OPTIMAL DESIGN FOR SOCIAL LEARNING

By

Yeon-Koo Che and Johannes Hörner

April 2015

COWLES FOUNDATION DISCUSSION PAPER NO. 2000

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Optimal Design for Social Learning

Yeon-Koo Che* Johannes Hörner†

first draft: November 25, 2012; current draft: April 28, 2015

Abstract

This paper studies the design of a recommender system for organizing social learning on a product. To improve incentives for early experimentation, the optimal design trades off fully transparent social learning by over-recommending a product (or "spamming") to a fraction of agents in the early phase of the product cycle. Under the optimal scheme, the designer spams very little about a product right after its release but gradually increases the frequency of spamming and stops it altogether when the product is deemed sufficiently unworthy of recommendation. The optimal recommender system involves randomly triggered spamming when recommendations are public—as is often the case for product ratings—and an information "blackout" followed by a burst of spamming when agents can choose when to check in for a recommendation. Fully transparent recommendations may become optimal if a (socially-benevolent) designer does not observe the agents' costs of experimentation.

Keywords: experimentation, social learning, mechanism design.
JEL Codes: D82, D83, M52.

1 Introduction

Most of our choices rely on recommendations by others. Whether selecting movies, picking stocks, choosing hotels or shopping online, the experiences of others can teach us to make better decisions. Internet platforms are increasingly acting as recommenders, enabling users to learn from other users on a scale never before imagined. *Amazon* (books) and *Netflix* (movies) are two well-known recommenders, but there are platforms that mediate social

*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.

learning among users for almost any "experience" good, including *Yelp* (restaurants), *TripAdvisors* (hotels), *RateMD* (doctors), and *RateMyProfessors* (professors), to name just a few. Search engines such as *Google*, *Bing* and *Yahoo* organize social discovery of relevant websites based on the "search experiences" of users. Social medias such as *Facebook* and *LinkedIn* do the same for the other quintessential "experience" good, friends, linking users to new potential friends based on the experiences of other friends.

While the economics of social learning are now well appreciated, the prior literature has focused only on its positive aspects.[1] A normative perspective—namely, how best to encourage users to experiment with new products and to disseminate their experiences to others—has not yet been developed. Incentivizing early users to experiment with products is challenging precisely because they cannot internalize the benefits that later users will reap. In fact, the availability of social learning may dampen the incentives for early experimentation because users might rather wait and free ride on the information provided by others in the future.[2] In other words, a recommender system may crowd out information production and undermine its very foundations. The absence of sufficient initial experimentation—known as the "cold start" problem—might lead to a collapse of social learning and to the death of products (even pre-launch) that are worthy of discovery by mainstream consumers. The cold start problem is particularly relevant for new and unestablished products, which are no longer the exception but rather the rule due to the proliferation of self production.[3]

This paper studies a recommender system that optimally achieves the dual purposes of social learning: production and dissemination of information. To maintain realism, we focus on non-monetary tools for achieving these purposes. Indeed, monetary transfers are seldom used to compensate for experimentation perhaps because it is difficult to verify whether a reviewer has performed genuine experimentation conscientiously and submitted an unbiased review.[4]

Our key insight is that the judicious use of the recommender system itself can incentivize consumer experimentation. To illustrate the idea, suppose that a recommender system—an online movie platform—generally recommends movies to users based on the reviews of

[1]For instance, the main question studied in the literature involves the circumstances under which observational learning leads to full revelation of the underlying state.

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

[3]For instance, once considered vanity publishing, self-publication has expanded dramatically in recent years with the availability of easy typesetting and e-books. 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 in its infancy, 3D printing and other similar technologies anticipate a future that will feature an even greater increase of self-manufactured products. The popularity of self production suggests a marketplace populated by such a large number of products/titles that they will not be easily recognized, which will heighten the importance of a recommender system even further.

[4]Attempts have been made in this regard that are limited in scope. For instance, the Amazon Vine Program rewards selected reviewers with free products, and LaFourchette.com grants discounts for (verified) diners that write reviews and make reservations via their site.

past users (truthful recommendations) but occasionally recommends new movies that require experimentation (fake recommendations or "spam").[5] As long as the platform keeps its users uninformed as to whether a recommendation is truthful or spam—and as long as it commits not to "spam" too much—users will happily follow the recommendation and perform the necessary experimentation in the process. Indeed, we show that an optimally designed recommender system trades off fully transparent social learning by over-recommending a product in the early phase of its life cycle.

Of course, the extent to which information acquisition can be motivated in this manner depends on the agent's cost of acquiring information and the frequency with which the platform provides truthful recommendations (as opposed to spam). Another important feature involves the dynamics of how the platform combines truthful recommendations with spam over time after the product's initial release (*e.g.*, a movie release date). For instance, it is unlikely that many users will have experimented with an unknown product immediately after its release, so recommending such a product during its early stage is likely to be met with skepticism. Therefore, to be credible, the platform must commit to truthful recommendations with sufficiently high probability in the early stage of the product's life cycle, which means that the designer will spam very little in the early stage and that learning will be slow. Over time, however, recommendation becomes credible, and hence the pace of learning will accelerate. This suggests that there will be nontrivial dynamics in the optimal recommendation strategy and in social learning.

The present paper explores how a recommender system optimally balances the tradeoff between experimentation and learning, what type of learning dynamics such a mechanism would entail and what implications these learning dynamics will have on welfare, particularly when compared with: (i) *no social learning* (where there is no platform supervision of learning) and (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 a salient normative benchmark that is important for any market design or public policy inquiry. From a more positive perspective, social welfare maximization might also result from platforms that compete a la Bertrand by charging membership fees and providing recommendations on a collection of products of varying vintages.[6]

[5] Throughout, the term "spam" means an unwarranted recommendation, more precisely a recommendation of a product that has yet to be found worthy of recommendation.

[6] One can imagine that each such platform provides recommendations for large collections of products of varying vintages. Providing recommendations on a number of products of different vintages means that the social welfare gain from optimal experimentation will be spread evenly across users arriving at different times. For instance, a user who "sacrifices" himself or herself for future users on certain products will be overcompensated by the benefits from past learning on other products. Hence, firms competing a la Bertrand on membership fees will be forced to offer the maximum intertemporal social welfare on each product.

While the optimal recommender system we identify rests on a normative justification, its main feature is consistent with certain observed practices. For instance, it is known that search engines such as *Google* periodically shuffle their rankings of search items to provide exposure to relatively new or unknown sites, which is highly consistent with the idea of distorting recommendations for under-exposed and lesser-known items. More generally, our finding is consistent with the casual observation that ratings of many products appear to be inflated.[7] Ratings are often inflated by sellers (as opposed to the platforms), who have every interest to promote their products even against the interests of consumers.[8] However, platforms have instruments at their disposal to control the degree of ratings inflation, such as filters that detect false reviews, requiring verification of purchase for positing reviews, and allowing users to vote for "helpful" reviews.[9] Our theory suggests that some tolerance for review inflation can result from the optimal supervision of social learning by a benevolent designer.

Our starting point is the standard "workhorse" model of experimentation that is borrowed from Keller, Rady and Cripps (2005). The designer offers a product to agents with unknown (binary) value. By consuming this product, a possibly costly choice, short-run agents might discover whether its value is high. Here, we are not interested in agents' incentives to truthfully report their experience to the designer: because they consume this product only once, they are willing to do so. We thus postulate that some exogenous fraction of the agents who consume will share the information with the designer (platform); hence, the greater the number of agents who consume, the more the designer learns about the product. The learning by the designer takes the form of conclusively good news about the product, with the news arriving at a rate that is proportional to the aggregate consumption/experimentation by agents. This structure might arise from an algorithm employed by a platform that aggregates the cumulative reviews from consumers into a simple "up" or "down" signal, for instance. Importantly, agents do not directly communicate with one another. The designer mediates the transmission of information, which raises a difficult problem: How should she control social learning so as to yield the right amount of experimentation?

In the baseline model in which agents cannot delay their check-in/consumption, the optimal policy of the designer involves revealing good news truthfully but recommending the product even without the news, or "spamming," as long as the designer's posterior belief remains above a certain threshold. Unlike *no social learning* and *full transparency*, spamming induces experimentation, but optimal policy keeps the frequency of spam below a certain

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

[8]Luca and Zervas (2014) suggest that as much as 16% of Yelp reviews are suspected to be fraudulent.

[9]Mayzlin et al. (2012) find that *Expedia*'s requirement that a reviewer must verify her stay to leave a review on a hotel resulted in fewer false reviews at *Expedia* compared with *TripAdvisor*, which has no such requirement.

maximal "capacity" necessary to maintain credibility. This creates learning dynamics that are single-peaked: the designer spams very little on a product immediately after its release but gradually increases the frequency of spamming over time until a critical belief threshold is reached, at which point she stops spamming altogether.

We then extend the model in several directions. First, we study public recommendations that are common in product ratings. Unlike private recommendations, public recommendations are revealed to all agents, and this limits the scope of the designer's ability to manipulate agents' beliefs. The optimal policy nevertheless induces experimentation, albeit at a slower average pace than under private recommendation. Notably, the designer randomly selects a time to trigger spam (*e.g.*, a "good" rating), and the spam then lasts until the designer's belief reaches a threshold (after which the spam is halted).

Second, we allow agents to choose when to "check-in" for a recommendation. Endogenous entry raises two novel issues. First, it is desirable to encourage early agents to wait for background information to accumulate sufficiently before making decisions. Second, those induced to experiment must be discouraged from delaying their check-in to free ride on others' experimentation. The optimal recommender policy now involves an information "blackout" for a duration of time after release of the product, effectively inducing agents to wait until some information accumulates, followed by a massive amount of spam that induces them to experiment in a short burst. Subsequently, spamming tapers off gradually, inducing agents to experiment at rates that decrease smoothly over time until it stops altogether—a feature crucial for controling free riding.

Third, we consider heterogeneity in the agents' costs of experimentation. If costs are observable, then the designer can tailor her recommendations to agents' costs levels, inducing them to experiment at different rates. If the costs are not observable, however, the designer's ability to tailor recommendations to the agents' costs is limited, which might lead the designer to employ a fully transparent recommendation policy, under reasonable circumstances.

In addition to the exponential bandit literature, our paper relates to several other strands of literature. First, our model can be viewed as introducing optimal design into the standard model of social learning (hence the title). In standard models (for instance, Bikhchandani, Hirshleifer and Welch, 1992; Banerjee, 1993; Smith and Sørensen, 2000), a sequence of agents takes actions myopically, ignoring their effects on the learning and welfare of agents in the future. Smith, Sørensen and Tian (2014) study altruistic agents who distort their actions to improve observational learning for posterity. Frick and Ishii (2014) examine how social learning affects the adoption of innovations of uncertain quality and explain the shape of commonly observed adoption curves. Our focus here is instead dynamic control of information to agents to incentivize their (unverifiable) experimentation. Such dynamic control of information is present in Gershkov and Szentes (2009), but that paper considers a very different environment, as there are direct payoff externalities (voting). Much more closely related to the present paper is a recent study by Kremer, Mansour and Perry (2014). They study the optimal mechanism for inducing agents to explore over multiple products. In their model, the quality of the product is ascertained once a single consumer buys it, so learning is instantaneous. By contrast, learning is gradual in our model, and controlling its dynamic trajectory is an important aspect of the mechanism design we focus on. We also study public recommendations as well as endogenous entry by agents, which have no analogues in their model.

Our paper also contributes to the literature on Bayesian persuasion that studies how a principal can credibly manipulate agents' beliefs to influence their behavior. Aumann, Maschler and Stearns (1995) analyze this question in repeated games with incomplete information, whereas Kamenica and Gentzkow (2011), Rayo and Segal (2010), and Ostrovsky and Schwarz (2010) study the problem in a variety of organizational settings. The current paper pursues the same question in a dynamic setting. In this regard, the current paper joins a burgeoning literature that studies Bayesian persuasion in dynamic settings (see Ely, Frankel and Kamenica (2015), Renault, Solan and Vieille (2014), Ely (2015), and Halac, Kartik, and Liu (2015)). The focus on social learning distinguishes the present paper from these other papers.[10]

Finally, the present paper is related to the empirical literature on the user-generated reviews (Jindal and Liu, 2008; Luca and Zervas, 2014; and Mayzlin et al. (2014)).[11] These papers suggest ways to empirically identify manipulations in the reviews made by users of internet platforms such as *Amazon, Yelp* and *TripAdvisor.* Our paper contributes a normative perspective on the extent to which the manipulation should be controlled.

2 Illustrative Example

We begin with a simple example that highlights the main themes of this paper: (1) the optimal policy trades off social learning to incentivize experimentation, and (2) the optimal policy involves slower experimentation and more transparency when agents are able to free ride.

Suppose a product, say a movie, is released at time $t = 0$, and a unit mass of agents arrive at each time $t = 1, 2$. The quality of the movie is either "good" ($\omega = 1$), in which case the movie yields a surplus of 1 to an agent, or "bad" ($\omega = 0$), in which case it yields a surplus of 0. The quality of the movie is unknown at the time of its release, with prior $p^0 := \Pr\{\omega = 1\} \in [0, 1]$. Watching the movie costs each agent $c \in (p^0, 1)$; thus, without

[10]Papanastasiou, Bimpikis and Savva (2014) show the insight of the current paper extends to the two-product context, but without fully characterizing the optimal mechanism.

[11]Dai, Jin, Lee and Luca (2014) offer a structural approach to aggregate consumer ratings and apply it to restaurant reviews from *Yelp*.

further information, the agents would never watch the movie.

At time $t = 0$, the designer receives a signal $\sigma \in \{g, n\}$ (from its marketing research, for example) about the quality of the movie with probabilities:

$$\Pr\{\sigma = g|\omega\} = \begin{cases} \rho_0 & \text{if } \omega = 1; \\ 0 & \text{if } \omega = 0, \end{cases}$$

and $\Pr\{\sigma = n|\omega\} = 1 - \Pr\{\sigma = g|\omega\}$. In other words, the designer receives good news only when the movie is good; but she also may receive no news (even) when the movie is good.[12]

Suppose that the designer has received no news at the beginning of $t = 1$ but that a fraction α of agents saw the movie at $t = 1$. Then, the designer again receives conclusively good news with probability:

$$\Pr\{\sigma = g|\omega\} = \begin{cases} \alpha & \text{if } \omega = 1; \\ 0 & \text{if } \omega = 0. \end{cases}$$

The feature that the signal becomes more informative with a higher fraction α of agents experimenting at $t = 1$ captures the learning benefit that they confer to the $t = 2$ agents.

The designer chooses her recommendation policy to maximize social welfare and does so with full commitment power. Specifically, she recommends the movie to a fraction of agents in each period based on her information at that point in time.[13] The designer discounts the welfare in period $t = 2$ by a factor $\delta \in (0, 1)$.

We consider two possible scenarios in terms of whether the agents arriving in $t = 1$ can wait until $t = 2$.

Exogenous Check-In. Suppose the agents arriving in $t = 1$ cannot delay their check-in/consumption to $t = 2$. The designer's optimal policy is then as follows. First, the

[12] Thus, it follows that the designer's posterior at time $t = 1$ on $\omega = 1$ is 1 with a probability of $\rho_0 p^0$ (in the event that she receives good news) and

$$p_1 = \frac{(1-\rho_0)p^0}{(1-\rho_0)p^0 + 1 - p^0},$$

with a probability of $1 - \rho_0 p^0$ (in the event that she receives no news).

[13] The designer would not gain from a stochastic recommendation policy. To see this, compare two choices: i) the designer recommends the movie to a fraction α of the agents, and ii) the designer recommends it to all agents with probability α. For agents in $t = 1$, the two options are the same in terms of welfare and thus in terms of incentives. For agents in $t = 2$, the former results in the learning of good news with probability $p^0(\rho_0 + (1-\rho_0)\alpha)$, whereas the latter does the same because when there is no good news at the beginning of $t = 1$, all agents are recommended the movie with probability α. This equivalence means that public recommendation entails no loss in this example; however, the equivalence does not hold in our general model.

designer is truthful at time $t = 2$, as lying can only reduce welfare and can never improve the incentive for experimentation at $t = 1$.

Consider now time $t = 1$. If good news has arrived by then, the designer would again recommend the movie to all agents at $t = 1$. Suppose that no news has been received by then and that the designer nevertheless recommends—or "spams"—to a fraction α of the agents. The agents receiving the recommendation cannot determine whether the recommendation is genuine or spam; instead, they would form a posterior:

$$P_1(\alpha) := \frac{\rho_0 p^0 + \alpha p^0(1-\rho_0)}{\rho_0 p^0 + (1-\rho_0 p^0)\alpha}.$$

If the designer spams to all agents (*i.e.*, $\alpha = 1$), then they will find the recommendation completely uninformative, and hence $P_1(\alpha) = p^0$. Since $p^0 < c$, they would never watch the movie. By contrast, if the designer spams rarely (*i.e.*, $\alpha \simeq 0$), then $P_1(\alpha) \simeq 1$, *i.e.*, they will be confident (and nearly certain) that the recommendation is genuine. Naturally, the agents receiving recommendations will definitely watch the movie in this case. Because the recommendation is more credible the less the designer spams, $P_i(\alpha)$ is decreasing in α. In particular, there is a maximal fraction $\hat{\alpha} =: \frac{(1-c)\rho_0 p^0}{c(1-\rho_0 p^0) - p^0(1-\rho_0)}$ of agents who can be induced to experiment.

Social welfare,

$$W(\alpha) := p^0(\rho_0 + (1-\rho_0)\alpha)(1-c)(1+\delta) - \alpha(1-p^0)c,$$

consists of the benefit from the good movie being recommended (the first term) and the loss borne by the $t = 1$ agents from a bad movie being recommended (the second term). In particular, the benefit contains the benefit that experimentation by the $t = 1$ agents confers to the $t = 2$ agents, which is captured by the term $p^0(1-\rho_0)\alpha(1-c)\delta$.

The optimal policy is to "spam" up to $\hat{\alpha}$, if W is increasing in α, *i.e.*, if the social value of experimentation at date 1 justifies the cost:

$$p^0 \geq \hat{p}^0 := \frac{c}{(1-\rho_0)(1+\delta)(1-c)+c}. \tag{1}$$

Note that the RHS is strictly less than c when $\rho_0 < \frac{\delta}{1+\delta}$. In that case, if $p^0 \in (\hat{p}^0, c)$, the designer will "spam" the agents at $t = 1$ to consume against their myopic interest.

Endogenous Check-In and Free-Riding. Suppose next that the agents arriving at $t = 1$ can wait until $t = 2$ to check in for a recommendation. (We assume that check-in for a recommendation is costly and that an agent thus checks in only once.) Agents may delay their check-in to free ride on experimentation by other agents. We will see how the free-riding concern affects the optimal recommendation policy.

As before, it is intuitively clear that the designer always recommends the movie in the event of good news at $t = 1$, as this improves welfare without sacrificing agents' incentives to experiment at time $t = 1$. Suppose that the designer recommends the movie to a fraction α of the agents at time $t = 1$ in the event that no news is received. Further, suppose at time $t = 2$ the designer provides a recommendation to a fraction γ_2 of agents in the event of good news and to a fraction α_2 in the event of no news. If an agent arriving at $t = 1$ checks in at $t = 1$, then his payoff is

$$U_1(\alpha) := p^0(\rho_0 + (1 - \rho_0)\alpha)(1 - c) - (1 - p^0)\alpha c.$$

If the agent waits and checks in at $t = 2$, he will receive

$$U_2(\alpha, \gamma_2, \alpha_2) := \delta p^0(\rho_0 + (1 - \rho_0)\alpha)\gamma_2(1 - c) - \delta(1 - p^0)\alpha_2 c.$$

Incentive compatibility then requires $U_1(\alpha) \geq U_2(\alpha, \gamma_2, \alpha_2)$.[14]

The designer's problem is then to solve

$$\max_{\alpha,\gamma_2,\alpha_2} U_1(\alpha) + U_2(\alpha, \gamma_2, \alpha_2)$$

subject to

$$U_1(\alpha) \geq U_2(\alpha, \gamma_2, \alpha_2).$$

At $t = 2$, the optimal policy is *full transparency* ($\gamma_2 = 1$ and $\alpha_2 = 0$).[15] At $t = 1$, assuming (1), the optimal policy is to spam up to a level $\alpha = \tilde{\alpha}(\delta)$ that satisfies $U_1(\tilde{\alpha}(\delta)) = U_2(\tilde{\alpha}(\delta), 1, 0)$. For any $\delta < 1$, $\tilde{\alpha}(\delta) > 0$, so the designer does spam. However, for any $\delta > 0$, $\tilde{\alpha}(\delta) < \hat{\alpha}$; *i.e.*, the designer spams less than under exogenous check in. This follows from the fact that the incentive constraint is now stronger: the designer should not only incentivize agents to experiment at $t = 1$ but also keep them from free-riding on others by delaying their check-in.[16] This requires the designer to spam less and induce less experimentation. In fact, as $\delta \to 1$, $\tilde{\alpha}(\delta) \to 0$. In other words, as $\delta \to 1$, the optimal policy converges to *full*

[14] If $\alpha = 0$, there will be no additional learning at $t = 1$, and the $t = 1$ agents will have no incentive to wait until $t = 2$. Hence, the inequality holds. If $\alpha > 0$, then the designer must be inducing some agents to check in at $t = 1$ and experiment; otherwise, $\alpha = 0$ combined with no waiting dominates that policy. To keep $t = 1$ agents from waiting, it must be that $U_1(\alpha) \geq U_2(\alpha, \gamma_2, \alpha_2)$.

[15] It is easy to see that at the optimum $\gamma_2 = 1$ and $\alpha_2 = 0$. Suppose not. Then, it must be the case that $\alpha > 0$; otherwise, the incentive constraint would not be binding and either raising γ_2 or lowering α_2 would increase the objective without violating the constraint. Given that $\alpha > 0$, suppose that one reduces α and simultaneously either raises γ_2 or lowers α_2 to hold $U_2(\alpha, \gamma_2, \alpha_2)$ constant. This will increase $U_1(\alpha)$ because $U_1(\cdot)$ is decreasing due to the assumption that $p^0 < c$.

[16] This latter result can be seen by the fact that $U_2(\tilde{\alpha}(\delta), 1, 0) > 0$, which means that the incentive constraint is stronger than simply ensuring that the agents in $t = 1$ break even.

transparency.

3 Model

Our model generalizes the example in terms of its timing and information structure. A product is released at time $t = 0$, and, for each time $t \geq 0$, a constant flow of unit mass of agents arrive and decide whether to consume the product. In the baseline model, the agents are short-lived—they make a one-time decision and then exit the market permanently. (We later extend the model to allow the agents to delay their entry after their arrival.) An agent incurs cost $c \in (0, 1)$ for consumption. This cost can be the opportunity cost of time spent or a price charged. The product is either "good," in which case an agent derives the (expected) surplus of 1, or "bad," in which case the agent derives the (expected) surplus of 0. The quality of a product is *a priori* uncertain but may be revealed over time.[17] At time $t = 0$, the probability of the product being good, or simply "the prior," is p^0. We shall consider all values of the prior, although the most interesting case will be $p^0 \in (0, c)$, which makes non-consumption myopically optimal.

Agents do not observe previous agents' decisions or their experiences. There is a **designer** who can mediate social learning by collecting information from previous agents and disclosing that information to current agents. We can think of the designer as an Internet platform, such as *Netflix*, *Google* or *Microsoft*, that has access to users' activities and reviews and makes product recommendations based on them. As is natural with these examples, the designer may obtain information from its own marketing research and other sources including agents' experiences themselves. For instance, there may be some flow of "fans" who experiment with the product at zero cost. We thus assume that some information arrives at a constant base rate $\rho > 0$ plus the rate at which agents experience the product. Specifically, if a flow of size μ consumes the product over some time interval $[t, t + \mathrm{d}t)$, then the designer learns during this time interval that the movie is "good" with probability $\lambda(\rho + \mu)\mathrm{d}t$, where $\lambda > 0$ is the arrival rate of good news and $\rho > 0$ is the rate at which the designer obtains the information regardless of the agents' behavior. Section 7.1 extends our model to allow for news to be (conclusively) bad, showing that our qualitative results continue to hold so long as the arrival rate of the good news exceeds that of the bad news. The designer begins with the same prior p^0, and the agents do not have access to "free" learning.

[17]The agents' preferences may involve an idiosyncratic component that is realized *ex post* after consuming the product; the quality then captures only their common preference component. The presence of an idiosyncratic preference component does not affect the analysis because each agent must decide based on the expected surplus he will derive from his consumption of the product. Idiosyncratic preferences would make the designer's interpretation of the agents' reviews nontrivial because a good product may receive a bad review and a bad product may receive a good review, which motivates our assumption that the arrival of the news signal is gradual.

The designer provides feedback on the movie to agents at each time based on the information she has learned by that time. Since agents' decisions are binary, without loss, the designer simply decides whether to recommend the product. The designer commits to the following policy: At time t, she recommends the movie to a fraction $\gamma_t \in [0,1]$ of agents if she learns that the movie is good, and she recommends or **spams** 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 agents, discounted at rate $r > 0$, over the (measurable) functions (γ_t, α_t).

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

$$\mu_t = \rho + \alpha_t.$$

Suppose that no news has arrived by $t + \mathrm{d}t$, then the designer's updated posterior at time $t + \mathrm{d}t$ must be

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

Rearranging and simplifying, the posterior must follow the law of motion:[18]

$$\dot{p}_t = -\lambda\mu_t p_t(1 - p_t), \tag{2}$$

with the initial value at $t = 0$ given by the prior p^0. Notably, the posterior falls as time passes, as "no news" leads the designer to form a pessimistic inference regarding the movie's quality.

In our model, agents do not directly observe the designer's information or her belief. However, they can form a rational belief about the designer's belief. They know that the designer's belief is either 1 or p_t, depending on whether good news has been received by time t. Let g_t denote the probability that the designer has received good news by time t. This belief g_t is pinned down by the martingale property, *i.e.*, that the designer's posterior must

[18]Subtracting p_t from both sides and rearranging, we obtain

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

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

on average equal the prior:

$$g_t \cdot 1 + (1 - g_t)p_t = p^0. \tag{3}$$

Notably, g_t rises as p_t falls, *i.e.*, the agents find it increasingly probable that the news has arrived as time progresses.

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

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

where (p_t, g_t) must follow the required laws of motion: (2) and (3), and $\mu_t = \rho + \alpha_t$ is the total experimentation rate and r is the discount rate of the designer.[19] The welfare consists of the discounted value of consumption—$1 - c$ in the event of good news and $p_t - c$ in the event of no news—for those the designer induces to consume.

In addition, for the policy (α, γ) to be implementable, the agents must have an incentive to follow the recommendation. Since the exact circumstances of the recommendation (whether the agents are recommended because of good news or despite no news) is kept hidden from the agents, their incentives for following the recommendation depend on their posterior regarding the information held by the designer. Specifically, an agent will have the incentive to consume the product, if and only if the posterior that the movie is good is no less than the cost:

$$\frac{g_t \gamma_t + (1 - g_t) \alpha_t p_t}{g_t \gamma_t + (1 - g_t) \alpha_t} \geq c. \tag{4}$$

(There is also an incentive constraint for the agents not to consume the product when it is not recommended by the designer. Because this constraint will not be binding throughout—as the designer typically desires more experimentation than do 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**: In this regime, the agents receive no information from the designer, and hence they decide solely based on the prior p^0. When $p^0 < c$, no agent ever consumes.
- **Full Transparency**: In this regime, the designer discloses her information—or her

[19] We allow the designer to randomize over (γ, α) although we show in the Appendix that such a policy is never optimal.

beliefs—truthfully to the agents. In our framework, full disclosure is implemented by the policy of $\gamma_t \equiv 1$ and $\alpha_t = \mathbf{1}_{\{p_t \geq c\}}$.

- **First-Best**: In this regime, the designer optimizes her policy without having to satisfy the incentive compatibility constraint (4).

To distinguish the current problem from the *first-best*, we call the optimal policy maximizing $\mathcal{W}$ subject to (2), (3) and (4), the **second-best** policy.

Before proceeding, we observe that it never pays the designer to lie if the news actually arrives.

LEMMA 1. *It is optimal for the designer to disclose the good news immediately. That is, an optimal policy has $\gamma_t \equiv 1$.*

Proof. If one increases γ_t, it can only increase the value of objective $\mathcal{W}$ and relax (4) (and does 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 $\gamma_t \equiv 1$ and focus on α as the sole policy instrument.

4 Optimal Recommendation Policy

We begin by further characterizing the process by which the designer's posterior, and the agents' beliefs about the 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." It then follows that (2) can be restated as:

$$\dot{\ell}_t = -\ell_t \lambda \mu_t, \quad \ell_0 = \ell^0 := \frac{p^0}{1-p^0}. \tag{5}$$

The agents' belief g_t concerning the arrival of news is determined via (3):

$$g_t = \frac{\ell^0 - \ell_t}{1-p^0}. \tag{6}$$

Immediately following the release of the product, news is very unlikely to have arrived, and hence $g_t \simeq 0$. As time elapses, the agents find it increasingly probable that news has arrived.

To see how this affects the designer's ability to "spam," substitute g_t into (4) to obtain:

$$\alpha_t \leq \bar{\alpha}(\ell_t) := \min\left\{1, \frac{\ell^0 - \ell_t}{k - \ell_t}\right\}, \tag{7}$$

if the normalized cost $k := c/(1-c)$ exceeds ℓ_t and $\bar{\alpha}(\ell_t) := 1$ otherwise. It is convenient to interpret $\bar{\alpha}(\ell_t)$ as the designer's *capacity* for spamming. This capacity depends on the prior p^0. If $p^0 \geq c$, then the designer can "freeze" the agents' belief at the prior p^0 by always spamming, and the prior is high enough to convince the agents to consume the product. Hence, $\bar{\alpha}(\ell_t) = 1$ for all ℓ_t. By contrast, if $p^0 < c$, then the spamming capacity is initially zero and increases gradually over time. Immediately after the product's release, the designer has nearly no ability to spam because good news could never have arrived instantaneously, and the agents' prior is unfavorable. Over time, however, $\bar{\alpha}(\ell_t)$ increases. In other words, even when no news is received, and ℓ_t falls as a result, the arrival of good news becomes increasingly probable, which allows the designer to develop her credibility over time and expands her capacity to spam.

Effectively, spamming "pools" recommendations across two very different circumstances: one in which the good news has arrived and one in which no news has arrived. Although the agents in the latter circumstance will never knowingly follow the recommendation, pooling the two circumstances for recommendation enables the designer to siphon the slack incentives from the former circumstance to the latter and to incentivize the agents to experiment, 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.

Substituting (6) into the objective function and using $\mu_t = \rho + \alpha_t$—and normalizing by ℓ—the *second-best* program is restated as follows:

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

subject to

$$\dot{\ell}_t = -\lambda(\rho + \alpha_t)\ell_t,\ \forall t,\ \text{and } \ell_0 = \ell^0, \tag{8}$$

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

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.

PROPOSITION 1. **(i)** *The* first-best *policy prescribes experimentation*

$$\alpha^{**}(p_t) = \begin{cases} 1 & \text{if } p_t \geq p^*; \\ 0 & \text{if } p_t < p^*, \end{cases}$$

where

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

and $v := \frac{1-c}{r}$ *denotes the continuation payoff upon the arrival of good news.*

(ii) *The* second-best *policy prescribes experimentation at*

$$\alpha^*(p_t) = \begin{cases} \bar{\alpha}(\frac{p_t}{1-p_t}) & \text{if } p_t \geq p^*; \\ 0 & \text{if } p_t < p^*. \end{cases}$$

(iii) *If* $p^0 \geq c$*, then the* second-best *policy implements the* first-best*, and if* $p^0 < c$*, then the* second-best *induces slower experimentation/learning than the* first-best*. Whenever* $p^0 > p^*$*, the* second-best *induces strictly more experimentation/learning than either* no social learning *or* full transparency.

The *first-best* and *second-best* policies have a cutoff structure: They induce maximal feasible experimentation, which equals 1 under *first-best* and equals the spamming capacity $\bar{\alpha}$ under the *second-best*, so long as the designer's posterior remains above the threshold level p^* and there is no experimentation otherwise. The optimal policies induce interesting learning trajectories, which are depicted in Figure 1 for the case of $p^0 < c$.

The optimality of a cutoff policy and the optimal level of the cutoff posterior can be explained by the main tradeoff facing the designer, namely the marginal benefit and cost of inducing additional experimentation at any given belief p:

$$\underbrace{\lambda pv\left(\frac{1}{(\lambda\rho/r)+1}\right)}_{\text{value of experimentation}} - \underbrace{c - p}_{\substack{\text{cost of} \\ \text{experimentation}}}. \tag{10}$$

The cost is the flow cost borne by the experimenting agents (the second term). The benefit is the social learning that the additional experimentation may generate (the first term): With probability p, the product is good, and experimentation will reveal this information at rate λ, which will enable the future generation of agents to collect the benefit of $v = (1-c)/r$. The term $\frac{1}{(\lambda\rho/r)+1}$ discounts this benefit by the rate at which the good news will be learned by "background learning" even with no experimentation. The optimal threshold posterior, p^*, which equates the benefits and the costs, is the same under either *first-best* and *second-best*

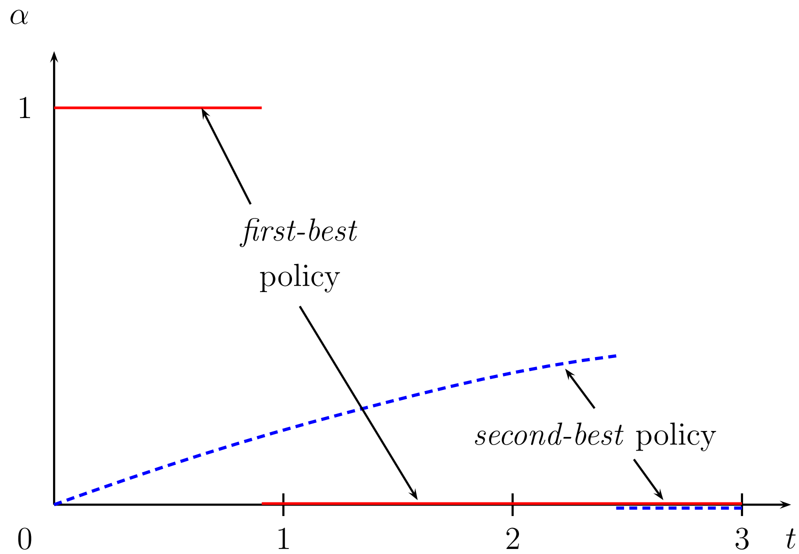

Figure 1: Path of α for $(c, \rho, p^0, r, \lambda) = (2/3, 1/4, 1/2, 1/10, 4/5)$.

because the opportunity cost of experimentation, *i.e.*, relying solely on background learning, is the same under both regimes.

If $p^0 \geq c$, the designer can implement the *first-best* policy by simply "spamming" *all* agents if and only if full experimentation is warranted under *first-best*. The agents comply with the recommendation because their belief is frozen at $p^0 \geq c$ under the policy. Admittedly, informational externalities are not particularly severe in this case because early agents will have an incentive to consume on their own. Note, however, that *full transparency* does not implement the *first-best* in this case, since agents will stop experimenting once p_t reaches c. In other words, spamming—or "obfuscation"—is crucial to achieving the *first-best*, even in this case.

In the more interesting case with $p^0 < c$, the *second-best* policy cannot implement the *first-best*. In this case, the spamming constraint for the designer is binding. As can be seen in Figure 1, spamming capacity is initially zero and increases gradually. Consequently, experimentation initially takes off very slowly and builds up gradually over time until the posterior reaches the threshold p^*, at which point the designer abandons experimentation. Throughout, the experimentation rate remains strictly below 1. In other words, learning is always slower under the *second-best* than under the *first-best*. Since the threshold belief is the same under both regimes, the agents are induced to experiment longer under the *second-best* than under the *first-best* regime, as Figure 1 shows.

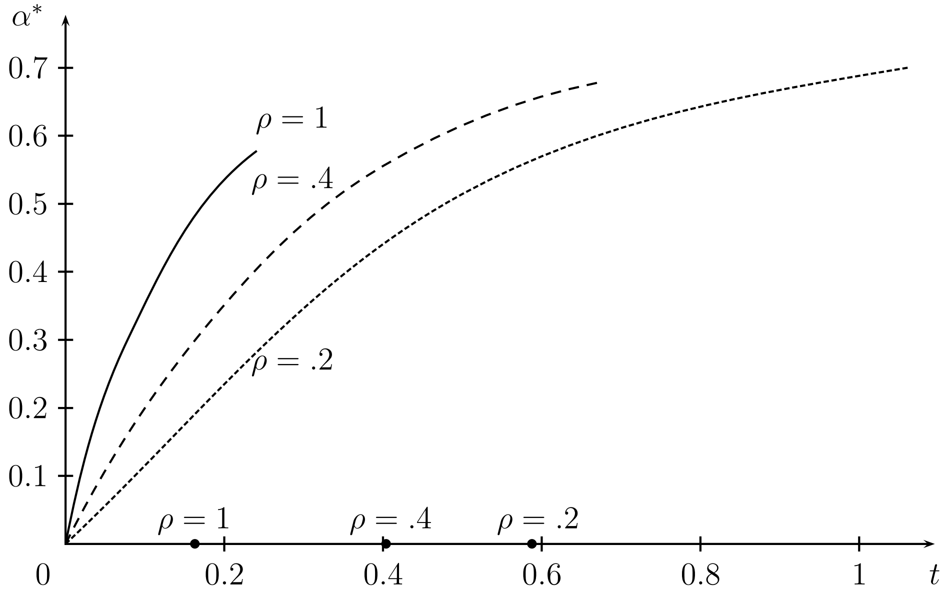

Figure 2: Optimal (*second-best*) spamming as a function of ρ (here, $(k, \ell_0, r, \lambda) = (2/5, 1/3, 1/2, 1)$). The dots on the x-axis indicate stopping times under *first-best*.

In either case, as long $p^0 > p^*$, the *second-best* policy implements strictly higher experimentation/learning than either *no social learning* or *full transparency*, strictly dominating both of these benchmarks.

Comparative statics reveal further implications. The values of (p^0, ρ) parameterize the severity of the cold start problem facing the designer. The lower these values, the more severe the cold start problems are. One can see how these parameters affect optimal experimentation policies and induced social learning.

Corollary 1. **(i)** *As p^0 rises, the optimal threshold remains unchanged and total experimentation/social learning increases under both the* first-best *and* second-best *policies. The learning speed remains the same in the* first-best *policy but rises in the* second-best *policy.*

(ii) *As ρ rises, the optimal threshold p^* rises and total experimentation/social learning declines under both the* first-best *and the* second-best *policies. The learning speed remains the same in the* first-best *policy but rises in the* second-best *policy.*

Unlike the *first-best* policy, the severity of the cold start problem affects the flow rate of experimentation under the *second-best* policy. Specifically, the more severe the cold start problem is, in the sense of (p^0, ρ) being smaller, the more difficult it becomes for the designer to credibly spam the agents, thereby reducing the flow rate of experimentation that the designer can induce. This has certain policy implications. Internet recommenders such as

Netflix and *Amazon* have the ability to increase ρ by investing resources in product research. Corollary 1-(ii) shows the sense in which such an investment "substitutes" for agents' experimentation: an increased free learning raises the opportunity cost of experimentation, calling for its termination at a higher threshold, under both *first-best* and *second-best* policies. In the *second-best* world, however, there is also a sense in which this investment "complements" experimentation: free learning makes spamming credible, and this allows the designer to induce a higher level of experimentation at each t. Figure 2 shows that this indirect effect can accelerate the social learning significantly: as ρ rises, the time it takes to reach the threshold belief is reduced much more dramatically under the *second-best* policy than under the *first-best*.

5 Public Recommendations

Thus far, we have considered private recommendations that may differ across agents and are not shared amongst them. Such personalized private recommendations are an important part of the Internet recommender system; *Netflix* and *Amazon* make personalized recommendations based on its users' past purchase/usage histories. Search engines are known to rank search items differently across users based on their past search behavior. At the same time, platforms also make public recommendations that are commonly observable to all users. Product ratings provided by *Amazon, Netflix, Yelp, Michelin*, and *Parker* are public and unpersonalized, to our knowledge.

Personalized recommendations benefit consumers when they have heterogenous tastes, but do personalized recommendations matter even when the users have the same preferences, as assumed herein? In particular, do they play any role in incentivizing experimentation? We answer this question by studying the (optimal) public recommendation policy. Plainly, any public recommendation can be made privately without any loss. Hence, a public recommendation cannot be strictly better than private recommendations. It is not obvious, however, that private recommendations are strictly better than public recommendations. Our main result is that although the designer can still induce agents to experiment under public recommendation, the experimentation is random and slower (on average) and thus welfare is strictly lower than under the optimal (private) recommendation.

To begin, consider the baseline model with $p^0 < c$. Recall *full transparency* induces no experimentation in this case. We first show that a public recommendation policy can induce some experimentation and improve upon *full transparency*. Consider a policy with the following structure:

- At $t = 0$, the designer begins by making no recommendation, which means that the only source of learning is free learning at rate ρ.

– If good news has arrived by $t > 0$, then the designer recommends the product and does so forever.

– If good news has not arrived by time $t > 0$ and the designer's posterior is above p^* (our *first-best/second-best* cutoff), then the designer "spams" (or recommends the product without good news) according to a distribution function $F(t)$ (to be defined). Once she recommends the product, she continues to recommend until her posterior reaches p^*.

– If good news has not arrived by t and the posterior is below p^*, then the designer recommends that the agents do not consume.

Proposition 2 below demonstrates that the optimal public recommendation policy has this structure. As with private recommendations, the designer reveals the good news when it arrives, but the way she "spams" is different. Given the public nature, the designer must effectively recommend to "all" agents when she does so at all. After the recommendation is made successfully, meaning that all agents choose to consume, the designer can "freeze" the agents' posterior by simply continuing to recommend the product, which allows the designer to implement the *first-best* policy from that point forward. Of course, the first time that a public recommendation is made cannot be arbitrary. For instance, if the designer begins the recommendation at time 0 with certainty (which is what she would like to do), then the agents' posterior will be p^0, and they will thus not follow the recommendation. The designer instead randomizes the time at which she begins spamming on the product.

We now study how the random spamming distribution $F(t)$ can be constructed to satisfy incentive compatibility. For ease of exposition, we assume $F(t)$ to be atomless. (This will be justified formally in the proof of Proposition 2 in the Appendix.) Let $h(t) := f(t)/(1-F(t))$ denote the hazard rate; *i.e.*, $h_t \mathrm{d}t$ is the probability that the designer spams for the first time during $[t, t + \mathrm{d}t)$. Then, by Bayes' rule, the posterior of an agent who receives the public recommendation to consume for the first time at t is given by:

$$P_t = \frac{p^0 e^{-\lambda\rho t}(\lambda\rho + h(t))}{p^0 e^{-\lambda\rho t}(\lambda\rho + h(t)) + (1-p^0)h(t)}.$$

The incentive constraint requires that $P_t \geq c$, which in turn yields:

$$h(t) \leq \frac{\lambda\rho\ell^0}{k - e^{\lambda\rho t} - \ell^0}.$$

It is intuitive (and formally shown) that the incentive constraint is binding at the optimal policy, which gives rise to a differential equation for F, alongside the boundary condition

$F(0) = 0$. Its unique solution is

$$F(t) = \frac{\ell^0(1 - e^{-\lambda\rho t})}{k - \ell^0 e^{-\lambda\rho t}}, \tag{11}$$

for all $t < t^* := -\frac{1}{\lambda\rho}\ln(\frac{k}{\ell^0}\frac{\lambda\rho+r}{\lambda(1+\rho)+r})$, the time at which the designer's posterior reaches the threshold belief p^*.

Clearly, $F(t) > 0$, for any $t \in (0, t^*)$, so the designer spams and the agents experiment with positive probability. Meanwhile, $F(t^*) = \frac{\ell^0-\ell^*}{k-\ell^*} < 1$, and hence there is a positive probability that the designer never spam—a property crucial for the designer to maintain her credibility. Learning under this policy can be either faster or slower depending on the realization of the random triggering of spam, but it is on average slower under public recommendation than under the optimal private recommendation. To see this more clearly, fix any $t \leq t^*$. Under optimal public recommendation, spam is triggered at s according to $F(s)$ and lasts until t, unless the posterior reaches p^*. Let $T(s)$ be the time at which the latter event occurs if spam was triggered at s. Then, the expected level of experimentation performed by time t under public experimentation is:

$$\begin{aligned}\int_0^t (\min\{t, T(s)\} - s)\mathrm{d}F(s) &\leq \int_0^t (t-s)\mathrm{d}F(s) \\ &= \int_0^t F(s)\mathrm{d}s = \int_0^t \frac{\ell^0 - \ell^0 e^{-\lambda\rho s}}{k - \ell^0 e^{-\lambda\rho s}}\mathrm{d}s < \int_0^t \frac{\ell^0 - \ell_s}{k - \ell_s}\mathrm{d}s = \int_0^t \bar{\alpha}(\ell_s)\mathrm{d}s,\end{aligned}$$

where ℓ_s is the likelihood ratio at time s under the optimal private recommendation. The first equality follows from integration by parts, and the inequality holds because $\ell_s = \ell^0 e^{-\lambda\int_0^s(\bar{\alpha}(\ell_{s'})+\rho)\mathrm{d}s'} < \ell^0 e^{-\lambda\rho s}$.

We now state our findings.

Proposition 2. *Under the optimal public recommendation policy, the designer recommends the product at time t if good news is received by that time. If good news is not received and a recommendation is not made by t, she triggers spam at a random time given by $F(t)$ in (11) and the spam lasts until t^* in the event that no good news arrives by that time. The experimentation speeds up until t^* on each realization of the policy (and thus on average). The induced experimentation under optimal public recommendation is on average slower—and the welfare attained is strictly lower—than under optimal private recommendation.*

Proof. The proof in the Appendix formulates the mechanism design problem fully generally, allowing the designer to communicate arbitrary messages upon receiving good news and no news. Although the argument is involved, the key observation is that once the agents are induced to consume the product, the designer can "freeze" the posteriors of the agents arriving subsequently and induce them to continue, which means that the *first-best* policy can be implemented following any history that credibly convinces the agents to consume the product. This fact allows us to formulate the optimal mechanism design problem as a stopping problem that is indexed by two stopping times, one specifying how soon after good news is received that the designer should recommend the product to the agents and one specifying how soon she should recommend it to the agents despite having received no news. The unusual feature of the problem is that the the incentive constraint must be obeyed, which forces the optimal stopping time (*i.e.*, triggering spam) to be random. We show that the above public recommendation policy solves this general mechanism design problem. $\square$

6 Endogenous Entry

Our model has thus far considered short-lived agents whose only decision is *whether* to consume a product given a recommendation. In practice, consumers also choose *when* to consume. In particular, they may delay their consumption decision until after more accurate recommendations become available. A recommender facing such patient consumers must guide consumers not only toward the right experimentation decision—and must make the recommendation persuasive in the sense discussed above—but also toward the right time for doing so. In particular, overcoming a possible free-rider problem becomes a new challenge.

To study this issue, we extend our baseline model in Section 4 as follows. As in our baseline model, agents arrive at a unit flow rate in each period, but upon arrival, say at t, an agent is free to choose any time $\tau \geq t$ to "check in" for a recommendation. In principle, agents may check in multiple times to obtain an improved recommendation, but such an incentive can be muted by ensuring that an agent receives the same recommendation each time he checks in, which is possible under personalized/private recommendation.[20] Agents would then never delay their consumption decision after checking in. As in Section 3, we focus on private recommendation.

The designer now controls two instruments: the mass X_t of agents the designer induces to check in by time t and the fraction α_t of agents to whom the designer spams among those who check in at t. Specifically, the designer faces the following problem:

$$[SB'] \qquad \sup_{\{\alpha_t, X_t\}} \int_{t\geq 0} e^{-rt} \left(\ell^0 - \ell_t - \alpha_t \left(k - \ell_t\right)\right) \mathrm{d}X_t$$

[20]For instance, an Internet recommender may program the same recommendation to appear at its Internet portal for any user who logs in with the same membership ID or the same IP address. Although motivated users may circumvent the system at some inconvenience, the stakes or benefits from this behavior are often not large enough for it to be worthwhile. However, multiple check-ins might be potentially beneficial from a welfare perspective because agents may reduce their wait time after good news has arrived; for instance, the designer may have the agents "continuously" check in for news. In practice, checking in for a recommendation can be costly. We assume that this cost is high enough for multiple check-ins to be undesirable.

subject to (7),

$$\dot{\ell}_t = -\lambda(\rho + \alpha_t \mathrm{d}X_t)\ell_t,\ \forall t,\ \text{ and } \ell_0 = \ell^0, \tag{12}$$

$$X_t \leq t,\ \forall t \tag{13}$$

$$t \in \arg\max_{t' \geq t} \left(\alpha_t \left(\ell_t - k \right) - \ell_t \right),\ \forall t \in supp(X). \tag{14}$$

The problem $[SB']$ has a similar objective function to $[SB]$, and involves the incentive constraint (7) and the law of motion (12), much as in $[SB]$. The difference is that the designer now controls the agents' check-in time according to a measure X_t. She does so subject to two constraints: (13) ensures that the the mass of agents who check in by time t cannot exceed the mass who have arrived by that time, and (3) requires that agents should have no incentive to delay their check-in beyond that desired by the designer.[21]

The following proposition characterizes the optimal recommendation policy and *full transparency* in the presence of endogenous entry.

Proposition 3. *Suppose that agents can delay checking in for a recommendation, and suppose that $\ell^0 < k$.*

(i) *Under* full transparency*, agents who arrive before $t^{FT} = -\frac{1}{\rho\lambda} \ln \frac{r}{r+\lambda\rho}$ all check in at t^{FT}, and those who arrive after t^{FT} check in as they arrive. Upon checking in, agents consume if and only if news has arrived by that time. In other words, no agent experiments under* full transparency*.*

(ii) *The* second-best *policy coincides with* full transparency *if ρ is sufficiently large for any p. The* second-best *policy differs from* full transparency *if $\rho \leq \frac{1}{\lambda}\left(\frac{r\ell^0 + \sqrt{r\ell^0}\sqrt{4k\lambda + \ell^0}}{2k} - r\right)$.*

(iii) *In case the* second-best *differs from* full transparency*, there exist $t^* > 0$ and $T > t^*$ such that the optimal mechanism induces the following behavior from the agents:*[22]

a. *(Blackout followed by a burst of spam) The designer offers no recommendation (even with good news) until t^*, which induces the agents who arrive before t^* to*

[21] Here, $supp(X)$ is support of X. More precisely,

$$supp(X) := \{t \in [0, \infty) \mid X_{t+\epsilon} - X_{\max\{t-\epsilon, 0\}} > 0,\ \forall \epsilon > 0\}.$$

We ignore the incentive for the agents never to check in too soon relative to what the designer intends, since the designer can simply withhold all recommendations to prevent agents from checking in for a period of time.

[22] As before, as soon as good news arrives, it is immediately shared with the agents, including during the process of atom split, which is described in a.

check in and experiment at t^* but sequentially with probability $\alpha(\ell)$ that increases along the locus $\alpha(\ell) := \frac{\ell^+ + \alpha_2(k-\ell^+)-\ell}{k-\ell}$, for some (α^+, ℓ^+) with $\ell^+ < \ell^{FT} := \frac{r\ell^0}{r+\lambda\rho}$.

b. *(Smoothly tapered spam) The agents who arrive after t^* check in as they arrive and experiment with probability α_t which declines over time and reaches 0 at $t = T$. The agents who arrive after T never experiment.*

Proof. The proof is available in the Appendix and the Supplementary Material. □

The intuition behind the results can be explained as follows. There are two conflicting welfare considerations involved in an agent's timing of check-in.

First, there is a social benefit from delaying agents' check-in time. By delaying the time at which agents obtain a recommendation, the designer allows an agent to make a decision based on more updated information. This benefit exists even without any experimentation because of free learning but declines over time due to discounting. Therefore, absent experimentation, there is an interior time $t^{FT} > 0$ such that the designer would seek to delay agents' check-in up to that time. This is precisely what happens under *full transparency*, which is socially optimal, assuming that no agent is induced to experiment.

The second consideration arises when the designer wishes to induce positive experimentation. Since experimentation is costly for agents who experiment and beneficial for later agents who do not experiment, the former have incentives to delay their checking in to free ride on other experimenters (even when delaying is not socially beneficial). This incentive problem can be seen clearly by the *second-best* policy in Section 4. In Figure 1, all agents who arrive before $t \simeq 2.4$ are told to experiment at a rate that leaves them with zero expected payoff, whereas those arriving slightly after 2.4 enjoy a strictly positive payoff; thus, all the former agents would wait until after that time. To overcome this free-rider problem, the designer must "bribe" the experimenting agents by reducing their experimentation and reducing the externalities enjoyed by later agents by prolonging experimentation. In particular, experimentation cannot fall discontinuously as in our baseline case, and all experimenting agents must enjoy strictly positive payoffs, which means that incentive constraint (7) is no longer binding.

Both of these considerations reduce the value of experimentation, particularly when ρ is large (significant free learning). Indeed, for a sufficiently large ρ, the *second-best* policy is reduced to *full transparency*. When ρ is sufficiently low, however, *full transparency* is again not optimal (part (ii) of Proposition 3). In this case, the *second-best* policy involves "spamming"—much as in our baseline model—but with qualitatively new features. Figure 3 depicts how experimentation and belief evolve over time under the *second-best* policy. As the belief ℓ_t declines over time, the relevant outcome (α_t, ℓ_t) moves from right to left in the figure.

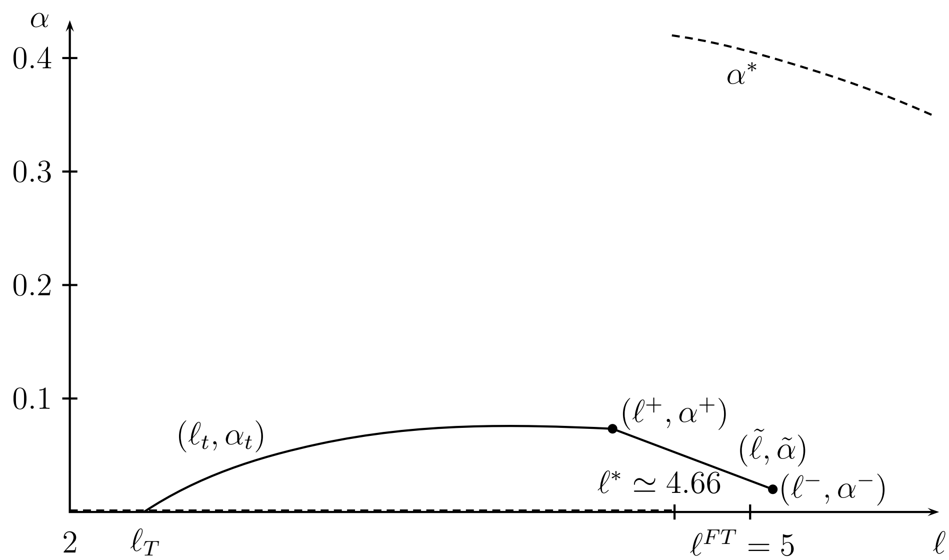

Figure 3: The probability of a recommendation as a function of ℓ (here, $(r, \lambda, \rho, k, \ell^0) = (1/2, 2, 1/5, 15, 9)$). In dotted line, the spamming probability α^* in the baseline model with exogenous entry, as a function of ℓ.

First, the agents who arrive before $t^* > 0$ are induced to delay their checking in for recommendation until t^*. This can be implemented by an information "blackout"—or the recommender refusing to recommend a product—for a duration of time after its release. This can be interpreted in practical terms as the recommender refusing to take a stance on the product. Given the initially unfavorable prior, all agents will then choose to wait during the blackout.

At time t^*, it is optimal for the agents to all check in at that moment; however, it is not optimal to have them experiment simultaneously. If the designer were to "split" the mass into two smaller masses and have one of them move immediately after the other, the former will benefit from the experimentation of the latter, and can be asked to experiment more without violating their incentives. Repeating this argument, one finds it optimal to divide the mass of agents into flows of them who check in sequentially—but without taking any real time—and to experiment with increasingly higher probability, in a manner that leaves them indifferent across this menu of $(\tilde{\ell}, \tilde{\alpha})$ offered sequentially at that instant.[23] Doing so

[23]The need to allow for sequential moves that do not take any real time arises from the feature of the continuous time: the soonest next time after any time t is not well defined. The Supplementary Material resolves this issue by enriching the policy space, specifically by adding a "virtual" clock that is used to sequence the check-in of agents who arrive at the same "real" time. See Simon and Stinchcombe (1989)

sequentially is optimal because it increases the overall experimentation that can be performed with the agents; Indifference among these agents gives rise to a locus $(\tilde{\ell}, \tilde{\alpha}) = (\tilde{\ell}, \alpha(\tilde{\ell}))$ that begins at (α^-, ℓ^-) and ends at (α^+, ℓ^+), as depicted in Figure 3. The mass of agents are all dissipated at a point (ℓ^+, α^+).

After t^*, the experimentation tapers off along a smooth locus, as depicted in Figure 3. Specifically, these agents check in as they arrive, and they experiment with probability that declines continuously at a rate that eliminates incentives for delaying entry. At T, the experimentation stops, and *full transparency* prevails from this point forward.

In sum, the optimal policy under endogenous entry exhibits qualitatively the same features as that from the baseline model (such as spamming and single-peaked experimentation), but it also introduces new features such as an information blackout and a gradual phase-out of experimentation. Endogenous entry also affects the amount of experimentation performed by the agents. Under exogenous entry, the agents who are induced to experiment do so at the level that leaves them with no rents. By contrast, under endogenous entry, the designer must leave them with positive rents to prevent them from free-riding on others. This means that, for each $\ell > \ell^*$, agents experiment less under endogenous entry than under exogenous entry; see the optimal experimentation α^* induced under the latter regime for comparison in Figure 3.[24] Meanwhile, agents are induced to experiment even when $\ell < \ell^*$ with endogenous entry (whereas no more experimentation occurs when $\ell < \ell^*$ with exogenous entry). This feature is in turn attributed to the free-riding problem. Inducing agents to experiment when $\ell < \ell^*$ entails net social loss (as was shown before); yet this distortion is necessary to prevent early agents who are induced to experiment when $\ell > \ell^*$ from delaying their check-in. In this sense, the distortion at a low belief is what makes experimentation at a higher belief implementable.[25]

---

and Ausubel (2004) for adopting similar enrichment of strategies to guarantee existence of an equilibrium. Intuitively, we can think of the optimal policy in the enriched space as the supremum over policies in which the designer can only provide one recommendation at each time t, or the limit of the optimal outcomes of the sequence of discrete time models as its time interval shrinks to zero. In such a model, the designer will divide the mass into an increasing number of smaller masses as the time interval shrinks to zero, but with total real time approaching zero in the limit. Note, however, that our optimal policy is a well-defined maximum if the optimal control problem is written with ℓ as the "time" index.

[24]This does not imply that learning is necessarily slower here in the temporal sense than under exogenous entry. The agents who arrive before t^* delay their check-in, and the free learning up to that point can motivate them to experiment more at $t \geq t^*$. This effect may offset the otherwise reduced incentive for experimentation.

[25]This feature is reminiscent of optimal second-degree price discrimination wherein a seller distorts a low-value buyer's purchase to extract rents from a high-value buyer.

7 Extensions

The model can be further extended to incorporate additional features. The detailed analysis is provided in the Supplementary Material; here, we illustrate the main ideas and results.

7.1 General Signal Structure

Thus far, our model assumed a simple signal structure which features only good news. This is a reasonable assumption for many products whose priors are initially unfavorable but can be improved dramatically through social learning. For some other products, however, social learning may involve discovery of poor quality. Our signal structure can be extended to allow for such a situation via "bad" news.[26] Specifically, news can be either good or bad, where good news means $\omega = 1$ and bad news reveals $\omega = 0$, and the arrival rates of the good news and bad news are respectively $\lambda_g > 0$ and $\lambda_b > 0$ conditional on the state.[27]

The relative arrival rate $\Delta := \frac{\lambda_g - \lambda_b}{\lambda_g}$ of good news proves crucial for both the evolution of the designer's belief and the optimal recommendation policy. Thus, we say that the signal structure is "good" news if $\Delta > 0$ and "bad" news if $\Delta < 0$. The posterior of the designer in the absence of any news evolves according to:

$$\dot{\ell}_t = -\ell_t \Delta \lambda_g \mu_t, \quad \ell_0 = \ell^0. \tag{15}$$

Intuitively, the belief becomes pessimistic in the good news environment and optimistic in the bad news environment as time progresses with no news.

As before, the designer reveals both good news and bad news truthfully. The more important question is whether the designer would recommend the product despite having no news. Here, the main thrust of our earlier analysis continues to hold. That is, the designer would like to spam the agents up to the maximum capacity allowed by incentive compatibility as long as her posterior remains above some threshold. In particular, the optimal policy in the good news case is virtually identical to that described in Proposition 1, in fact with λ in the optimal threshold p^* being replaced by λ_g.[28]

[26] See Keller and Rady (2015) for the standard bad news model of strategic experimentation.

[27] More precisely, if a flow of size μ consumes the product over some time interval $[t, t + \mathrm{d}t)$, then during this time interval the designer learns that the movie is "good" with probability $\lambda_g(\rho + \mu)\mathrm{d}t$ and "bad" with probability $\lambda_b(\rho + \mu)\mathrm{d}t$.

[28] At first glance, it may be surprising that the arrival rate λ_b of the bad news does not affect the optimal threshold p^* in the case of good news, and vice versa. The reason is that the tradeoff captured in (10) does not depend on the arrival rate of bad news. At the same time, the arrival rate of bad news affects both the duration and the rate of incentive-compatible experimentation. As (15) shows, 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, which means that the agents are induced to experiment for longer (until news arrives or the threshold

The optimal recommendation policy in the bad news environment has the same cutoff structure, although the threshold posterior is somewhat different. The formal result, the proof of which is available in the Supplementary Material, is as follows:

Proposition 4. *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 experimentation at the maximum incentive-compatible level thereafter for any $p > p_b^*$,*[29] *where $p_b^* > p_b^{**}$. In other words, the* second-best *policy triggers experimentation at a later date and at a lower rate than the* first-best *policy.*

Although the structure of the optimal recommendation policy is similar between the good news and bad news cases, the intertemporal trajectory of experimentation is quite different. Figure 4 depicts an example with $\Delta < 0$ and a sufficiently low prior belief. Initially, the designer finds the prior to be too low to trigger recommendation, and she never spams as a result. However, as time progresses without receiving any news (good or bad), her belief improves gradually, and as her posterior reaches the optimal threshold, she begins spamming at the maximal capacity allowed by incentive compatibility. One difference here is that the optimal *second-best* threshold differs from that of the *first-best*. The designer has a higher threshold, so she waits longer to trigger experimentation under the *second-best* policy than she would under the *first-best* policy. This is due to the difference in the tradeoffs at the margin between the two regimes. Although the benefit from not triggering experimentation is the same between the two regimes, the benefit from triggering experimentation is lower in the *second-best* regime due to the constrained experimentation that follows in the regime.

7.2 Heterogenous Costs

Thus far, agents have been assumed to be homogenous, which has greatly simplified the analysis and made the question of the observability of the agents' costs moot. To discuss

p^* is reached), holding constant the per-period experimentation rate. Furthermore, the spamming capacity can be seen to increase with λ_b, as the commitment never to recommend in the event of bad news means that a recommendation is more likely to have been the result of good news. Hence, experimentation increases in two different senses when λ_b rises.

[29] The maximally incentive-compatible level is $\bar{\alpha}(\ell_t) := \min\left\{1, \frac{\left(\frac{\ell_t}{\ell_0}\right)^{-\frac{1}{\Delta}} - 1}{k - \ell_t}\ell_t\right\}$.

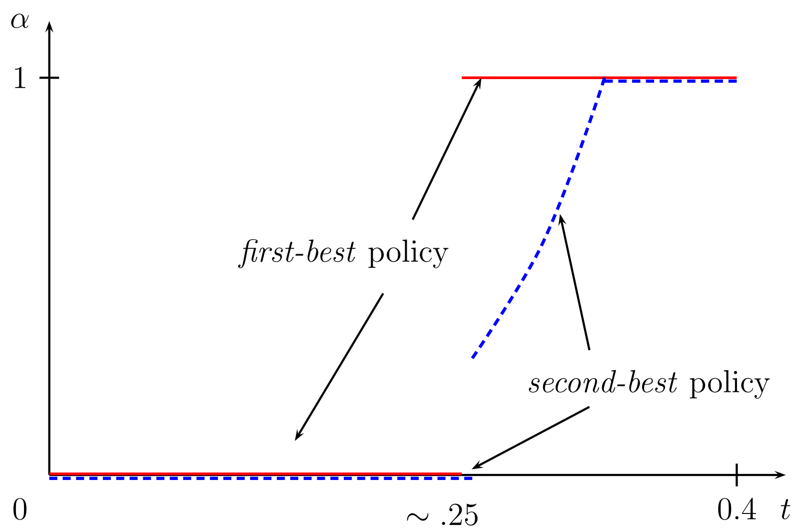

Figure 4: Path of α for $\Delta < 0$ and $(c, \rho, p^0, r, \lambda_g, \lambda_b) = (1/2, 1, 2/7, 1/10, 1, 2)$).

how our findings generalize, consider first a simple two-type example. Suppose that there are two cost types, c_L and c_H, with $c_H > c_L > p^0$, arising with probabilities q_L and $q_H = 1 - q_L$, respectively. For concreteness, we assume that

$$c_H/(1-c_H) = 12/10, c_L/(1-c_L) = 11/10, p^0 = 1/2, \text{ and } q_H = q_L = 1/2.$$

Furthermore, $\rho = r = \lambda = 1$. We contrast two scenarios. In the first, the types are observable. In the second, they are not.

7.2.1 Observable types

Here, we assume that the designer perfectly observes or infers the type of the agent. In practice, such an inference may be possible from a user's past consumption history. For instance, the frequencies of downloading or streaming movies may indicate a user's (opportunity) cost of experimentation, and the online platform may use that information to tailor its recommendation to the user. With observable costs, the designer can spam an agent with a maximum probability that depends on his type, namely,

$$\bar{\alpha}_i(\ell_t) := \min\left\{1, \frac{\ell^0 - \ell_t}{k_i - \ell_t}\right\},$$

for $i = L, H$, where $k_i := c_i/(1-c_i)$.

The optimal policies, described fully in the Supplementary Material, are characterized by two cutoffs, $p_2 < p_1 < c$, in the designer's belief such that the designer induces both types of agents to experiment if $p > p_1$, only low-cost type agents to experiment if $p \in (p_2, p_1]$, and no agents to experiment if $p < p_2$. The agents induced to experiment do so up to the maximal probability, which is one under the *first-best* regime and $\bar{\alpha}_i(\ell)$ for type $i = L, H$ under the *second-best* regime.

As in Section 4, the last threshold is equal to

$$p_2 = p_2^{**} = c_L \left(1 - \frac{rv}{\rho + r(v + 1/\lambda)}\right),$$

under both *first-best* and *second-best* policies. In the current example, $p_2^{**} \simeq .425$. The first threshold differs between under the two regimes. In our example, the threshold is $p_1^{**} = 0.451$ under the *first-best* policy and $p_1^* = .448$ under the *second best* policy. In other words, the high-cost agents are induced to experiment longer (in terms of the belief) under the *second-best* policy than under the *first-best* policy. This is attributed to the differing consequences of triggering the phase in which only low-cost agents experiment. In the *first-best* case, the designer can rely on all low-cost agents to experiment from that point on; in the *second-best*, she can only credibly ask a fraction of them to experiment. Her alternative channel for experimentation appears bleaker, and as a result, she is forced to ask high-cost agents to experiment for lower beliefs than in the *first-best* case. As before, since experimentation occurs at a lower rate under the *second-best* policy, it takes longer to reach each belief threshold: it takes $t_1^* \simeq 1.15$ and $t_2^* \simeq 1.22$ to reach the first and second threshold under *second-best*, whereas it takes $t_1^{**} \simeq .1$ and $t_2^{**} \simeq .17$ under the *first-best* policy.

This feature is not specific to the particular numerical example we picked. In the Supplementary Material, we prove that, for the case of good news, and two types, the high-cost agents are spammed at beliefs below the point at which they should stop under the *first-best* policy.

PROPOSITION 5. *Both the* first-best *and* second-best *policies are characterized by a pair of thresholds* $0 \leq \ell_L \leq \ell_H \leq \ell^0$*, such that (i) all agents are asked to experiment with maximum probability for* $\ell \geq \ell_H$*, (ii) only low-cost agents experiment (with maximum probability) for* $\ell \in [\ell_L, \ell_H)$*, and (iii) no agent experiments for* $\ell < \ell_L$*. Furthermore, the belief thresholds are* $\ell_L^{**} = \ell_L^*$ *and* $\ell_H^{**} \geq \ell_H^*$ *under* first-best *and* second-best *policies, respectively, with a strict inequality whenever* $\ell^0 > \ell_H^{**}$.

Alternatively, one could allow for a continuum of types. Although we are unable to solve for the optimal policy with full generality, it is possible to solve for the limit policy as $r \to 0$ in the case of the uniform distribution (and pure good news). Figure 5 depicts the optimal policies for some parameter values. As time passes, the belief decreases, as does the threshold cost type above which it is optimal for the designer not to spam (highest curve in Figure 5). Types below this threshold are spammed to different extents: those whose cost

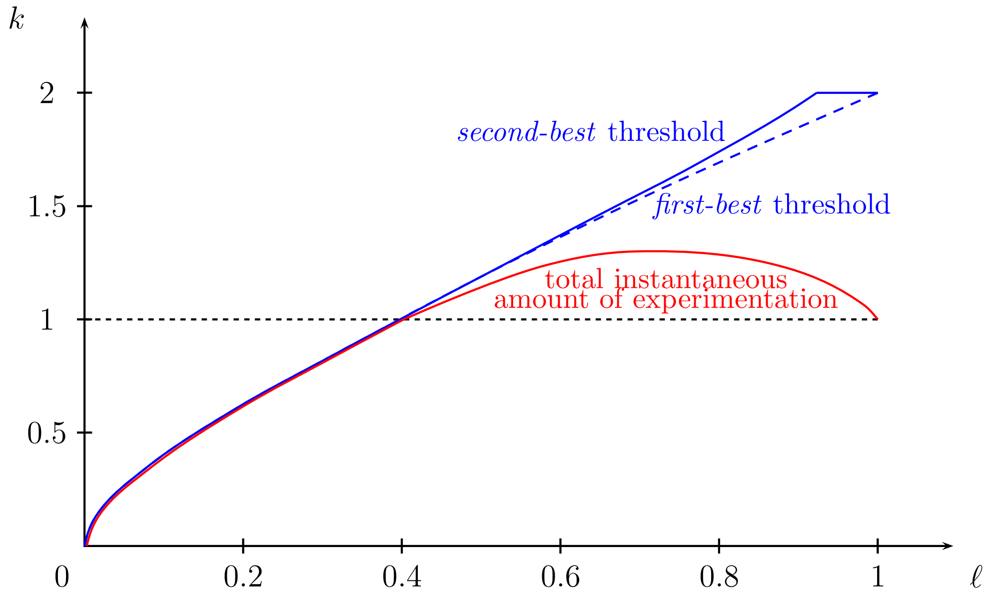

Figure 5: Optimal policy with a continuum of cost (here, $p^0 = 1/2, k \sim U[0, 2]$).

is below the prior ($c \leq p^0 = 0.5$) are spammed for sure, while those whose cost is above the prior are spammed to the maximum extent that makes them still willing to follow the recommendation. Hence, the total rate of experimentation is lower than the highest curve and equal to the lower curve: at the initial instant, agents with $c \leq p^0$ are "fully" spammed whereas agents with higher costs cannot be spammed at all. As time passes and the belief decreases, higher cost agents can be spammed (although fewer of them are targeted), and experimentation increases. Eventually, however, the designer focuses on spamming only lower and lower types only, leading to a decline in rate of experimentation. When this threshold coincides with the cost type that can be spammed with a probability of one (the type with $c = .5$), the *first-best* policy can be followed. Until then, however, the *second-best* threshold is above the *first-best* threshold, as indicated by the dotted line: a given cost type (such that $c > .5$) is spammed at beliefs below those at which he would no longer be asked to experiment in the *first-best*.[30]

7.2.2 Unobservable Costs

If the designer cannot infer agents' costs, then her ability to tailor recommendations is severely limited. Specifically, consider our two-type example but assume that the type is unobservable to the agent. Because the designer cannot use transfers to elicit truth-telling, and because all types of agents prefer a more honest recommendation, the designer cannot

[30] Keep in mind, however, that under the *first-best* policy, types below the dotted threshold are experimenting with probability 1, while under the *second-best* policy, this is not the case for cost types $c \geq 0.5$.

effectively screen agents' types. In fact, the designer can only choose one probability of spamming α and has effectively three options regarding its value at any given point in time:[31]

1. No spamming: The designer chooses $\alpha = 0$.

2. Spamming only to the low-cost agents: The designer spams to a fraction $\alpha = \bar{\alpha}_L$ of agents. A low-cost agent follows that recommendation, while a high-cost agent does not. Thus, this means, in particular, that at any such time a high-cost agent will not consume even when the designer sends a genuine recommendation after good news.

3. Spamming to both types of agents: The designer spams to fraction $\alpha = \bar{\alpha}_H$ of agents. Both types of agents then follow the recommendation, but since $\bar{\alpha}_H < \bar{\alpha}_L$, experimentation will proceed at a relatively slow rate—although it will be spread across all agents, as opposed to only the low-cost agents.

Intuitively, the designer triggers the options in the following order: First, she spams to both types, then only to the low type, and finally to none. It turns out that in this example the designer skips the first phase and begins with $\alpha = \bar{\alpha}_L$ incentivizing only the low-cost type to experiment. She does that until her belief reaches $p_2 := .468$, which occurs at time $t_2 = .109$, assuming that no news has been received by that time. Compared to the observable case, the designer stops experimenting much sooner (.109 vs. 1.22!), and at a higher belief (.468 vs. .425) than when types are observable. Furthermore, she stops spamming high-cost agents altogether—or rather, they do not listen to the designer's recommendation and will not consume.

Clearly, experimentation is much costlier when types are unobservable. Given spamming at $\bar{\alpha}_H$, not all low-cost agents are induced to experiment and in fact some low-cost agents receive a "do not consume" recommendation while some high-cost agents receive a "consume" recommendation. This mismatch makes this option costly (making it suboptimal to use it it at least in our example). Meanwhile, spamming at rate $\bar{\alpha}_L$ is costly because the designer must then forego consumption by high-cost agents altogether—even if he has found out that the product is of high quality. This makes it more expensive to prolong experimentation with low-cost agents and results in a shorter experimentation spell than when types are observable.

In sum, unobservability makes spamming costly and pushes the designer toward greater transparency. In fact, as we show in the Supplementary Material, this push toward transparency is exacerbated with more types. Namely,

[31] While this claim sounds intuitive, it requires proof, which follows from the general result stated below.

Proposition 6. *When agents' costs are uniformly distributed and unobservable to the designer,* full transparency *is optimal.*[32]

8 Conclusion

Early experimentation is crucial for users to discover and adopt potentially valuable products on a large scale. The purpose of the present paper has been to understand how a recommendation policy can be designed to promote such early experimentation. There are several takeaways from the current study.

First, "spamming" on a product—namely, recommending a product that has yet to be found worthy of recommendation—can turn users' belief favorably toward the product and can thus incentivize experimentation by early users. Accordingly, spamming can be part of the socially optimal design of a recommendation system.

Second, spamming is effective only when it is properly underpinned by genuine learning. Spamming can leverage genuine learning to amplify the incentive but cannot work without genuine learning. This insight has two useful implications. First, the recommender would be well-advised to scale up spamming only gradually, beginning at a low level upon the release of a product, when an agent could not credibly believe that much had been learned about the product. Excessive spamming or highly inflated reviews at an early stage can backfire and harm the recommender's reputation. Second, a recommender could raise the level of genuine learning by investing in independent reviews of a product. Such an investment in product reviews can directly contribute to the learning on a product (which can then be shared with future consumers). But it can also substantially increase the credibility with which the designer can persuade agents to experiment. This indirect benefit could very well be as important as the direct benefit.

The optimal recommendation policy involves randomly triggered spamming when recommendation is public, as is often the case with product ratings. And in case users can choose when to check in for a recommendation, the optimal policy features an early blackout (or a suppressed recommendation) followed by a flurry of spamming and its gradual phase-out—a dynamics that encourages early agents to wait for sufficient information to accumulate but not to free ride on other agents' experimentation. Several extensions show that the general thrust of our findings is robust to heterogeneity among users with respect to their costs of experimentation and to the nature of the information learned from experimentation.

[32]It is important to note that this result is obtained without any restriction on the (finite) number of messages the designer may employ and that the designer is allowed to randomize over *paths* of recommendation. That is, she is allowed to "flip a coin" at time 0—unbeknownst to the agents—and decide on a particular function $(\alpha_t)_t$ as a function of the outcome. Such "chattering" controls are often useful but not in this instance.

Throughout, we have assumed that the recommender pursues a benevolent policy and does so with full commitment. These assumptions served to identify a social optimal policy and can be justified if a recommender competes on a fixed fee (*e.g.*, membership fee) and offers recommendations on a mix of products with varying vintages. Nevertheless, it is worth exploring what happens if these assumptions are violated. If the recommender is benevolent but unable to commit, she will spam excessively to the point of losing credibility, and users will then simply ignore the recommendation and not consume. Consequently, no experimentation occurs, and the outcome coincides with *full transparency*. If the recommender can commit but maximizes consumption (in the event that the recommender earns revenue proportional to consumption),[33] then the recommender will follow a policy similar to our optimal policy except that she will never stop recommending even after a critical level of belief is passed. The fuller implications of relaxing these assumptions and other aspects of a recommendation policy remains a subject for future research.

References

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

[2] Ausubel, L., 2004. "An Efficient Ascending-Bid Auction for Multiple Objects," *American Economic Review*, **94**, 1452–1475.

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

[4] 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.

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

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

[7] Dai, W., G. Jin, J. Lee and M. Luca, 2014. "Optimal Aggregation of Consumer Ratings: An Application to Yelp.com," working paper, Harvard Business School.

[33]Revenue may be tied to consumption, because a platform may charge a fee for each book or movie downloaded but also from the advertising profit associated with streaming or downloading.

[8] Ely, J., 2015. "Beeps," mimeo, Northwestern University.

[9] Ely, J., A. Frankel and E. Kamenica, 2015. "Suspense and Surprise," *Journal of Political Economy*, **123**, 215–260.

[10] Frick, M. and Y. Ishii, 2014. "Innovation Adoption by Forward-Looking Social Learners," working paper, Harvard.

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

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

[13] Halac, M., N. Kartik, and Q. Liu, 2014. "Contests for Experimentation," mimeo, Columbia University.

[14] 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.

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

[16] Keller, G. and S. Rady, 2015. "Breakdowns," *Theoretical Economics*, **10**, 175–202.

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

[18] Kremer, I., Y. Mansour, and M. Perry, 2014. "Implementing the "Wisdom of the Crowd,"" *Journal of Political Economy* **122**, 988–1012.

[19] Luca, M. and G. Zervas, 2014. "Fake It Till You Make It: Reputation, Competition and Yelp Review Fraud," mimeo, Harvard Business School.

[20] Mayzlin, D., Y. Dover and J. Chevalier, 2014. "Promotional Reviews: An Empirical Investigation of Online Review Manipulation," *American Economic Review*, **104**, 2421–2455.

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

[22] Papanastasiou, Y., K. Bimpikis and N. Savva, 2014. "Crowdsourcing Exploration," mimeo, London Business School.

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

[24] Renault, J., E. Solan, and N. Vieille, 2014. "Optimal Dynamic Information Provision," arXiv:1407.5649 [math.PR].

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

[26] Simon, L.K. and M. Stinchcombe, 1989. "Extensive Form Games in Continuous Time: Pure Strategies," *Econometrica*, **57**, 1171–1214.

[27] Smith, L. and P. Sørensen, 2000. "Pathological Outcomes of Observational Learning," *Econometrica*, **68**, 371–398.

[28] Smith, L., P. Sørensen, and J. Tian, 2014. "Informational Herding, Optimal Experimentation, and Contrarianism," mimeo, University of Wisconsin.

Appendix

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

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

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

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

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

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 $i = SB, FB$,

$$\sup_{u(\ell)} \int_0^{\ell^0} e^{-rt(\ell)} \left(1 - \frac{k}{\ell} - \left(\rho\left(1 - \frac{k}{\ell}\right) + 1\right) u(\ell)\right) \mathrm{d}\ell, \tag{16}$$

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

$$t'(\ell) = -\frac{u(\ell)}{\lambda\ell},$$
$$u(\ell) \in \mathcal{U}^i(\ell),$$

where the admissible set for the control is $\mathcal{U}^{SB}(\ell) := [\frac{1}{\rho+\bar{\alpha}(\ell)}, \frac{1}{\rho}]$ 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:

$$\mathcal{H}(t, u, \ell, \nu) = e^{-rt(\ell)} \left(1 - \frac{k}{\ell} - \left(\rho\left(1 - \frac{k}{\ell}\right) + 1\right) u(\ell)\right) - \nu\frac{u(\ell)}{\lambda\ell}. \tag{17}$$

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

$$\phi(\ell) := \lambda e^{-rt(\ell)}\ell\left(\rho\left(1 - \frac{k}{\ell}\right) + 1\right) + \nu(\ell) = 0, \tag{18}$$

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

Furthermore,

$$\nu'(\ell) = -\frac{\partial\mathcal{H}(t, u, \ell, \nu)}{\partial t} = re^{-rt(\ell)}\left(\left(1 - \frac{k}{\ell}\right)(1 - \rho u(\ell)) - u(\ell)\right) \ (\ell - \text{a.e.}). \tag{19}$$

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

Note that

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

or using the formulas for t' and ν',

$$\phi'(\ell) = \frac{e^{-rt(\ell)}}{\ell} \left(r(\ell - k) + \rho\lambda k + \lambda(\rho(\ell - k) + \ell) \right), \tag{20}$$

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,

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

Also, $\phi(0) = -\lambda e^{-rt(\ell)}\rho k < 0$. So $\phi(\ell) < 0$ for all $0 < \ell < \ell^*$, for some threshold $\ell^* > 0$, and $\phi(\ell) > 0$ for $\ell > \ell^*$. 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 $\ell < \ell^*$ to $1/(\rho + \bar{\alpha}(\ell))$ in the *second-best* problem and to $1/(\rho + 1)$ in the *first-best* problem for $\ell > \ell^*$. It remains to determine the switching point ℓ^* (and establish uniqueness in the process).

For $\ell < \ell^*$,

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

so that

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

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

$$\nu'(\ell) = -\frac{r}{\rho} C_1 \ell^{\frac{r}{\rho\lambda}},$$

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

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

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

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

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

$$\frac{k}{\ell^*} = 1 + \frac{\lambda}{r + \rho\lambda},$$

which leads to the formula for p^* 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-best* 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). This amounts to showing that 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 concave in t (the state variable), for all ℓ. To this end, it suffices to show that the terms inside the big parentheses in (17) are negative for all $u \in \mathcal{U}^i$, $i = FB, SB$. This is indeed the case:

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

where the inequality follows from the linearity of the expression in $u(\ell)$ and the fact that $u(\ell) \in \mathcal{U}^i \subset [\frac{1}{\rho+1}, \frac{1}{\rho}]$, for $i = FB, SB$. The concavity of maximized Hamiltonian in t, and thus sufficiency of our candidate optimal solution, then follows. $\square$

Proof of Proposition 2. Write h_t^P for the public history up to time t, and h_t for the private history of the designer — which includes whether or not she received positive feedback by time t. Let $p(h_t)$ denote the designer's belief given her private history.

- Suppose that, given some arbitrary public history h_t^P, the agent is willing to buy at t. Then, they are willing to buy if nothing more is said afterwards. To put it differently, the designer can receive her incentive-unconstrained *first-best* after such a history, and since this is an upper bound on her payoff, we might assume that this is what she does (full experimentation as long as she wishes after such a history).

- It follows that the only public histories that are non-trivial are those after which the agents are not yet willing to buy. Given h_t, the designer chooses (possibly randomly) a stopping time τ, which is the time at which she first tell the agent to buy (she then gets her *first-best*). Let $F(\tau)$ denote the distribution that she uses to tell them to buy at time τ conditional on her not having had good news by time τ; let $F_t(\tau)$ denote the distribution that she uses if she had positive news precisely at time $t \leq \tau$. We will assume for now that the designer emits a single "no buy" recommendation at any given time; we will explain why this is without loss as we proceed.

- Note that, as usual, once the designer's belief $p(h_t)$ drops below p^*, she might as well use "truth-telling:" tell agents to abstain from buying unless she has received conclusive news. This policy is credible, as the agent's belief is always weakly above the designer's belief who has not received positive news, conditional on h_t^P. And again, it gives the designer her *first-best* payoff, so given that this is an upper bound, it is the solution. It follows immediately that $F(t^*) > 0$, where t^* is the time it takes for the designer's belief to reach p^* absent positive news, given that $\mu_t = \rho$ until then. If indeed $F(t) = 1$ for some $t \leq t^*$, then the agent would not be willing to buy conditional on being told to do so at some time $t \leq \max\{t' : t' \in \operatorname{supp}(F)\}$. (His belief would have to be no more than his prior for some time below this maximum, and this would violate $c > p^0$.) Note that $F_t(t^*) = 1$ for all $t \leq t^*$: conditional on reaching time t^* at which, without good news, the designer's belief would make telling the truth optimal, there is no benefit from delaying good news if it has occurred. Hence, at any time $t > t^*$, conditional on no buy recommendation (so far), it is common knowledge that the designer has not received good news.

- The final observation: whenever agents are told to buy, their incentive constraint must be binding (unless it is common knowledge experimentation has stopped and the designer has learnt that the state is good). If not at some time t, the designer could increase $F(t)$ (the probability with which she recommends to buy at that date conditional on her not having received good news yet): she would get her *first-best* payoff from doing so; keeping the hazard rate $F(\mathrm{d}t')/(1 - F(t'))$ fixed at later dates, future incentives would not change.

Let
$$H(\tau) := \int_0^\tau \int_0^t \lambda\rho e^{-\lambda\rho s}(1 - F(s))\mathrm{d}s F_s(\mathrm{d}t).$$
This (non-decreasing) function is the probability that the agent is told to buy for the first time at some time $t \leq \tau$ given that the designer has learnt that the state is good at some earlier date $s \leq t$. Note that H is constant on $\tau > t^*$, and that its support is the same as that of F. Because $H(0) = 0$, $F(0) = 0$ as well.

Let $P(t)$ denote the agent's belief conditional on the (w.l.o.g., unique) history h_t^P such that he is told to buy at time t for the first time. We have, for any time t in the support of F,
$$P(t) = \frac{p^0\left(H(\mathrm{d}t) + e^{-\rho\lambda t}F(\mathrm{d}t)\right)}{p^0\left(H(\mathrm{d}t) + e^{-\rho\lambda t}F(\mathrm{d}t)\right) + (1 - p^0)F(\mathrm{d}t)}.$$
Indifference states that
$$P(t) = c, \text{ or } L(t) = k,$$

where $L(t)$ is the likelihood ratio

$$L(t) = \ell^0 \frac{H(\mathrm{d}t) + e^{-\rho\lambda t}F(\mathrm{d}t)}{F(\mathrm{d}t)}.$$

Combining, we have that, for any t in the support of F,

$$\left(k/\ell^0 - e^{-\rho\lambda t}\right) F(\mathrm{d}t) = H(\mathrm{d}t).^{34} \tag{21}$$

This also holds for any $t \in [0, t^*]$, as both sides are zero if t is not in the support of F. Note that, by definition of H, using integration by parts,

$$H(\tau) = \int_0^\tau \lambda\rho e^{-\lambda\rho t}(1 - F(t))F_t(\tau)\mathrm{d}t.$$

Integration by parts also yields that

$$\int_0^\tau (k/\ell^0 - e^{-\rho\lambda t})F(\mathrm{d}t) = (k/\ell^0 - e^{-\rho\lambda\tau})F(\tau) - \int_0^\tau \lambda\rho e^{-\lambda\rho t}F(t)\mathrm{d}t.$$

Hence, given that $H(0) = F(0) = 0$, we may rewrite the incentive compatibility constraint as, for all $t \leq t^*$,

$$(k/\ell^0 - e^{-\rho\lambda\tau})F(\tau) = \int_0^\tau \lambda\rho e^{-\lambda\rho t}((1 - F(t))F_t(\tau) + F(t))\mathrm{d}t,$$

and note that this implies, given that $F_t(\tau) \leq 1$ for all $t, \tau \geq t$, that

$$(k/\ell^0 - e^{-\rho\lambda\tau})F(\tau) \leq \int_0^\tau \lambda\rho e^{-\lambda\rho t}\mathrm{d}t = 1 - e^{-\lambda\rho\tau},$$

so that

$$F(t) \leq \frac{1 - e^{-\lambda\rho t}}{k/\ell^0 - e^{-\rho\lambda t}}, \tag{22}$$

an upper bound that is achieved for all $t \leq t^*$ if, and only if, $F_t(t) = 1$ for all $t \leq t^*$.

Before writing the designer's objective, let us work out some of the relevant continuation

[34] If multiple histories of "no buy" recommendations were considered, a similar equation would hold after any history h_t^P for which "buy" is recommended for the first time at date t, replacing $F(\mathrm{d}t), H(\mathrm{d}t)$ with $\tilde{F}(h_t^P), \tilde{H}(h_t^P)$; $\tilde{F}(h_t^P)$ is then the probability that such a history is observed without the designer having received good news by then, while $\tilde{H}(h_t^P)$ stands for the probability that it does with the designer having observed good news by then, yet producing history h_t^P. Define then $F, H : \mathbf{R}_+ \to \mathbf{R}_+$ as (given t) the expectation $F(t)$ (resp. $H(t)$) over all public histories h_t^P for which t is the first time at which "buy" is recommended. Taking expectations over such histories h_t^P gives (21). The remainder of the proof is unchanged.

payoff terms. First, t^* is given by our familiar threshold defined by the belief $\ell_{t^*} = k\frac{\lambda\rho+r}{\lambda(1+\rho)+r}$; given that, until t^*, conditional on no buy recommendation, experimentations occurs at rate ρ, it holds that $e^{-\lambda\rho t^*} = \ell_{t^*}/\ell^0$.

From time t^* onward, if the designer has not recommended to buy yet, there cannot have been good news. Experimentation only occurs at rate ρ from that point on. This history contributes to the expected total payoff the amount

$$p^0(1-F(t^*))e^{-(r+\lambda\rho)t^*}\frac{\lambda\rho}{r+\lambda\rho}\frac{1-c}{r}.$$

Indeed, this payoff is discounted by the factor e^{-rt^*}; it is only positive if the state is good, and it is reached with probability $p^0(1-F(t^*))e^{-\lambda\rho t^*}$: the probability that the state is good, the designer has not received any good news and has not given a buy recommendation despite not receiving any good news. Finally, conditional on that event, the continuation payoff is equal to

$$\int_0^\infty \lambda\rho e^{-rs-\lambda\rho s}\mathrm{d}s\cdot\frac{1-c}{r} = \frac{\lambda\rho}{r+\lambda\rho}\frac{1-c}{r}.$$

Next, let us consider the continuation payoff if the designer emits a buy recommendation at time $\tau \le t^*$, despite not having received good news. As mentioned, she will then experiment at maximum rate until her belief drops below p^*. The stopping time $\tau + t$ that she will pick must maximize her expected continuation payoff from time τ onward, given her belief p_τ, that is,

$$W(\tau) = \max_t\left\{p_\tau\left(1-\frac{r}{\lambda\rho+r}e^{-(\lambda(1+\rho)+r)t}\right)\frac{1-c}{r} - (1-p_\tau)(1-e^{-rt})\frac{c}{r}\right\}.$$

The second term is the cost incurred during the time $[\tau, \tau+t]$ on agents when the state is bad. The first is the sum of three terms, all conditional on the state being good: (i) $(1-e^{-rt})(1-c)/r$, the flow benefit on agents from experimentation during $[\tau, \tau+t]$; (ii) $(1-e^{-\lambda(1+\rho)t})e^{-rt}(1-c)/r$, the benefit afterwards in case good news has arrived by time $\tau+t$; (iii) $e^{-(r+\lambda(1+\rho))t}\frac{\lambda\rho}{r+\lambda\rho}(1-c)/r$, the benefit from the free experimentation after time $\tau+t$, in case no good news has arrived by time $\tau+t$. Taking first-order conditions, this function is uniquely maximized by

$$t(\tau) = \frac{1}{\lambda(1+\rho)}\ln\left(\frac{\ell_\tau}{k}\frac{\lambda(1+\rho)+r}{\lambda\rho+r}\right).$$

Note that we can write $W(\tau) = p_\tau W_1(\tau) - (1-p_\tau)W_0(\tau)$, where $W_1(\tau)$ ($W_0(\tau)$) is the benefit (resp., cost) from the optimal choice of t given that the state is good (resp., bad). Plugging in the optimal value of t gives that

$$w_1(\tau) := rW_1(\tau)/(1-c) = 1 - \frac{r}{\lambda\rho + r}\left(\frac{\ell_\tau}{k}\frac{\lambda(1+\rho)+r}{\lambda\rho+r}\right)^{-1-\frac{r}{\lambda(1+\rho)}},$$

and

$$w_0(\tau) := rW_0(\tau)/c = 1 - \left(\frac{\ell_\tau}{k}\frac{\lambda(1+\rho)+r}{\lambda\rho+r}\right)^{-\frac{r}{\lambda(1+\rho)}}.$$

Note that, given that absent any good news by time t, we have $\ell_t = \ell^0 e^{-\rho t}$. It follows that

$$\begin{aligned} k(1-w_0(t)) - \ell^0 e^{-\lambda\rho t}(1-w_1(t)) &= k\left(1 - \frac{r}{\lambda(1+\rho)+r}\right)\left(\frac{k}{\ell_t}\frac{\lambda\rho+r}{\lambda(1+\rho)+r}\right)^{\frac{r}{\lambda(1+\rho)}} \\ &= Ke^{\frac{r\rho}{1+\rho}t}, \end{aligned} \tag{23}$$

with

$$K := k\frac{\lambda(1+\rho)}{\lambda(1+\rho)+r}\left(\frac{k}{\ell^0}\frac{\lambda\rho+r}{\lambda(1+\rho)+r}\right)^{\frac{r}{\lambda(1+\rho)}}.$$

For future reference, note that, by definition of ℓ_{t^*},

$$Ke^{\frac{r\rho}{1+\rho}t^*} = k\frac{\lambda(1+\rho)}{\lambda(1+\rho)+r}\left(\frac{k}{\ell_{t^*}}\frac{\lambda\rho+r}{\lambda(1+\rho)+r}\right)^{\frac{r}{\lambda(1+\rho)}} = k\frac{\lambda(1+\rho)}{\lambda(1+\rho)+r}. \tag{24}$$

We may finally write the objective. The designer wishes to choose $\{F, (F_s)_{s=0}^{t^*}\}$ so as to maximize

$$\begin{aligned} J = p^0 &\int_0^{t^*} e^{-rt}\left(\frac{1-c}{r}H(\mathrm{d}t) + e^{-\rho\lambda t}W_1(t)F(\mathrm{d}t)\right) \\ &- (1-p^0)\int_0^{t^*} e^{-rt}W_0(t)F(\mathrm{d}t) + p^0(1-F(t^*))e^{-(r+\lambda\rho)t^*}\frac{\lambda\rho}{r+\lambda\rho}\frac{1-c}{r}. \end{aligned}$$

The first two terms are the payoffs in case a buy recommendation is made over the interval $[0, t^*]$ and is split according to whether the state is good or bad; the third term is the benefit accruing if no buy recommendation is made by time t^*.

Multiplying by $\frac{r}{1-c}\frac{e^{rt^*}}{1-p^0}$, this is equivalent to maximizing

$$\int_0^{t^*} e^{-r(t-t^*)}\left(\ell^0 H(\mathrm{d}t) + \ell^0 e^{-\rho\lambda t}w_1(t)F(\mathrm{d}t) - kw_0(t)F(\mathrm{d}t)\right) + \ell^0(1-F(t^*))e^{-\lambda\rho t^*}\frac{\lambda\rho}{r+\lambda\rho}.$$

We may use (21) (as well as $\ell^0 e^{-\lambda\rho t^*} = \ell_{t^*} = k\frac{\lambda\rho+r}{\lambda(1+\rho)+r}$) to rewrite this as

$$\int_0^{t^*} e^{-r(t-t^*)}\left(k(1-w_0(t)) - \ell^0 e^{-\rho\lambda t}(1-w_1(t))\right)F(\mathrm{d}t) + (1-F(t^*))\frac{\lambda\rho k}{\lambda(1+\rho)+r}.$$

Using (23), and ignoring the constant term $\frac{\lambda\rho k}{\lambda(1+\rho)+r}$ (irrelevant for the maximization) this gives

$$e^{rt^*}K\int_0^{t^*} e^{-\frac{r}{1+\rho}t}F(\mathrm{d}t) - \frac{\lambda\rho k}{\lambda(1+\rho)+r}F(t^*),$$

or, integrating by parts and using that $F(0)=0$, as well as (24),

$$e^{rt^*}\frac{rK}{1+\rho}\int_0^{t^*} e^{-\frac{r}{1+\rho}t}F(t)\mathrm{d}t + \left(k\frac{\lambda(1+\rho)}{\lambda(1+\rho)+r} - k\frac{\lambda\rho}{\lambda(1+\rho)+r}\right)F(t^*),$$

or finally (using (24) once more to eliminate K)

$$\frac{\lambda k}{\lambda(1+\rho)+r}\left(\int_0^{t^*} re^{-\frac{r}{1+\rho}(t-t^*)}F(t)\mathrm{d}t + F(t^*)\right).$$

Note that this objective function is increasing pointwise in F. Hence, it is optimal to set F as given by its upper bound given by (22), namely, for all $t \le t^*$,

$$F(t) = \frac{\ell^0(1-e^{-\lambda\rho t})}{k - \ell^0 e^{-\rho\lambda t}},$$

and for all $t \le t^*$, $F_t(t) = 1$. $\square$

Proof of Proposition 3. **Part (i):** Under *full transparency*, agents who check in at t realize payoff

$$\mathcal{U}_t = e^{-rt}(\ell_0 - \ell_t).$$

Note that the timing at which agents check in is irrelevant for belief updating (because those who check in never experiment), so

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

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

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

Part (ii): 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 $t_2 = t_1 + \Delta$:

$$\ell^0 - \ell_1 - \alpha_1(k - \ell_1) = e^{-r\Delta}(\ell^0 - \ell_2), \tag{25}$$

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 (25) 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

$$\rho \leq \frac{1}{\lambda}\left(\frac{r\ell^0 + \sqrt{r\ell^0}\sqrt{4k\lambda + \ell^0}}{2k} - r\right).$$

Part (iii) is more involved and is provided in Supplementary Material. $\square$

Supplementary Material: "Optimal Design for Social Learning"

Yeon-Koo Che and Johannes Hörner

April 28, 2015

1 Endogenous Entry: Proof of Proposition 3

Given the optimality of revealing good news whenever it is received, we can without loss focus on the agents' check in times and their experimentation decisions as the two control variables. The optimal policy is therefore described by a process $(X_t, \alpha_t)_t$ where $X_t \leq t$ is the right-continuous measure of all agents that have checked in by time t, and $\alpha_t \in [0,1]$ is the probability of consumption (or experimentation) by agents who check in at time t, conditional on having received no news by t. As will be seen below, the set of all such processes is not rich enough to admit an optimal mechanism. We thus consider a richer space of policies.

Specifically, we allow for the possibility that a mass of agents check in instantaneously—that is, without elapse of any real time—but sequentially experiment at rates that depend on the ℓ. Formally, we enrich the policy space so that whenever X_t jumps at t (so a mass of agents check in at t), we allow the designer to run a second "virtual" clock $s \in [0, m_t]$, where $m_t := X_t - X_t^-$, where $X_t^- := \lim_{t' \uparrow t} X_{t'}$. This virtual clock does not take any real time, but can be used to sequence the agents' check in and experimentation decisions.[1]

Formally, we use (X_t^s, α_t^s) to denote the enriched policy, where X_t^s is the mass of agents who check in by real time t and virtual time s, and α_t^s is the experimentation probability for agents who check in at (t, s). For consistency, we require $X_t^s \geq X_{t'}^-$ and $X_t^{m_t} = X_t$. The virtual clock can be used to split the atom into flows of agents checking in sequentially,

[1]The need for the virtual clock, or the enrichment of the strategy space more generally, arises from the peculiarity of the continuous time game that the soonest next time after any time t is not well defined. To avoid nonexistence, therefore, it is sometimes necessary to allow for sequential moves that do not take any real time. See Simon and Stinchcombe (1989) and Ausubel (2004) for adopting similar enrichment of strategies to guarantee existence of an equilibrium. Note also that the need for the virtual clock disappears once we formulate the policy as a function of ℓ, in which case a split mass is captured by a time "state" variable $t(\ell)$ that is constant for an interval of ℓ's.

in which case X_t^s admits density for all $s \in (0, m_t)$. In such a case, we shall without loss assume that $X_t^s = s + X_t^-$. The virtual clock also admits the entire mass of agents checking in simultaneously, in which case $X_t^s = X_t$ for all s. The current framework also allows for the mass to be split in finite number of lumps.

Throughout, we fix an optimal policy (X, α) and derive properties it must satisfy.

Lemma 1. *At the optimal policy, $t^* := \inf\{t \geq 0 \mid X_t > 0\} > 0$. That is, a mass of agents are induced to wait for a strictly positive amount of time before they check in.*

Proof. Let (ℓ_t, α_t) denote the optimal policy. There exists $T > 0$ such that the surplus accruing to agents checking in at T:

$$e^{-rT}(\ell_0 - \ell_T - \alpha_T(k - \ell_T))$$

is strictly positive, or else the policy will be even inferior to the full transparency. To induce agents to check in any $t < T$, we must have

$$e^{-rt}(\ell_0 - \ell_t - \alpha_t(k - \ell_t)) \geq e^{-rT}(\ell_0 - \ell_T - \alpha_T(k - \ell_T)).$$

The LHS of this inequality is no greater than $\ell_0 - \ell_t$, which goes to zero as $t \to 0$. This proves that there exists $t^* > 0$ such that no agents will check in at $t < t^*$. $\square$

Lemma 2. *Suppose $\hat{t} \in supp(X)$ and $\hat{t} > X_{\hat{t}}^- := \lim_{t \uparrow \hat{t}} X_t$. Then, X jumps at $\hat{t}$.*

The proof consists of two steps.

Step 1. *Suppose X does not jump at $\hat{t}$. Then, there exists $\epsilon > 0$ such that positive density of agents check in $t \in (\hat{t}, \hat{t} + \epsilon)$ and experiment along the locus*

$$\breve{\alpha}(\ell_t) := 1 - \frac{r(k - \ell^0)}{r(k - \ell_t) - \lambda \ell \rho},$$

and the belief evolves according to $\dot{\ell} = -\lambda(\rho + \breve{\alpha}(\ell_t))\ell_t$.

Proof. For $\epsilon > 0$ sufficiently small, X_t admits density $x_t > 0$ for $t \in (\hat{t}, \hat{t} + \epsilon)$. Consider the agents who check in during this time interval. We shall consider the implication of the property of the optimal policy that the designer cannot be better off by redistributing within this time interval while maintaining their incentive to comply with the redistribution. Fix any $t \in (\hat{t}, \hat{t} + \epsilon)$, and let $t' = t + 2\mathrm{d}t < t^* + \epsilon$, and the designer reduces the flow of the agents who check in during $[t, t + \mathrm{d}t)$ by $\delta < x_t$ and increases the flow of the agents who check in during $[t + \mathrm{d}t, t + 2\mathrm{d}t)$ by the same δ, while ensuring that the agents are indifferent over the check in times during that interval. This operation is feasible for sufficiently small

$\delta \in (0, x_t)$. For the original policy to be optimal, such an operation should not lower the posterior (in likelihood) $\ell_{t'}$ (since a lower $\ell_{t'}$ means a higher learning benefit which is strictly preferred under commitment).

We thus require that the operation cannot lower $\ell_{t'}$. To study the effect of the operation on $\ell_{t'}$, let $\alpha_1 := \alpha_t, \alpha_2 := \alpha_{t+\mathrm{d}t}, \alpha_3 := \alpha_{t+2\mathrm{d}t}$, and $\ell_1 := \ell_t$, $\ell_2 := \ell_{t+\mathrm{d}t}, \ell_3 := \ell_{t+2\mathrm{d}t}$. A few conditions must be satisfied:

- Agents are indifferent over consuming at time t and $t + \mathrm{d}t$, *i.e.*,

$$\ell_0 - \ell_1 - \alpha_1(k - \ell_1) = e^{-r\mathrm{d}t}(\ell_0 - \ell_2 - \alpha_2(k - \ell_2)).$$

- Since the flow of consumers picking time t is $(x - \delta)$ while the flow at time $t + \mathrm{d}t$ is $(x + \delta)$, the beliefs at ℓ_2 and ℓ_3 must satisfy

$$\ell_2 = \ell_1 e^{-\lambda(\rho+(x-\delta)\alpha_1)\mathrm{d}t},$$

  and

$$\ell_3 = \ell_2 e^{-\lambda(\rho+(x+\delta)\alpha_2)\mathrm{d}t}.$$

We now fix ℓ_1 and α_1 (and hence the utility for a regular consumer to choose the first instant), and solve this system for ℓ_2, ℓ_3, α_2, as a function of δ, α_1, ℓ_1. Differentiate ℓ_3 with respect to δ, and evaluate the derivative at $\delta = 0$ to get:

$$\operatorname{sgn}\left(\left.\frac{\partial \ell_3}{\partial \delta}\right|_{\delta=0}\right) := \operatorname{sgn}\Big((\breve{\alpha}(\ell_t) - \alpha_t)(\mathrm{d}t)^2 + o((\mathrm{d}t)^2)\Big).$$

This shows that if $\alpha_t > \breve{\alpha}(\ell_t)$, there exists $\delta \in (0, x_t)$ and $t' > 0$ sufficiently small such that the redistribution of agents lowers the posterior at t', a contradiction to its optimality. Hence, we conclude that $\alpha_t \leq \breve{\alpha}(\ell_t)$ for any $t \in T_0$.

Next suppose $\alpha_t < \breve{\alpha}(\ell_t)$. Then, since $X_t < t$ for all $t \in (\hat{t}, \hat{t} + \epsilon)$, a perturbation from the optimal policy considered above, with $\delta < 0$ (meaning shifting forward the check-in time of some flow of agents), is feasible and will also lower the posterior, contradicting the optimality of the original policy.

Given the experimentation policy follows the locus $\breve{\alpha}(\cdot)$, the belief must follow the law of motion. □

Step 2. *X must jump at $\hat{t}$.*

Proof. By Step 1, for some $\epsilon > 0$, a positive flow of agents check in at each $t \in (\hat{t}, \hat{t} + \epsilon)$ and experiment along the locus $\breve{\alpha}(\ell_t) := 1 - \frac{r(k-\ell^0)}{r(k-\ell_t)-\lambda\ell\rho}$. For the optimal policy to be incentive compatible, the agents must be indifferent along the locus $\breve{\alpha}(\cdot)$. In particular, the payoff of agents who check in at $t \in (\hat{t}, \hat{t}+\epsilon)$,

$$e^{-rt}(\ell_0 - \ell_t - \breve{\alpha}(\ell_t)(k - \ell_t)), \tag{1}$$

must be constant in t, where the belief ℓ_t evolves according to $\dot{\ell} = -\lambda(\rho + \breve{\alpha}(\ell_t))\ell_t$. One can check that this is not the case. In particular, the payoff (1) decreases in t, which contradicts incentive compatibility. $\square$

Lemma 3. *If X jumps at $\hat{t}$, then $X_{\hat{t}}^s$ admits no atom at $s \in [0, m_{\hat{t}}]$. In other words, any atom of $X_{[\cdot]}$ is split into a flow of agents checking in continuously according to the virtual clock.*

Proof. Suppose to the contrary that $X_{\hat{t}}^{[\cdot]}$ jumps at $\hat{s} \in [0, m_{\hat{t}}]$, and let $m := X_{\hat{t}}^{\hat{s}} - \lim_{s\uparrow\hat{s}} X_{\hat{t}}^s$ be the mass of agents who experiment simultaneously with probability $\alpha > 0$ at the virtual clock $s = \hat{s}$. These agents enjoy the expected payoff of

$$e^{-r\hat{t}}(\ell^0 - \ell_{\hat{t}}^{\hat{s}} + \alpha(k - \ell_{\hat{t}}^{\hat{s}})),$$

where $\ell_{\hat{t}}^{\hat{s}} := \lim_{s\uparrow\hat{s}} \ell_{\hat{t}}^s$ is the belief just before $(\hat{t}, \hat{s})$, and $\ell_{\hat{t}}^{\hat{s}+\mathrm{d}s} := \ell_{\hat{t}}^{\hat{s}} e^{-\lambda(\alpha m + \mathcal{O}(\mathrm{d}s))}$ is the belief just after, for small $\mathrm{d}s > 0$. Consider now a deviation in which the designer splits mass $\delta \in (0, m)$ out of mass m and move it by $\mathrm{d}s$ in the virtual clock. After the experimentation by the first mass $m - \delta$, the belief becomes

$$\hat{\ell}_{\hat{t}}^{\hat{s}+\mathrm{d}s} = \ell_{\hat{t}}^{\hat{s}} e^{-\lambda(\alpha(m-\delta)+\mathcal{O}(\mathrm{d}s))},$$

which is smaller than, and bounded away from, $\ell_{\hat{t}}^{\hat{s}}$ for any $\mathrm{d}s > 0$. The second mass δ of agents can be induced to experiment at rate $\hat{\alpha}$ such that

$$e^{-r\hat{t}}(\ell^0 - \hat{\ell}_{\hat{t}}^{\hat{s}+\mathrm{d}s} + \hat{\alpha}(k - \hat{\ell}_{\hat{t}}^{\hat{s}+\mathrm{d}s})) = e^{-r\hat{t}}(\ell^0 - \ell_{\hat{t}}^{\hat{s}} + \alpha(k - \ell_{\hat{t}}^{\hat{s}})).$$

Since $\hat{\ell}_{\hat{t}}^{\hat{s}+\mathrm{d}s}$ is smaller than and is bounded away from $\ell_{\hat{t}}^{\hat{s}}$, it follows that $\hat{\alpha} - \alpha$ is bounded away from zero for any $\mathrm{d}s$. Hence, the deviation results in the belief,

$$\ell_{\hat{t}}^{\hat{s}} e^{-\lambda(\alpha(m-\delta)+\hat{\alpha}\delta+\mathcal{O}(\mathrm{d}s))} < \ell_{\hat{t}}^{\hat{s}} e^{-\lambda(\alpha m+\mathcal{O}(\mathrm{d}s))} = \ell_{\hat{t}}^{\hat{s}+\mathrm{d}s},$$

for $\mathrm{d}s$ small enough. This is a contradiction to the optimality of the original policy. We therefore conclude that $X_{\hat{t}}^s$ is atomless in s. $\square$

Remark 3. *The argument of Lemma 3 also implies that the optimal policy is not well defined without the enriching of the space. Lemmas 1 and 2 imply that a positive of mass of agents is induced to wait before they check in, and check in at some time $t^* > 0$. But it is never* *optimal for them to check in all simultaneously. The argument given in the proof of Lemma 3 suggests that it is optimal to split the mass and move them later, with an arbitrarily small delay.*

We next investigate an atom X. As Lemma 3 suggests, any atom must be split. We derive a necessary condition for the endpoint of the atom split.

Lemma 4. *Let $(\alpha_t^s, \ell_t^s)_{s\in[0,m_t]}$ be the process of experimentation and beliefs associated with a split atom at t. Then, defining*

$$\bar{\alpha}(\ell) := \frac{r\ell^0 - \ell(\rho\lambda + r)}{rk - \ell(\rho\lambda + r)},$$

it must hold that there exists a (unique) (α_t^s, ℓ_t^s) with $s \in [0, m_t]$ such that $\ell_t^s = \hat{\alpha}(\ell_t^s)$.

Proof. Let there be a mass at t, with $X_t - X_t^- > 0$, and Lemma 3, the mass is split with (α^-, ℓ^-) and (α^+, ℓ^+) denoting start and end points of the split mass. We shall consider a variation which moves the a small segment of the mass forward or backward slightly, and ask when such a variation is profitable. Specifically, we move a segment $\{X_t^s\}_{s'}^t$ of agents with sufficiently small mass $m := X_t^t - X_t^{s'}$ forward in time by a small interval $\mathrm{d}t > 0$ subject to the constraint that the moved agents enjoy the same discounted payoffs, and that the belief at time $t + \mathrm{d}t$ remain the same as before, namely $\ell^+ - \lambda\rho\ell^+\mathrm{d}t$. We then derive the belief at (t, s'), denoted ℓ', that would allow for such a move to be feasible. If $\ell' > \ell_t^{s'}$, then this means that starting from $\ell_t^{s'}$, it is indeed possible to move the segment that would result in the belief at $t + \mathrm{d}t$ being strictly lower than ℓ^+, a welfare improvement. So, the optimality of the original policy will require that $\ell' \leq \ell_t^{s'}$. We shall show that this requirement produces a condition: $\alpha^+ \leq \bar{\alpha}(\ell^+)$.

To begin, consider the variation. First, we require the agents involved in the moved segment to be indifferent to the move. In particular, the agents at the end point of the moved split must enjoy the utility equal to:

$$\mathcal{U}_t = e^{-rt}\left(\ell^0 - \ell^+ - \alpha^+(k - \ell^+)\right).$$

Upon differentiating, this means that

$$(k - \ell^+)\mathrm{d}\alpha^+ = ((1 - \alpha^+)\rho\lambda\ell^+ - r(\ell^0 - \ell^+ - \alpha^+(k - \ell^+)))\mathrm{d}t, \tag{2}$$

using that $\mathrm{d}\ell^+ = -\lambda\rho\ell^+\mathrm{d}t$, which follows from the requirement that the belief at $t + \mathrm{d}t$ must be equal to $\ell^+ - \lambda\rho\ell^+\mathrm{d}t$, the level that would prevail at $t + \mathrm{d}t$ had there been no variation.

Meanwhile, the agents involved in the split atom must be indifferent along the process $(\alpha^\sigma, \ell^\sigma)_{\sigma\in[0,m]}$, where σ is the new virtual clock that is run to sequence the agents who are moved. Let $(\alpha_1, \ell_1) := (\alpha^0, \ell^0)$ be the start point of the moved agents. The end point is $(\alpha^m, \ell^m) = (\alpha^+, \ell^+)$. The indifference means that

$$\ell_0 - \ell^\sigma - \alpha^\sigma(k - \ell^\sigma) = \ell_0 - \ell^+ - \alpha^+(k - \ell^+),$$

for all $\sigma \in [0, m]$. Hence,

$$\alpha^\sigma = \alpha(\ell^\sigma) := \frac{h - \ell^\sigma}{k - \ell^\sigma}, \tag{3}$$

where $h := \ell^+ + \alpha^+(k - \ell^+)$.

Now, the equation

$$\frac{\mathrm{d}\ell^\sigma}{ds} = -\lambda \ell^\sigma \alpha(\ell^\sigma)$$

can be solved for the virtual time $s(\ell)$ that it takes to reach a given belief:

$$s'(\ell) = -\frac{k - \ell}{\lambda \ell (h - \ell)},$$

which gives, along the curve,

$$s(\ell) = C - \frac{k \ln \ell + (h - k)\ln(h - \ell)}{\lambda h}, \quad C \in \mathbf{R}.$$

Now recall that $s(\ell^+) - s(\ell_1) = m$. We therefore obtain:

$$m = \frac{k \ln \ell_1 + (h - k)\ln(h - \ell_1)}{\lambda h} - \frac{k \ln \ell^+ + (h - k)\ln(h - \ell^+)}{\lambda h}. \tag{4}$$

We may derive $\mathrm{d}\ell_1$ from this expression, by totally differentiating with respect to α^+ and ℓ^+, using (2). This yields (as a change for a given $\mathrm{d}t$)

$$\mathrm{d}\ell_1 = \frac{\ell_1 (h - \ell_1) \left(\frac{k\left(h - \ell^0\right) r \left(\ln\left(\frac{h - \ell^+}{h - \ell_1}\right) + \ln\left(\frac{\ell_1}{\ell^+}\right)\right)}{h} - \frac{h\lambda\rho(h - \ell_1)\left(k - \ell^+\right) - \left(h - \ell^0\right)\left(\ell^+ - \ell_1\right) r (h - k)}{(h - \ell^+)(h - \ell_1)} \right)}{h (k - \ell_1)} \mathrm{d}t + o(\mathrm{d}t).$$

We now consider doing this for a small change m. That is, we are considering delaying by $\mathrm{d}t$ the experimentation performed by a small mass m (that is, picking a small measure of those agents supposed to check in, and delaying this checking-in –making sure they are willing to wait). To be clear, the change in m is small (so that Taylor expansions apply to (4)), but given this m, we take $\mathrm{d}t$ to be small (so that the previous differential holds approximately). Expanding ℓ_1 from (4) in m gives that (in terms of m)

$$\ell_1 = \ell^+ + \lambda \alpha^+ \ell^+ m + o(m).$$

This is the impact of the mass m on the belief at the end of the splitting. We now delay their experimentation by $\mathrm{d}t$, and use the expression for $\mathrm{d}\ell_1$. Because some background learning would have occurred after the original splitting during the $\mathrm{d}t$ interval, to evaluate the new belief at t, we must account for the learning. To obtain the effect on the new belief at t consistent with the move, we are moving backward in time, so we must add $\rho\lambda\ell_1\mathrm{d}t$ (we had already subtract $\lambda\rho\ell^+\mathrm{d}t$).[2] In sum, the effect on the belief that must prevail at t for the above variation to be possible must equal:

$$\mathrm{d}\ell_1 + \lambda\rho\ell_1\mathrm{d}t = \frac{\ell^+}{k - \ell^+}\left((\alpha^+ k - \ell^0)r + (1-\alpha^+)\ell^+(r + \lambda\rho)\right)\lambda m + o(m).$$

If this expression were positive, this means that the new belief at t consistent with this move is higher than the original belief at (t, s'), meaning that it would be possible to move the agents in a way that keeps the incentives of all agents intact and yields a lower belief at $t + \mathrm{d}t$. Since the latter move would be strictly profitable for the designer, the optimality of the original policy requires the expression above to be nonpositive, or

$$\alpha^+ \leq \frac{r\ell^0 - \ell^+(\rho\lambda + r)}{rk - \ell^+(\rho\lambda + r)}.$$

Similar reasoning applies at the start of the atom splitting, pushing backward in time by $\mathrm{d}t$ a small mass m of experimenters. For this not be profitable, we then get

$$\alpha^- \geq \frac{r\ell^0 - \ell^-(\rho\lambda + r)}{rk - \ell^-(\rho\lambda + r)}.$$

Combining the inequalities, we conclude that the atom splitting must cross the locus $(\ell, \hat{\alpha}(\ell))$. It is readily checked from (3) that the slope of the locus $\alpha_t^s(\ell)$ is larger than the slope $\mathrm{d}\alpha/\mathrm{d}\ell$ along the locus $(\alpha_t^s, \ell_t^s)_{s\in[0,m_t]}$, so that they cross only once. □

Lemma 5. *Suppose $\alpha_t > 0$ at $t > 0$.*[3] *Then, the discounted utility of the agents who check in before cannot be strictly higher.*

Proof. We will prove that the discounted utility of an agent at time t is not boundedly lower than those who check in immediately before. The argument can be extended, as mentioned below, to show that the agents' discounted utility does not decline over time. Suppose that the utility of an agent at time t is boundedly lower than the utility of an agent who has not checked in at time $t - \epsilon$, for all $\epsilon > 0$. The reasoning below assumes a gradual check in over some interval $[t - \Delta, t + \Delta]$, but can be adjusted (in case there is a mass point). Let α denote

[2] The change of experimentation due to agents arriving during the interval $[t, t + \mathrm{d}t)$ and possibly asked to experiment is of order $\mathrm{d}t \cdot \mathrm{d}m$ and so ignored.

[3] In case X has a mass point at t, we select $\alpha_t := \sup_s \alpha_t^s$.

the probability of recommendation at times $[t-\Delta, t)$. More precisely, let $\alpha = \lim_{\tau \uparrow t} \alpha_\tau$ and let $\Delta > 0$ be such that $|\alpha_\tau - \alpha| < \varepsilon$ for all $\tau \in [t-\Delta, t)$, for some fixed $\varepsilon > 0$ arbitrarily small.)

Similarly, let α' denote the probability of recommendation at times $[t, t+\Delta)$ (again, take limits in the obvious way.) Because the utility is boundedly lower at times in $[t, t+\Delta)$, we may increase the probability α by some $\delta > 0$ and decrease α' by the same amount, so that (i) the total experimentation (and hence belief) at time $t+\Delta$ is the same before and after the change, (ii) agents supposed to check in at times $[t-\Delta, t)$ prefer to do so than to check at times $[t, t+\Delta)$. Clearly, agents checking in at time $[t, t+\Delta)$ gain from this change and so will still check in.[4]

We ask, when does such a change increase welfare? By construction, it does not affect what happens before (agents before certainly don't want to wait now) nor after (the belief at $t+\Delta$ hasn't changed). Hence, the change in payoff is

$$(\ell_0 - \ell - (\alpha+\delta)(k-\ell) + e^{-r\Delta}(\ell_0 - \ell_\Delta - (\alpha' - \delta)(k - \ell_\Delta)) + o(\Delta),$$

where ℓ, ℓ_Δ are the beliefs at the beginning of each subinterval. Taking derivatives with respect to δ and then a Taylor expansion with respect to Δ, evaluated at $\delta = 0$, gives that the derivative equals

$$(\ell(1+\rho+r) - rk)\Delta + o(\Delta),$$

where we normalize λ to 1 (Alternatively, replace all occurrences of r by r/λ). Hence, such a change is profitable if $\ell > \frac{rc}{1+\rho+r}$. Thus, we may assume otherwise.

If the change in utility is not bounded below by some constant, the same reasoning applies, but $\delta = \delta(\tau)$ must be chosen so that $\delta(t) = 0$, $\delta(\tau) > 0$ for $\tau < t$, and $\delta(\tau) > 0$ for $\tau > t$, such that agents do not wish to change their check-in time over these intervals. If there is an atom at t, then there must be an black-out immediately before t, and a similar reasoning applies for moving a small mass m' of split atom backward in time and raise their experimentation by small δ. Both extensions are omitted.

Let us now recall that the total continuation payoff is given by

$$J = \int_{s \geq t} e^{-rs}(\ell^0 - \ell_s - \alpha_s(k - \ell_s))\mathrm{d}s.$$

Because $\dot{\ell} = -(\rho+\alpha)\ell_s$, we can substitute to obtain an expression that only depends on $\ell, \dot{\ell}$, integrate by parts to eliminate the terms involving $\dot{\ell}$, and ignoring constants, obtain that,

[4] Adding small $\delta > 0$ to α does not violate the incentive constraint for consumption since the constraint is not binding.

up to a constant, J is equal to

$$
\begin{aligned}
J &= \int_{s\geq t} e^{-rs}(\ell^0 - \ell_s - \alpha_s(k-\ell_s))\mathrm{d}s \\
&= \int_{s\geq t} e^{-rs}(\ell^0 - \ell_s + (\rho + \frac{\dot{\ell}}{\lambda \ell_s})(k-\ell_s))\mathrm{d}s \\
&= \int_{s\geq t} e^{-rs}\left(\ell^0 - (1+\rho)\ell_s - \frac{\dot{\ell}_s}{\lambda} + \frac{\dot{\ell}}{\lambda \ell_s}k\right)\mathrm{d}s \\
&= \int_{s\geq t} e^{-rs}(\ell^0 - (1+\rho)\ell_s)\mathrm{d}s - e^{-rt}\left(\frac{\ell_s}{\lambda} - \frac{\ln \ell_s}{\lambda}k\right)_t^\infty - \int_{s\geq t} e^{-rs} r\left(\frac{\ell_s}{\lambda} - \frac{\ln \ell_s}{\lambda}k\right)\mathrm{d}s \\
&= -\frac{1}{\lambda}\int_{s\geq t} e^{-rs}\left(\lambda(1+\rho)+r)\ell_s - rk\ln \ell_s\right)\mathrm{d}s + \text{Const.}
\end{aligned}
$$

The derivative of the integrand with respect to ℓ is

$$rk - (\lambda(1+\rho)+r)\ell > 0.$$

Hence, given t as defined above, we note that this derivative is positive: if we replace the trajectory $\{\ell_\tau : \tau \geq t\}$ by a trajectory $\{\hat{\ell}_\tau : \tau \geq t\}$, with $\hat{\ell}_\tau \geq \ell_\tau$ (with a strict inequality for some non-zero measure interval of times), the payoff increases. Hence, decrease α at time t (or rather, fix a time higher than, but arbitrarily close to t and decrease α at that time), and adjust α_τ for all later $\tau \in T_0$ so that incentives to check in do not change. This requires non-positive changes in α_τ (which can be expressed in terms of a differential equation), and results in a higher trajectory $\hat{\ell}$, and hence an increase in payoff. We thus obtain a contradiction. □

Lemma 6. *$X_t = 0$ for $t < t^*$ and $X_t = t$ for $t \geq t^*$.*

Proof. By Lemma 1, $t^* := \inf\{t \geq 0 | X_t > 0\} > 0$. Since $X_{t^*}^- = 0 < t^*$, by Lemma 2, X has a mass point at t^*. Further, by Lemma 3, the atom at t^* must be split. Let (α^+, ℓ^+) be the end point of the split atom at t^*. Then, by Lemma 4, $\alpha^+ \leq \hat{\alpha}(\ell^+)$. One can show hat $\hat{\alpha}(\cdot)$ is steeper than $\alpha(\cdot)$.

We next show that for any $t > t^*$, $X_t = t$. Suppose to the contrary that there exists $t' > t$ such that $X_{t'} < t'$. Then, by Lemmas 2–4, X has an atom at t', and it is split, and the splitting crosses the locus $\hat{\alpha}(\cdot)$. But this is impossible, since by Lemma 5, the agents' utility never strictly decrease for $t \in [t^*, t']$, and this means that during that interval, α can never rise at a faster rate in ℓ than $\alpha(\ell)$ (the locus followed in the first split atom) does. Since $X_{t'} = t'$ for all $t' > t$, and since X is required to be right continuous, the claim follows. □

Armed with the lemmas, we now complete our characterization of the optimal policy.

Proof of Proposition 3 and the solution algorithm. Lemmas 4-6 pin down the structure of the optimal policy: there exist times $t^* > 0$ and $T > t^*$ such that agents who arrive before t^* wait until t^*; the accumulated mass is split at t^*; and then the agents who arrive after t^* check in upon arrival and experiment at a rate that falls to zero at T. All agents' discounted payoff is constant for all argents arriving prior to T, and the agents arriving after T enjoys payoff according to *full transparency regime*. This structure, along with further necessary conditions, enables us to derive an one-dimensional family of optimal policies indexed by the belief $\bar{\ell}$ at which agents' experimentation stops fully.

Initially, we fix both $\bar{\ell}$ and t^*. The variables to be determined are $(\ell^-, \alpha^+, \ell^+, T)$, where T is such that $\ell_T = \bar{\ell}$. Several conditions are derived to determine these variables. First, since only background learning occurs during the blackout, the (designer's) belief just prior to the splitting must satisfy

$$\ell^- = \ell^0 e^{-\lambda\rho t^*}. \tag{5}$$

Second, the agents associated with split mass must be indifferent across the locus $(\alpha_t^s, \ell_t^s)_{s\in[0,t^*]}$. This requires (3). This, together with the fact that the entire mass of t^* is split, gives rise to an equation (4) with $m = t^*$, or

$$t^* = \frac{k\ln\ell^- + (h-k)\ln(h-\ell^-)}{\lambda h} - \frac{k\ln\ell^+ + (h-k)\ln(h-\ell^+)}{\lambda h}, \tag{6}$$

where $h := \ell^+ + \alpha^+(k - \ell^+)$.

Third, by Lemma 6, agents check in as they arrive during the the smooth tapering phase, and by Lemma 5, they must experiment at levels that make them all indifferent. Hence, during the $t \in (t^*, T)$, (α_t, ℓ_t) must satisfy the indifference condition:

$$e^{-rt}(\ell^0 - \ell_t - \alpha_t(k - \ell_t)) = e^{-rT}(\ell^0 - \bar{\ell}), \tag{7}$$

and the belief evolution condition:

$$\ell_t = \bar{\ell} e^{\lambda \int_t^T (\rho + \alpha_s)\mathrm{d}s}. \tag{8}$$

Given $(t^*, \bar{\ell})$, conditions (7) and (8) uniquely determine (α^+, ℓ^+) as a function of T.[5]

In sum, the conditions (5)–(8) uniquely pin down $T, \ell^-, \alpha^+, \ell^+$ as functions of $(\bar{\ell}, t^*)$. Next, we hold $\bar{\ell}$ fixed and characterize the condition for optimal choice of t^*. In particular, we use the fact that for a fixed $\bar{\ell}$, T must be minimized. This follows from the fact that the earlier time T at which a given belief $\bar{\ell}$ is reached without affecting the payoff of the

[5]Essentially, given any $T > t^*$, the two conditions give rise to a differential equation that runs from T backward to t^*, with the initial value $(\alpha_T, \ell_T) = (0, \bar{\ell})$, which admits a unique solution (α_t, ℓ_t), the end point of which is (α^+, ℓ^+).

agents who arrive before, the higher the payoff is for all agents arriving afterwards, and thus the higher the overall welfare is. This means that, as t^* is varied slightly by $\mathrm{d}t$, subject to the constraint that all agents who arrive before T should enjoy the same payoff, the optimal choice of T remains constant. Since the variation should not alter T at the optimum, as t^* is raised by $\mathrm{d}t$, the last phase is shortened by length $\mathrm{d}t$. Hence, the variation affects ℓ^+, the belief at which (going backward) the differential equation ends, by:

$$\mathrm{d}\ell^+ = -\lambda(\rho + \alpha^+)\ell^+\mathrm{d}t. \tag{9}$$

Next, the variation keeps the payoff of early agents constant. Totally differentiating the indifference condition, we obtain the change in α^+ arising from the variation:

$$r(\ell_0 - \ell^+ - \alpha^+(k - \ell^+))\mathrm{d}t = -(1 - \alpha^+)\mathrm{d}\ell^+ - (k - \ell^+)\mathrm{d}\alpha^+. \tag{10}$$

Now, since the initial phase has been lengthened by $\mathrm{d}t$, we should have

$$\ell^- + \mathrm{d}\ell^- = \ell_0 e^{-\lambda\rho(t^* + \mathrm{d}t)},$$

or

$$\frac{\mathrm{d}\ell^-}{\ell^-} = -\lambda\rho\mathrm{d}t. \tag{11}$$

Finally, substituting from (5) into (6) to eliminate t^*, we get

$$\frac{\ell^+ + \alpha^+(k - \ell^+)}{\rho}\ln\frac{\ell^-}{\ell_0} = k\ln\frac{\ell^+}{\ell^-} + (1 - \alpha^+)(k - \ell^+)\ln\left(1 + \frac{\ell^+ - \ell^-}{\alpha^+(k - \ell^+)}\right), \tag{12}$$

The effect of the variation on the end points of split mass can be obtained by totally differentiating (12). The resulting equation, after (9)–(11) are substituted into it, must hold for any small $\mathrm{d}t$. This gives rise to another condition, which is too long and cumbersome to include here. This condition, together with the earlier observations, pins down $(t^*, T, \ell^-, \alpha^+, \ell^+)$ as functions of only one variable, $\bar{\ell}$. (Incidentally, one can vary $\bar{\ell}$ and trace out the locus of (ℓ^+, α^+) at which the atom splitting terminates under the optimal policy, yielding a dashed locus in Figure 2.)

Let $t^*(\bar{\ell}), \alpha^+(\bar{\ell}), \ell^+(\bar{\ell}), T(\bar{\ell})$ be the key variables of the optimal policy as functions of $\bar{\ell}$. The resulting welfare is:

$$\mathcal{U}(\bar{\ell}) := T(\bar{\ell})e^{-rt^*(\bar{\ell})}(\ell^0 - \ell^+(\bar{\ell}) - \alpha^+(t^*)(k - \ell^+(\bar{\ell}))) + \int_{T(\bar{\ell})}^{\infty} e^{-rt}(\ell^0 - \ell_t)\mathrm{d}t.$$

where

$$\ell_t := \bar{\ell}e^{-\lambda\rho(t - T(\bar{\ell}))}, \forall t \geq T(\bar{\ell}).$$

To characterize the optimal policy, it remains to choose $\bar{\ell}$ to maximize $\mathcal{U}(\bar{\ell})$. A closed form solution on the optimal $\bar{\ell}$, or a simple characterization, is difficult to come by.[6] But a numerical solution for the optimal $\bar{\ell}$, and thus the optimal policy, can be obtained. Figure 2 in the paper describes the optimal policy for the case $(r, \lambda, \rho, k, \ell^0) = (1/2, 2, 1/5, 15, 9)$.

The optimal policy employs a "blackout" until $t^* \simeq 1.45$, and the mass accumulated by then is then split at time t^* into (sequential) check in; this mass experiments along the locus that starts from $(\alpha^-, \ell^-) \simeq (0.017, 5.04)$ and ends at $(\alpha^+, \ell^+) \simeq (.066, 4.52)$. After time t^*, the agents check in as they arrive. They experiment along the locus of (α, ℓ)'s that begins at (α^+, ℓ^+) and tapers gradually down to $(\alpha, \ell) \simeq (0, 2.72)$. $\square$

2 General Signal Structure: Proof from Section 7.1

Here, we extend our model to allow for both good news and bad news. 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 of background learning.

The designer commits to the following policy: At time t, she recommends the movie to a fraction $\gamma_t \in [0, 1]$ of agents 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 to fraction $\alpha_t \in [0, 1]$ if no news has arrived by t. Clearly,

$$\mu_t = \rho + \alpha_t.$$

The designer's belief evolves according to

$$\dot{p}_t = -(\lambda_g - \lambda_b)\mu_t p_t(1 - p_t), \tag{13}$$

with the initial value $p_0 = p^0$. It is worth noting that the evolution of the posterior depends on the relative arrival rates of the good news and the bad news. 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 posterior rising. We label the former case **good news** case and the latter **bad news** case. Recall that main body of the paper treats the special case of $\lambda_b = 0$, a pure good news case.

Let g_t and b_t denote the probability that the designer's belief is 1 and 0, respectively.

[6] In particular, we have no proof that this constrained maximization admits a unique solution $\bar{\ell}$. There might be (presumably non-generic) parameter configurations for which this is the case, in which case there would be multiple optimal policies.

Given the experimentation rate μ_t, these probabilities evolve according to

$$\dot{g}_t = (1 - g_t - b_t)\lambda_g \mu_t p_t, \tag{14}$$

with the initial value $g_0 = 0$, and

$$\dot{b}_t = (1 - g_t - b_t)\lambda_b \mu_t (1 - p_t), \tag{15}$$

with the initial value $b_0 = 0$.[7] Further, these beliefs must form a martingale:

$$p_0 = g_t \cdot 1 + b_t \cdot 0 + (1 - g_t - b_t)p_t. \tag{16}$$

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: (13), (14), (15), and (16), where $\mu_t = \rho + \alpha_t$ is the total experimentation rate and r is the discount rate of the designer.[8]

Given policy (α, β, γ), conditional on being recommended to watch the movie, the agent 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 less than the cost, or

$$\frac{g_t \gamma_t + (1 - g_t - b_t)\alpha_t p_t}{g_t \gamma_t + b_t \beta_t + (1 - g_t - b_t)\alpha_t} \geq c. \tag{17}$$

The following is immediate:

Lemma 7. *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 (17) (and do not affect other constraints). $\square$

[7] 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, and to the first-order,

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

Dividing these equations by $\mathrm{d}t$ and taking the limit as $\mathrm{d}t \to 0$ yields (14) and (15).

[8] More precisely, the designer is allowed to randomize over the choice of policy (α, β, γ) (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.

Using $\ell_t = \frac{p_t}{1-p_t}$, (13) can be restated as:

$$\dot{\ell}_t = -\ell_t \Delta \lambda_g \mu_t, \quad \ell_0 := \frac{p_0}{1-p_0}, \tag{18}$$

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 8. *If $\Delta \neq 0$, then*

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

Proof. Let $\kappa_t := p_0/(p_0 - g_t)$. Note that $\kappa_0 = 1$. Then, it follows from (14) and (16) that

$$\dot{\kappa}_t = \lambda_g \kappa_t \mu_t, \quad \kappa_0 = 1. \tag{19}$$

Dividing both sides of (19) by the respective sides of (18), 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 (16). □

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.

Next, substitute g_t and b_t into (17) to obtain:

$$\alpha_t \leq \bar{\alpha}(\ell_t) := \min\left\{1, \frac{\left(\frac{\ell_t}{\ell_0}\right)^{-\frac{1}{\Delta}} - 1}{k - \ell_t}\ell_t\right\}, \tag{20}$$

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 9. *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 $\bar{\alpha}(\ell_t) \in [0, 1)$.*[9]

Proof. We shall focus on

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

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

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

$$1 - (\ell/\ell_0)^{\frac{1}{\Delta}} \geq 0 \text{ and } k\ell^{\frac{1}{\Delta}-1}\ell_0^{-\frac{1}{\Delta}} > 1. \tag{21}$$

Second,

$$\tilde{\alpha}'(\ell) = \frac{(\ell_0/\ell)^{\frac{1}{\Delta}} h(\ell, k)}{\Delta(k - \ell)^2}, \tag{22}$$

where

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

Third, (21) implies that

$$\frac{\mathrm{d}h(\ell, k)}{\mathrm{d}\ell} = 1 - k\ell^{\frac{1}{\Delta}-1}\ell_0^{-\frac{1}{\Delta}} < 0, \tag{23}$$

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

$$h(0, k) = -k(1 - \Delta) = -k\frac{\lambda_b}{\lambda_g} \leq 0. \tag{24}$$

It follows from (23) and (24) that $h(\ell, k) < 0$ for any $\ell \in (0, k)$ and $\tilde{\alpha}(\ell) \in [0, 1)$. By (22), this last fact implies that $\tilde{\alpha}'(\ell) < 0$ if $\Delta > 0$ and $\tilde{\alpha}'(\ell) > 0$ if $\Delta < 0$, as was to be shown. □

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

Substituting the posteriors from Lemma 8 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

$$\dot{\ell}_t = -\Delta \lambda_g (\rho + \alpha_t) \ell_t, \tag{25}$$

proposition

$$0 \leq \alpha_t \leq \bar{\alpha}(\ell_t). \tag{26}$$

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.

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

The analysis is similar to that for Proposition 1 in the paper. As in the paper, we first switch the roles of variables so that we treat ℓ as a "time" variable and $t(\ell) := \inf\{t | \ell_t \leq \ell\}$ as the state variable, interpreted as the time it takes for a posterior ℓ to be reached. Up to 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 $i = SB, FB$,

$$\sup_{u(\ell)} \int_0^{\ell^0} e^{-rt(\ell)} \ell^{\frac{1}{\Delta}-1} \left(1 - \frac{k}{\ell} - \left(\rho\left(1 - \frac{k}{\ell}\right) + 1\right) u(\ell)\right) \mathrm{d}\ell, \tag{27}$$

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 $\mathcal{U}^{SB}(\ell) := [\frac{1}{\rho+\bar{\alpha}(\ell)}, \frac{1}{\rho}]$ 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).

The characterization of the solution is summarized in the proposition which extends Proposition 1 of the paper for the general good news case.

Proposition 1. *The second-best policy prescribes, absent any news, the maximal experimentation at $\alpha(p) = \bar{\alpha}(\frac{p}{1-p})$ 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 No Social Learning nor Full Transparency can. If $p_0 < c$, then the second-best induces a slower experimentation/learning than the first-best.

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

$$\mathcal{H}(t, u, \ell, \nu) = e^{-rt(\ell)} \ell^{\frac{1}{\Delta}-1} \left(1 - \frac{k}{\ell} - \left(\rho\left(1 - \frac{k}{\ell}\right) + 1\right) u(\ell)\right) - \nu \frac{u(\ell)}{\Delta\lambda_g \ell}. \tag{28}$$

The necessary optimality conditions state that there exists an absolutely continuous function

$\nu : [0, \ell^0]$ such that, for all ℓ, either

$$\phi(\ell) := \Delta\lambda_g e^{-rt(\ell)} \ell^{\frac{1}{\Delta}} \left(\rho \left(1 - \frac{k}{\ell}\right) + 1\right) + \nu(\ell) = 0, \tag{29}$$

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

Furthermore,

$$\nu'(\ell) = -\frac{\partial \mathcal{H}(t, u, \ell, \nu)}{\partial t} = re^{-rt(\ell)} \ell^{\frac{1}{\Delta}-1} \left(\left(1 - \frac{k}{\ell}\right)(1 - \rho u(\ell)) - u(\ell)\right) \; (\ell - \text{a.e.}). \tag{30}$$

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

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',

$$\phi'(\ell) = e^{-rt(\ell)} \ell^{\frac{1}{\Delta}-2} \left(r(\ell - k) + \rho\Delta\lambda_g k + \lambda_g (\rho(\ell - k) + \ell)\right), \tag{31}$$

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,

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

Also, $\phi(0) \leq 0$ (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 $1/(\rho + \bar{\alpha}(\ell))$ 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_g^* (and establish uniqueness in the process).

For $\ell < \ell_g^*$,

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

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 $C_1 > 0$; or else $C_1 = 0$ and $t(\ell) = \infty$ for every $\ell \in (0, \ell_g^*)$, 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). This amounts to showing that 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 concave in t (the state variable), for all ℓ. To this end, it suffices to show that the terms inside the big parentheses in (28) are negative for all $u \in \mathcal{U}^i$, $i = FB, SB$. This is indeed the case:

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

where the inequality follows from the linearity of the expression in $u(\ell)$ and the fact that $u(\ell) \in \mathcal{U}^i \subset [\frac{1}{\rho+1}, \frac{1}{\rho}]$, for $i = FB, SB$. The concavity of maximized Hamiltonian in t, and thus sufficiency of our candidate optimal solution, then follows. □

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

The analysis is qualitatively the same for the general bad news case. 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 $\mathcal{U}^{FB}(\ell) := [\frac{1}{\rho+1}, \frac{1}{\rho}]$. Again, a solution exists Filippov-Cesari theorem (Cesari, 1983).

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 $\bar{\alpha}(\frac{p}{1-p})$ 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.

Proof. 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

$$\psi(\ell) := -\phi(\ell) = \Delta\lambda_g e^{-rt(\ell)} \ell^{\frac{1}{\Delta}} \left(\rho\left(1 - \frac{k}{\ell}\right) + 1\right) - \nu(\ell) = 0, \tag{32}$$

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 (30). 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 (32) that

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

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 (32).

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

$$\int_{\ell_b^*}^{\infty} e^{-rt^{**}(\ell)} f(\ell)\mathrm{d}\ell \geq 0, \tag{33}$$

(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

$$\int_{\ell_b^*}^{\infty} e^{-rt(\ell)} f(\ell)\mathrm{d}\ell = 0. \tag{34}$$

Multiplying $e^{rt^{**}(\tilde{\ell})}$ on both sides of (33) gives

$$\int_{\ell_b^*} e^{-r(t^{**}(\ell)-t^{**}(\tilde{\ell}))} f(\ell)\mathrm{d}\ell \geq 0. \tag{35}$$

Likewise, multiplying $e^{rt(\tilde{\ell})}$ on both sides of (34) gives

$$\int_{\ell_b^*}^{\infty} e^{-r(t(\ell)-t(\tilde{\ell}))} f(\ell)\mathrm{d}\ell = 0. \tag{36}$$

Subtracting (35) from (36) gives

$$\int_{\ell_b^*} \left(e^{-r(t(\ell)-t(\tilde{\ell}))} - e^{-r(t^{**}(\ell)-t^{**}(\tilde{\ell}))}\right) f(\ell)\mathrm{d}\ell \leq 0. \tag{37}$$

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 (37).

For sufficiency, the same argument as with Proposition 1 establishes that the maximized Hamiltonian will necessarily be concave in t, which implies optimality of our candidate solution, by Arrow's sufficiency theorem. □

2.3 "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 $\bar{\alpha}_t$ if $p_0 \geq p_0^*$, and no experimentation if $p_0 < p_0^*$, where $p_0^* := p_g^* (= p_b^{**})$ and $\bar{\alpha}_t$ (given in the Appendix) is increasing and convex in t and reaches 1 when $t^* = \frac{k-\ell_0}{\lambda_b(\ell_0 - k\rho)} \ln \frac{\ell_0}{k\rho}$. The first-best policy has the same structure with the same threshold posterior, except that $\bar{\alpha}_t$ is replaced by 1. The first-best is implementable if and only if $p_0 \geq c$ or $p_0 < p_0^*$.*

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

$$\begin{aligned}
\mathcal{W} &= \int_{t\geq 0} e^{-rt}\left(g_t(1-c) + \frac{p_0 - c}{p_0}\alpha_t(p_0 - g_t)\right)\mathrm{d}t \\
&= \int_{t\geq 0} e^{-rt}\left(g_t(1-c) + \frac{p_0 - c}{p_0}\left(\frac{\dot{g}_t}{\lambda_g} - (p_0 - g_t)\rho\right)\right)\mathrm{d}t
\end{aligned}$$

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

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

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

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 $\bar{\alpha}_t$ from the incentive compatibility constraint, plug back into the differential equation for g_t and get, by solving the ordinary differential equation,

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

and finally

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

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). □

3 Heterogeneous Costs: Proofs from Section 7.2

3.1 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 $\bar{c} := q_H c_H + q_L c_L$. 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 $\eta : [0, \ell_0] \to \mathbf{R}$, 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.

We now characterize the threshold beliefs under first-best and second-best policies. Throughout, we shall use superscript $**$ to denote the first-best and superscript $*$ to denote the second-best policy. By the principle of optimality, the threshold ℓ_L must coincide with $\ell^* = \ell^{**}$ in the case of only one type of regular consumers (with cost c_L). To compare ℓ_H^* and ℓ_H^{**}, we proceed as in the bad news case, by noting that, in either case,

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

and

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

Hence,

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

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 $\ell_H^* < \ell_H^{**}$. □

3.2 Uniform Cost: Derivation of the Optimum

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 $t \leq t_1$, or $\gamma_t \geq \ell_0$. 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

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

This suffices to represent the solution, as we have done in Figure 5 in the main body. □

3.3 Proof of Proposition 5

We allow the designer 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.

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

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 1 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.

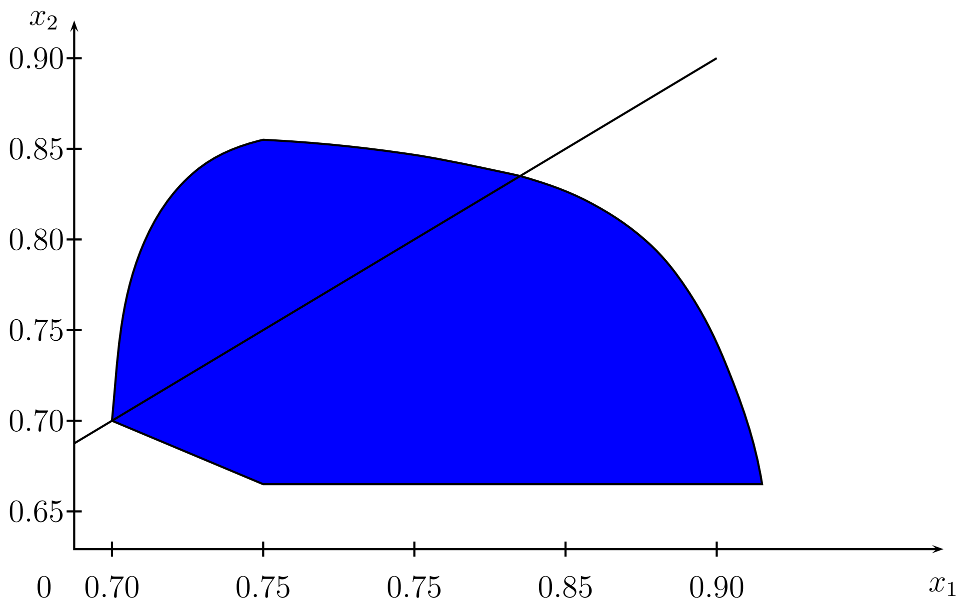

Figure 1: Region B_2 of feasible (x_1, x_2) in the case $J = 2$ (here, for $\ell_0 = 5, \ell_1 = 3, \ell_2 = 2, \mu_1 = 2/3$.

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, Theorem 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

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

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. $\square$

References

[1] Ausubel, L., 2004. "An Efficient Ascending-Bid Auction for Multiple Objects," *American Economic Review*, **94**, 1452–1475.

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

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

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

[5] Simon, L.K. and M. Stinchcombe, 1989. "Extensive Form Games in Continuous Time: Pure Strategies," *Econometrica*, **57**, 1171–1214.