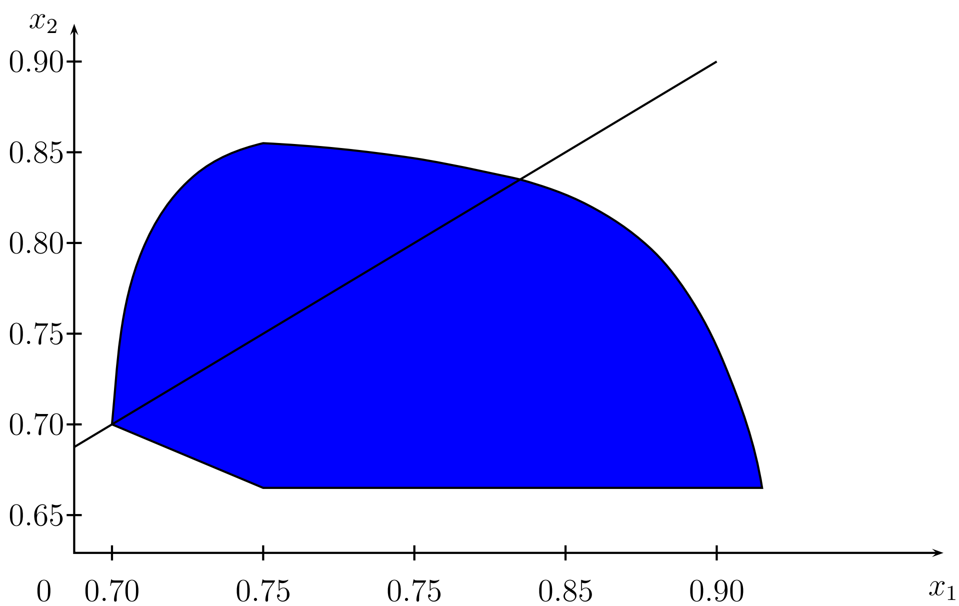

Figure 5: Region B_2 of feasible (x_1, x_2) in the case $J = 2$ (here, for $\ell_0 = 5, \ell_1 = 3, \ell_2 = 2, \mu_1 = 2/3$.

In terms of our notation, the optimum value of a non-randomized strategy is

$$W^S(\ell) = -e^{\frac{r}{\ell}} E_{1+r}\left(\frac{r}{\ell}\right).$$

We claim that the solution to the optimal control problem is given by the "separating" strategy, given μ and $\mathbf{l} = (\ell_1, \ldots, \ell_K)$, *for the case* $J = 2$ to begin with. That is,

$$W(\mathbf{l}) = W^S(\mathbf{l}) := -\sum_j \mu_j W^S(\ell_j).$$

To prove this claim, we invoke a verification theorem (see, for instance, Thm. 5.1 in Fleming and Soner, 2005). Clearly, this function is continuously differentiable and satisfies the desired transversality conditions on the boundaries (when $\ell_j = 0$). We must prove that it achieves the maximum. Given the structure of B_2, we have to ensure that for every state ℓ and feasible variation $(\partial x_1, \partial x_2)$, starting from the policy $x = x^S$, the cost increases. That is, we must show that

$$\sum_j \mu_j \left(1 + \ell_j \frac{\mathrm{d}W^S(\ell_j)}{\mathrm{d}\ell_j}\right) \partial x_j \geq 0,$$

for every ∂x such that (i) $\partial x_2 \geq 0$, (ii) $\partial x_2 \geq -\frac{\mu_1}{\mu_2}\frac{1+\ell_1}{1+\ell_2}\partial x_1$. (The first requirement comes from the fact that x^S minimizes x_2 over B_2; the second comes from the other boundary line of B_2 at x^S.) Given that the result is already known for $J = 1$, we already know that this is true for the special cases $\partial x_j = 0, \partial x_{-j} \geq 0$. It remains to verify that this holds when

$$\partial x_2 = -\frac{\mu_1}{\mu_2}\frac{1+\ell_1}{1+\ell_2}\partial x_1,$$

i.e. we must verify that, for all $\ell_1 \geq \ell_2$,

$$(1+\ell_1)\ell_2 \frac{\mathrm{d}W^S(\ell_j)}{\mathrm{d}\ell_2} - (1+\ell_2)\ell_1 \frac{\mathrm{d}W^S(\ell_j)}{\mathrm{d}\ell_1} \geq \ell_2 - \ell_1,$$

or rearranging,

$$\frac{\ell_2}{1+\ell_2}\left(\frac{\mathrm{d}W^S(\ell_j)}{\mathrm{d}\ell_2} - 1\right) - \frac{\ell_1}{1+\ell_1}\left(\frac{\mathrm{d}W^S(\ell_j)}{\mathrm{d}\ell_1} - 1\right) \geq 0,$$

which follows from the fact that the function $\ell \mapsto \frac{\ell}{1+\ell}\left(\frac{\mathrm{d}\left[re^{\frac{r}{\ell}}E_{1+r}\left(\frac{r}{\ell}\right)\right]}{\mathrm{d}\ell} - 1\right)$ is decreasing.

To conclude, starting from $\ell_1 = \ell_2 = \ell_0$, the value of μ is irrelevant: the optimal strategy ensures that the posterior beliefs satisfy $\ell_1 = \ell_2$. Hence, the principal does not randomize.

The argument for a general J is similar. Fix $\ell_0 \geq \ell_1 \geq \cdots \geq \ell_J$. We argue below below that, at x^S, all possible variations must satisfy, for all $j'=1,\ldots,J$,

$$\sum_{j=j'}^{J}\mu_j(1+\ell_j)\partial x_j\geq 0,$$

It follows that we have

$$\sum_j \mu_j\left(1+\ell_j\frac{\mathrm{d}W^S(\ell_j)}{\mathrm{d}\ell_j}\right)\partial x_j=\frac{\ell_1}{1+\ell_1}\left(\frac{\mathrm{d}W^S(\ell_1)}{\mathrm{d}\ell_1}-1\right)\sum_{j'=1}^{J}\mu_{j'}(1+\ell_{j'})\partial x_{j'}+$$
$$\sum_{j=1}^{J-1}\left(\frac{\ell_{j+1}}{1+\ell_{j+1}}\left(\frac{\mathrm{d}W^S(\ell_{j+1})}{\mathrm{d}\ell_{j+1}}-1\right)-\frac{\ell_j}{1+\ell_j}\left(\frac{\mathrm{d}W^S(\ell_j)}{\mathrm{d}\ell_j}-1\right)\right)\sum_{j'=j+1}^{J}\mu_{j'}(1+\ell_{j'})\partial x_{j'}\geq 0,$$

by monotonicity of the map $\frac{\ell}{\ell+1}\left(\frac{\partial W^S(\ell)}{\partial \ell}-1\right)$, as in the case $J=2$.

To conclude, we argue that, from x^S, all variations in B_J must satisfy, for all j',

$$\sum_{j=j'}^{J}\mu_j(1+\ell_j)\partial x_j\geq 0.$$

In fact, we show that all elements of B satisfy

$$\sum_{j=j'}^{J}\mu_k\left((1+\ell_j)x_j-\ell_j\right)\geq 0,$$

and the result will follow from the fact that all these inequalities trivially bind at x^S. Consider the case $j'=1$, the modification for the general case is indicated below. To minimize

$$\sum_{j=1}^{J}\mu_j(1+\ell_j)x_j,$$

over B_J, it is best, from the formula for x_j (or rather, γ^R that are involved), to set $\gamma^{R'}=1$ for some R' for which $\alpha_j^{R'}=0$, all j. (To put it differently, to minimize the amount of experimentation conditional on the low posterior, it is best to disclose when the posterior belief is one.) It follows that

$$\begin{aligned}&\sum_j \mu_j\left[(1+\ell_j)x_j-\ell_j\right]\\ =\ &\sum_j \mu_j\left[(1+\ell_j)\sum_R \alpha_j^R\frac{\sum_{k'}\mu_{j'}\ell_{j'}\alpha_{j'}^R}{\sum_{j'}\mu_{k'}(1+\ell_{j'})\alpha_{j'}^R}-\ell_j\right]\end{aligned}$$

$$
\begin{aligned}
&= \sum_R \mu_j(1+\ell_j)\alpha_j^R \frac{\sum_{j'} \mu_{j'}\ell_{j'}\alpha_{j'}^R}{\sum_{j'} \mu_{j'}(1+\ell_{j'})\alpha_{j'}^R} - \sum_j \mu_j \ell_j \\
&= \sum_R \sum_{j'} \mu_{j'}\ell_{j'}\alpha_{j'}^R - \sum_j \mu_j\ell_j = \sum_{j'} \mu_{j'}\ell_{j'} \sum_R \alpha_{j'}^R - \sum_j \mu_j \ell_j = 0.
\end{aligned}
$$

The same argument generalizes to other values of j'. To minimize the corresponding sum, it is best to disclose the posterior beliefs that are above (*i.e.*, reveal if the movie is good, or if the chosen j is below j'), and the same argument applies, with the sum running over the relevant subset of states.

$\square$