Rewarding Failure

Aditya Kuvalekar* Nishant Ravi †

June, 2019

Abstract

We explore when and how to reward failure in a dynamic principal-agent relationship with experimentation. The agent receives flow rents from experimentation, and divides his time between searching for evidence of success and failure about the underlying project. The principal commits in advance to rewards conditional on the type of evidence. At each instant, the principal makes a firing decision. We show that the principal's optimal equilibrium features a stark reward structure—either the principal does not reward failure at all or rewards success and failure equally.

Keywords: dynamic agency, experimentation

JEL codes: C73, D83, D86, M51.

1. Introduction

Rewarding employees for failed ideas is becoming increasingly common. For example, Google X, an ambitious R&D division of Google, was in the news for this practice.[1] Google X is not alone in doing so. P&G has a "heroic failure award", TATA, an Indian conglomerate, has a "dare to try award".[2] A common ingredient in these situations is that employees have the freedom to develop innovative ideas to assess their potential without a fear of failure. In an interview to BBC, Astro Teller of Google X says, "*If you*

*Universidad Carlos III de Madrid. Email: akuvalek@eco.uc3m.es

†University of Pennsylvania. Email: nishant@upenn.edu

‡Nishant Ravi is deeply indebted to his advisers George Mailath, Steve Matthews, and Mallesh Pai for constant guidance and encouragement. We would like to thank Heski Bar-Isaac, Aislinn Bohren, Yan Chen, Joyee Deb, Rahul Deb, William Fuchs, Marina Halac, Rohit Lamba, Annie Liang, Elliot Lipnowski, Andrew Postlewaite, Maher Said, Johannes Schneider, Rakesh Vohra, Yuichi Yamamoto and audience members at University of Carlos III de Madrid and University of Pennsylvania,. Aditya Kuvalekar gratefully acknowledges support from the Ministerio Economia y Competitividad, Maria de Maeztu grant (MDM 2014-0431), and Comunidad de Madrid, MadEco-CM (S2015/HUM-3444).

[1]For example, https://www.bbc.com/news/technology-25880738

[2]https://www.forbes.com/sites/jacobmorgan/2015/03/30/why-failure-is-the-best-competitive-advantage/ e.g.

don't reward failure, people will hang on to a doomed idea for fear of the consequences. That wastes time and saps an organisation's spirit."

In fact, when employees enjoy the freedom to develop their ideas, rewards for failures may be all the more important to incentivize them to *look for failure*. For example, consider a computer scientist working on building a prediction algorithm that improves on existing algorithms by $x\%$. One approach may be to try different machine learning strategies towards "success", i.e. meeting the target. But, what if such an improvement is theoretically impossible? For example, a contest called "Heritage Health Prize" was supposed to award $3mn for a prediction algorithm for patient hospitalization. However, after two years and over 35,000 entries, reportedly, no team managed to achieve the target error of 0.4 or lower.[3] Could this be because this error was theoretically impossible to achieve? Proving so would require a completely different strategy—that of providing a theoretical proof of the impossibility. It seems natural that, upon furnishing such a proof of "failure", the scientist should be rewarded.

Despite this seemingly compelling rationale, the practice of rewarding employees for failures is far from ubiquitous. One reason could be that rewarding employees for failed projects imposes an additional cost for an organization that can be justified only if the gain from a successful project is sufficiently large. For example, talking about the ideas undertaken at Google X, Teller says, "*...these ideas are about huge, transformative, disruptive change, not the marginal, incremental change of a conventional business.*".

So, why, when and how should an innovative organization reward an employee for failure? These questions are the focus of our paper.

Toward this goal, we develop and study a novel dynamic principal-agent framework facing a project of unknown feasibility. We answer the questions posed above. First, the reason to reward failures is to incentivize employees to look for failures. Second, the costs of such rewards may be worth incurring if, for example, the upside potential of the underlying project is large, but not otherwise. Also, it may be optimal to reward employees for failure if the employer and the employee begin their interaction with high optimism about the project's feasibility, but not otherwise. Third, the organization should either not reward employees for failure at all or reward success and failure equally.

Summary of the framework and results: We study a continuous time interaction between an employer, the principal and an employee, the agent who face a project of unknown quality, either good or bad. Both players are equally informed about the project quality. At the start, the principal commits to two rewards—a reward amount for "success"—a conclusive evidence that the project is good, (e.g. algorithm that meets the improvement target) and a reward for "failure"—a conclusive evidence that the project is bad (e.g. theoretical proof of the impossibility of an improvement). After accepting the reward structure, the agent executes his innovation strategy—deciding

[3] See http://blog.kaggle.com/2013/06/03/powerdot-awarded-500000-and-announcing-heritage-health-prize-2-0/ and https://www.prnewswire.com/news-releases/hpn-announces-team-powerdot-wins-500000-as-current-leader-209924971.html

how to split his time between looking for success and failure. We model this as splitting his unit resource (time) at each instant between two exponential bandits, called arms. One is a "success" arm that produces a signal "(S)uccess" at an arrival rate proportional to the resources allocated to it if and only if the project is good. The other is a "failure" arm that produces a signal "(F)ailure" at an arrival rate proportional to the resources allocated to it if and only if the project is bad. A signal on either arm reveals the project quality and ends the game. Signals are public, but the agent's allocation choice is not observable to the principal. A success provides the knowledge needed to implement the project which results in a lump-sum payoff to the principal. Failure is costless in and of itself but the principal pays the reward upon failure to the agent. The principal can terminate the relationship at any instant by firing the agent. Also, flow costs of experimentation are borne by the principal while the agent earns flow rents while experimenting.

Both parties have the same discounting rate and have outside options that are normalized to zero. We are interested in the pure strategy Nash equilibria—principal's firing strategy (a deadline of firing the agent absent success or failure) and agent's experimentation strategy, that are mutual best responses. In particular, we focus on the principal optimal equilibria—equilibria that deliver the highest value to the principal.

The initial reward structure has two restrictions: First, the agent cannot be forced to pay the principal under any circumstance. Second, the agent's reward upon obtaining success is at least as much as the flow rent of the agent. This assumption is motivated by our interpretation of success as the principal adopting the project and employing the agent to implement it. After having produced success, the agent continues to receive at least the flow rent he receives during experimentation implying that the reward to the agent on producing success cannot be less than the flow rent.[4]

We briefly discuss the fundamental ingredient of our modeling approach. Searching for success or failure are two distinct activities that produce conclusive evidence. Note that this framework makes it fundamentally different from the majority of literature on agency problems with experimentation where, typically, conclusive evidence of only one type is available.[5] We have in mind situations like the following: Consider a computer scientist trying to build an algorithm with a set target or trying to prove, theoretically, that the target is infeasible, or a person investigating a crime trying to find incriminating evidence such as a camera footage establishing a person's guilt, or an irrefutable alibi establishing innocence. Or, imagine a research assistant helping a professor on a conjecture that one can either prove or disprove via a counterexample. The common features in these examples are that there is an objective state of the

[4]Since we normalize the rate of discounting to 1, the present discounted value of receiving the flow rent in perpetuity is the flow rent itself.

[5]For example, Keller et al. (2005) and the subsequent literature on fully revealing "success", Keller and Rady (2015) for conclusive "failures". An exception is Che and Mierendorff (2016) who study a single agent decision problem with both, conclusive success and failure technologies. We discuss it in detail later.

world (true or false, innocent or guilty) and the strategies towards proving either state are fundamentally different.

We first explore when will the agent be willing to look for failure? In this regard, Proposition 3 shows that, whenever the reward for failure is strictly smaller than the flow rent, the agent will not search for failure on-path in any equilibrium. The intuition behind the result is the following. For any such reward and a deadline T by the principal, the agent's best response has a simple structure:[6] look for success up to some $t_1 \leq T$ and, if no success arrives until then, switch to looking for failure for the remaining time.[7] Given this agent behavior, the principal would rather fire the agent at t_1 as she only incurs costs (flow costs of experimentation, reward for failure) after t_1 with no benefits.

Proposition 3 shows that, consistent with the prevailing arguments behind rewarding failures, it is indeed the case that the employees may "*hang on to a doomed idea for fear of the consequences*" (termination in this context). However, it also shows that if we must reward failures, such a reward must be at least as much as the flow rent, i.e. like an employment protection.

To complete the analysis, we depart from the game momentarily to ask the following question: If the principal could choose the experimentation strategy herself with the condition that the reward must equal the flow rent for either a success or a failure, what should be her optimal strategy, henceforth policy? We answer this question in Proposition 4.[8] Unsurprisingly, the optimal policy is Markov in players' belief (probability that the project is good) and can take one of three forms in fig.1. Notice that, as the arrows indicate, when looking for success (failure), no signal makes the players update less (more) favourably about the project quality. As a result, beliefs move down (up).

Proposition 5 shows that each of the above can be implemented as an equilibrium. In (1), the "S-only policy", the principal looks for success until the beliefs drift below a cutoff (p_S) where she quits to take her outside option. The principal can implement this outcome by offering no reward for failure (Proposition 2).

In (2), the "FS policy", the principal looks for failure when the beliefs are between $\underline{p}$ and p^f, and looks for a success above p^f. At p^f, she "freezes beliefs"—allocates resources across looking for success and failure in a way that, absent any signal, the beliefs remain there. Therefore, conditional on reaching p^f, the belief does not move until one of success or failure arrives. To implement this outcome, the principal would

[6] That a best response exists and has this structure is not obvious. However, we omit the discussion of the intuition here. See Lemma 3 and the discussion therein for the reasoning.

[7] To be precise, this is true only if T is finite. If T is infinite, the agent will never search for failure. In that case, Proposition 1 yields that employing the agent forever will not be an equilibrium.

[8] We omit the technical details about the characterization of the optimal policy for brevity. But, we would like to point out that rather than reasoning through the usual HJB equations, we prove optimality using necessary conditions for an optimal control, whose existence is guaranteed by standard results in the theory of stochastic control. Interested reader can follow the discussion after Proposition 4 for more details.

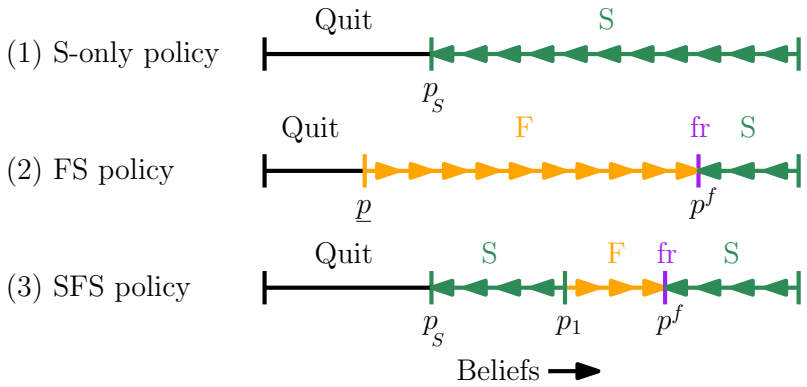


Figure 1: Three possible optimal policies

set the reward for success and failure equal to the flow rent and offer an infinite firing deadline. This way, the agent is indifferent across any allocation, and therefore, can follow the principal optimal one.

In (3), the "SFS policy", the principal looks for success in two regions—on $(p^f, 1)$ and $(\underline{p}, p^f)$. In the middle region (p_1, p^f), she looks for failure. Moreover, at p^f, the principal reezes beliefs. Here, the implementation as an equilibrium depends on the initial belief. Notice that if the initial belief is above p_1, the principal's behavior is as in the FS policy and, therefore, the implementation is the same as above. On the other hand, if the initial belief is below p_1, then the optimal policy is to keep looking for success. As before, this can be implemented by not rewarding failure.

Therefore, either the principal should reward success and failure equally or not reward failure at all. Moreover, it may be optimal for the principal to reward for failure if the initial prior is sufficiently high (as in the SFS case) but not otherwise.

Coming back to the points made at the beginning, Proposition 7 confirms Teller's insight. If the gain from a success of a good project is sufficiently high then it is optimal to reward the agent for failure but not otherwise.[9] This may perhaps be the reason why, except for giants like Google X, most other companies cannot afford to reward for failures.

In fact, there may be another reason why we are witnessing more companies rewarding failures. Many modern industries, even in manufacturing, rely on simulating complex systems before deploying them. Simulations are a much faster way of learning about possible flaws or a failure. In the context of our model, this translates to a high arrival rate of the failure arm. And indeed, as Proposition 8 shows, if the failure arm has a sufficiently high arrival rate then it may be optimal to reward for failures. This may also, perhaps, be a reason behind why the "bug bounty programs"—incentives to employees to find bugs in codes—are gaining popularity.[10] One could view such

[9] To be precise, it also requires that the failure arm has a sufficiently high arrival rate and the flow cost of experimentation is high enough.

[10] For example https://techcrunch.com/2018/02/07/googles-bug-bounty-programs-paid-out-almost-3m-in-2017

programs as incentives for employees to report failures. in the form of bugs. One could argue that finding bugs or flaws in a code is easier, captured through higher λ_b, compared to finding a flaw in some component of a manufacturing unit.

Below, we briefly discuss the related literature before presenting the model and the analysis. Most proofs are relegated to the appendix.

Related Literature: On the problem of rewarding the agent for failure, the literature has focused on incentivizing the agent to reveal failure that he observes privately. For example, Levitt and Snyder (1997) show that rewarding for failure may be optimal when the agent receives a private signal about the project quality. Hidir (2017) and Chade and Kovrijnykh (2016)) are examples of dynamic contracting problems where the agent has the freedom to disclose negative news. We complement this literature by showing that, even though both actions and signals are public, the inability of the parties to write contracts contingent on actions can deter the agent from searching for failures. Note that the choice of specifically searching for failure is absent in the above mentioned papers. Manso (2011) shows that, in a two period setting with full commitment, motivating an agent to innovate may require tolerating or even rewarding early failures. Like the ones mentioned above, this model also does not allow for a technology to search for failures.

Our model builds on the exponential bandit models of Keller et al. (2005) and Keller and Rady (2015), which study good and bad news arms (success and failure arms in our context) resepectively.

Technically, the paper closest to ours is Che and Mierendorff (2016), henceforth CM. They study a single agent decision problem (as opposed to a two player game we have) of experimentation where the agent has the choice to look for good news and bad news. Besides the game, our single agent setting (Proposition 4), is also fundamentally different in a an important way. In CM, the decision maker can "adopt a project" by quitting at any instant. The decision maker receives the expected value of the project based on her belief at that instant. In our model, quitting without a signal yields the outside option regardless of the beliefs. This difference generates substantially different dynamics, e.g. the contradictory learning in their model does not happen in our single agent version.

In a related single agent decision problem, Damiano et al. (2017) introduce an auxiliary learning process that allows for looking for both good and bad news while experimenting on a one arm bandit in lines of Keller et al. (2005).

Garfagnini (2011) and Guo (2016) also study a delegation game between a principal and an agent where the agent carries out experimentation. While the contracting and payoff environment differs, the key distinction is our focus on how the agent's incentives shape the dynamics when the choice of both good and bad news is available. This tradeoff is absent in both Garfagnini (2011) and Guo (2016). As an agency problem of collective experimentation, this paper also relates to Kuvalekar and Lipnowski (2018). However, the efforts there are ranked in the sense of Blackwell (1953) making the

agent's choice, when not getting fired, straightforward—choose the least informative action. Since the success and failure arms are not ranked in the sense of Blackwell (1953), the dynamics are richer in our environment. Halac et al. (2016), Bergemann and Hege (2005) and Hrner and Samuelson (2013) are other instances of contracting problems with delegated expeirmentation with moral hazard and (or) adverse selection.

Recently, the question of information acquisition in the presence of multiple information sources has been pursued among others by Che and Mierendorff (2016), Liang et al. (2017), Liang and Mu (2018), Fudenberg et al. (2017), and Mayskaya (2017). In contrast, in this paper we explore information acquisition from multiple sources of information in a principal-agent setting where the incentives of the two parties differ.

2. Model

Players: There are two players, a principal (she) and an agent (he). Time t is continuous and runs from 0 to ∞. The principal hires the agent to work on a project of unknown quality. The quality of the project is good, $\theta = 1$, or bad, $\theta = 0$. At time 0 both players have a common prior on the underlying project quality: $\mathbb{E}_0\theta = p_0 \in (0, 1)$.

Actions: At each instant, the principal chooses whether to fire ($a_t = 0$) or not to fire the agent ($a_t = 1$). Firing is irreversible and ends the game. Conditional on not firing, the agent divides a unit resource between a "success" arm and a "failure" arm. The agent's allocation to the success arm at time t is $\gamma_t \in [0, 1]$, and $(1 - \gamma_t)$ is the allocation to the failure arm.

Information: The principal *does not observe* the agent's allocation choice. The agent's allocation affects the arrival rate of two exponentially distributed signals (news). The success arm can generate a signal called S (success). The failure arm can generate a signal F (failure). The arrival rate of an S signal is $\lambda_g\gamma_t\theta$, and that of an F signal is $\lambda_b(1 - \gamma_t)(1 - \theta)$. Both signals are publicly observed. Also, notice that either signal, S or F is conclusive: the realization of S(F) gives both players the belief $p = 1(p = 0)$. We denote by $y_t \in \{\phi, S, F\}$ the signal up to time t, where ϕ denotes no signal. Note that since the agent's allocation is not observed by the principal, players may have different posterior belief about θ conditional on no signal.

Payoffs: At the beginning of the relationship, the principal commits to a reward structure which specifies a payment of R^S to the agent if an S signal arrives and R^F if an F signal arrives. When employed, the agent receives an exogenously specified fixed flow wage $w > 0$ from the principal. The principal incurs a flow cost of $c > w$ which we interpret as the cost of performing experimentation and the wage paid to the agent. If S arrives, the game ends with the principal receiving a lump-sum payoff of Γ. If F arrives, the game ends with the principal receiving a lump-sum payoff of $-R^F$ due to the promised reward to the agent. Both players discount future payoffs at rate r. Counting time in different units we normalize r to 1.

The terminal payoffs are:

1. If principal fires the agent, both players receive 0.
2. If S arrives, the principal receives $\Gamma - R^S$ and the agent receives R^S.
3. If F arrives, the principal receives $-R^F$ and the agent receives R^F.

2.1. Strategies

Since the only relevant history is one involving no signal and the agent not having been fired, all the strategies are specified under this condition. The agent's strategy, which is his allocation to the success arm is a $[0,1]$ valued measurable process $(\gamma_t)_{t\geq 0}$, i.e. $\gamma : \mathbb{R}_+ \to [0,1]$ is a measurable map. Let $\mathcal{G}$ denote the space of such $[0,1]$ valued measurable maps. The principal's strategy is a *deterministic* deadline $T \in [0,\infty]$ at which she fires the agent.[11] If the agent plays $\gamma \in \mathcal{G}$ and the principal offers a deadline $T \in \bar{\mathbb{R}}_+$, we will denote this strategy profile by (γ, T).

2.2. Learning

Let $P_t := \mathbb{E}_{t,\gamma}\theta$ be the agent's posterior probability that $\theta = 1$ at time t. In the absence of an S or F signal, the agent's belief will evolve according to Bayes' rule:[12]

$$\frac{\mathrm{d}P_t}{\mathrm{d}t} = [(1-\gamma_t)\lambda_b - \gamma_t\lambda_g]P_t(1-P_t). \tag{1}$$

On the other hand, since the principal does not observe γ_t, she may have different belief about θ at time t, off-path. Suppose the principal expects the agent to play a strategy $\tilde{\gamma}$ Let $\mathbb{E}_{t,\tilde{\gamma}}\theta = \tilde{P}_t$. Similar to before, her beliefs evolve as in (1) with γ_t replaced by $\tilde{\gamma}_t$. Obviously, in equilibrium, $\gamma = \tilde{\gamma}$, and therefore $P_t = \tilde{P}_t$.

Note that using (1) we can show that $\dot{P}_t = 0$ when $\gamma_t = \gamma^f := \frac{\lambda_b}{\lambda_b+\lambda_g}$. That is, beliefs do not move in the absence of a conclusive signal if the agent allocates γ^f to the success arm. We call γ^f as the freezing allocation and when agent chooses γ^f at some belief p, we say that "the agent freezes beliefs at p". Similarly, if $\gamma_t = 1$ (0), we say that the agent is "looking for success (failure)".

2.3. Equilibrium

In equilibrium, it is necessary that $\tilde{\gamma} = \gamma$. Therefore, we will assume that to be the case here on. Consequently, $P_t = \tilde{P}_t$ for all t. Of course, in defining an equilibrium,

[11] To be precise, the principal fires the agent whenever $t \geq T$. Therefore, if the principal was supposed to fire the agent at T but did not until some $t > T$, we assume that she will fire the agent at t.

[12] Since beliefs are a martingale, we have that $\lambda_g\gamma_t P_t \mathrm{d}t + (1 - [\lambda_g\gamma_t P_t + \lambda_b(1-\gamma_t)(1-P_t)]\mathrm{d}t)(P_t + \dot{P}_t\mathrm{d}t) = P_t$. Dividing by $\mathrm{d}t$ we obtain (1).

we would need to ensure the optimality of γ from the agent's perspective given the deadline T and the reward structure (R, F).

Define, $\tau := \inf\{t \geq 0 : y_t \in \{S, F\}\} \wedge T$. Suppose $P_t = p$. Given the strategy profile (T, γ), the agent's expected payoff at time t is,

$$U(t, p, \gamma, T|R^S, R^F) := \mathbb{E}_{\gamma,p}\left[(1 - e^{-(\tau - t)})w\mathrm{d}u + e^{-(\tau - t)}\left[\mathbb{1}_{y_\tau = S}R^S + \mathbb{1}_{y_\tau = F}R^F\right]\right]$$

Similarly, the principal's expected payoff at time t is,

$$\Pi(t, p, \gamma, T|R^S, R^F) := \mathbb{E}_{\gamma,p}\left[(1 - e^{-(\tau - t)})(-c) + e^{-(\tau - t)}\left[\mathbb{1}_{y_\tau = S}(\Gamma - R^S) + \mathbb{1}_{y_\tau = F}(-R^F)\right]\right]$$

By dividing both players' payoffs by w, we can set, without loss of generality, $w = 1$.[13] We make the following assumptions.

Assumption 1. *We assume that $R^F \geq 0$ and $R^S \geq 1$.*

That is, any reward to the agent must be non-negative and in particular the amount the principal pays to the agent upon obtaining an S signal, is no less than the discounted value of the agent's wage. We interpret success as the principal adopting the project and employing the agent to work on it. The agent should thus continue to receive at least the flow rents he receives during the experimentation stage.

Assumption 2. $\frac{\lambda_g(\Gamma - 1) - c}{1 + \lambda_g} > 0$.

Assumption 2 says that if $\theta = 1$, the principal finds it worthwhile to employ the agent to experiment by paying a reward of 1, the lowest possible reward upon an S signal.

Assumption 3. $c > 1$.

The above assumption makes it possible for the agent to look for failure when $R^F = 1$ which, as we shall see, is the minimum reward the principal needs to give to incentivize the agent to search for failure. If $c \leq 1$, it means that the principal would rather have the agent search for success forever over offering a reward of 1 for failure.

Suppose $P_0 = p$. Given (T, R, F) the agent solves the following problem:

$$U^*(0, p, T, |R^S, R^F) := \sup_{\hat{\gamma} \in \mathcal{G}} U(0, p, T, \hat{\gamma}|R^S, R^F) \tag{AP}$$

$\gamma \in \mathcal{G}$ is a best response given (T, R^S, R^F) for the agent if $U(0, p, T, \gamma, |R^S, R^F) = U^*(0, p, T|R^S, R^F)$.[14] Similarly, deadline T is a best response for the principal given (γ, R^S, R^F) if,

$$T \in \mathrm{argmax}_{\hat{T} \in \mathbb{R}_+} \Pi(0, p, \hat{T}, \gamma, |R^S, R^F). \tag{PP}$$

[13] By doing so, the agent's wage becomes 1, while his terminal payoff becomes S/w and F/w depending on the signal. For the principal, the flow cost is c/w and the terminal payoffs are $\Gamma/w - S/w$ and $-F/w$ depending on the signal.

[14] By the dynamic programming principle, if a control is optimal at 0, then it is optimal at every $t > 0$.

Since the principal observes nothing other than a conclusive signal S or F which ends the game, she merely chooses a deadline to end the game in the absence of a signal *assuming* that the agent follows γ. That is, her problem is essentially a static problem.

Finally, we define the notion of (Nash) equilibrium in our setting.

Definition 1. *Suppose $P_0 = p$. Given a reward structure (R^S, R^F), an equilibrium is a strategy profile $(\gamma, T) \in \mathcal{G} \times \mathbb{R}_+$ such that:*

1. Agent optimality: *γ is a best response for the agent given (T, R^S, R^F)*
2. Principal optimality: *Deadline T is a best response for the principal given (γ, R^S, R^F).*

Define, $\mathcal{E}(R^S, R^F) := \{(\gamma, T) : (\gamma, T)$ constitutes an equilibrium given $(R^S, R^F).\}$, and $\Pi^*(p|R^S, R^F) := \sup_{(\hat{\gamma},\hat{T})\in\mathcal{E}(R^S,R^F)} \Pi(0, p, \hat{\gamma}, \hat{T}|R^S, R^F)$.

Notation 1. *Henceforth, we will drop the dependence on R^S, R^F and T for $U(\cdot)$ and $\Pi(\cdot)$ as the dependence will be obvious in each section.*

As mentioned in the introduction, the two main questions we seek to answer are the following:

1. Should the principal ever set $R^F > 0$? After all, R^F is a cost to the principal whose only use is to save future experimentation costs.
2. If yes, what should R^F be optimally?

Toward answering this question, we will first show, in Section 3.2, that the agent will not look for failure in any equilibrium if $R^F < 1$. Therefore, the choice for the principal is simple—either set $R^F = 0$ and make no use of the failure arm, or set $R^F \geq 1$ and have the agent look for failure sometimes.

3. Results

3.1. When the agent only looks for success

Suppose $R^S = 1, R^F = 0$. Let $v^1 := \Gamma - R^S$. We pose a simple question—what should the principal do if the agent was restricted to only looking for success, i.e. $\gamma_t = 1$ for all t? Let this agent strategy be denoted by γ^1. For the principal, this is a simple optimal stopping problem problem as in the planner's problem (Theorem 1) of Keller et al. (2005) (henceforth KRC).[15]

[15] There is a slight difference in the payoff structure as our success arm produces only one lump-sum payoff if $\theta = 1$ while in KRC the success (good news) arm produces multiple lump-sum payoffs if $\theta = 1$. The expressions, therefore, have to be appropriately adjusted for an exact mapping.

Define, $V_S : [0,1] \to \mathbb{R}$ as,

$$V_S(\cdot) := \sup_{\hat{T} \in \mathbb{R}_+} \Pi(0, \cdot, \gamma^1, \hat{T} | (1,0)).$$

This is a standard optimal stopping problem whose solution is simple—the principal would continue employing the agent whenever the beliefs are above a belief p_S and fire below. p_S is given below:

$$p_S := \frac{c}{\lambda_g v^1} \tag{2}$$

Proposition 1. *The optimal stopping time for the principal is defined as* $\tau := \inf\{s \geq 0 : P_s \notin (p_S, 1)\}$. *Principal fires the agent iff* $t > \tau$.

When $\gamma_t = 1$, the law of motion for beliefs, (1), yields $\dot{P}_t = -\lambda_g P_t(1 - P_t)\mathrm{d}t$. It is straightforward to verify that

$$T_G(p) := \frac{1}{\lambda_g}\left[\log\left(\frac{p}{1-p}\right) - \log\left(\frac{p_S}{1-p_S}\right)\right] \tag{3}$$

is the amount of time it takes for $P_t = p_S$ given $P_0 = p > p_S$ in the absence of an S signal.

3.2. When $R^S = 1, R^F < 1$

In analyzing this case, the first obvious question is, what would happen if there was no reward for failure ($R^F = 0$)? Lemma 1 characterizes the principal optimal equilibrium—the equilibrium where the principal obtains her highest payoff given $(R^S, R^F) = (1, 0)$. The equilibrium has an intuitive structure—the agent never searches for failure and the principal fires agent when the belief P_t reaches p_S. The proof of Lemma 1 is in the appendix but the underlying intuition is straightforward. Suppose the principal gives a deadline $T < \infty$. Notice that by searching for success for the entire duration, the agent is guaranteed to receive the flow wage of \$1 for the entire duration of T if $\theta = 0$. Moreover, if $\theta = 1$, with some probability an S signal will arrive yielding a value of 1 to the agent. On the other hand, if the agent allocates some fraction of his time searching for failure, then an F signal may arrive if $\theta = 0$. In this case, he will get terminated and would lose his subsequent wages. Moreover, by searching for failure, he also lowers the probability of receiving a success if $\theta = 1$. Therefore, the unique best response for any deadline is to keep searching for success until the deadline. Proposition 1 delivers that a deadline of $T_G(P_0)$ is optimal in this case.

Lemma 1. $\Pi^*(\cdot|1,0) = V_S(\cdot)$.

The natural follow-up question is, can the principal do better—induce search for failure—by offering a reward for failure? We answer this question in two steps. First, we analyze an environment with $R^F \in (0,1)$.

A first, reasonably obvious, result in this regard is when $T = \infty$. Lemma 2 shows that in this case, the agent's unique best response is to look for success, i.e. γ^1. The intuition is exactly as in Lemma 1. By searching for failure, if an F signal arrives, the agent forgoes some wages. On the other hand, by searching for success, the agent is guaranteed a payoff of 1 (infinite horizon flow wage discounted appropriately). Therefore, if the agent looks for a success only, the principal would best respond by having a deadline of T_G, i.e. (γ^1, ∞) cannot be an equilibrium *regardless* of the value of R^F so long as $R^F < 1$. We omit the proof of Lemma 2 in light of the similarity with the proof of Lemma 1.

Lemma 2. *When $R^F < 1 \leq R^S$, there is no equilibrium with $T = \infty$.*

However, what would happen if $T < \infty$? Since $R^S \geq 1 > R^F$, searching for success would still be more attractive if either signal was equally likely. But, when P_t is sufficiently low, this would not be the case. To see the intuition, let us for a moment think in discrete time of length $\Delta \approx 0$. Consider the "last period" before T, say, $t = T - \Delta$. Suppose there is no signal until then and the agent is deciding what type of news to look for. What If $\lambda_b(1 - P_t) >> \lambda_g P_t$? In the next "instant", an F signal is far more likely than an S signal. Naturally, the agent would rather collect an expected reward of $\lambda_b(1 - P_t)R^F\Delta$ than $\lambda_g P_t \Delta R^S$.

Therefore, unlike in the case with $R^F = 0$, even with small rewards, the principal may be able to induce a search for failure. However, as Proposition 2 shows, the principal cannot benefit from this. In any equilibrium with $T < \infty$, the agent exclusively searches for a success until getting fired.

Proposition 2. *Suppose $R^F < 1 \leq R^S$. If a (γ, T) equilibrium exists with $T < \infty$, it involves $\gamma_t = 1$ for all $t < T$.*

It would seem that Proposition 2 and the discussion before it—that the agent would find looking for failure lucrative when T is finite and $R^F \in (0, 1)$—contradict each other. But that is not quite the case. Indeed, Lemma 3 confirms the intuition that the agent will search for failure when $R^F \in (0, 1)$. But, it also shows that his best response has a rather simple structure: Look exclusively for a success until some time T_1 and then switch to looking exclusively for a failure.

Lemma 3. *Suppose $\gamma \in \mathcal{G}$ is a best response for (AP) with $R^F \in (0, 1)$ and $T < \infty$. Then, $\exists T' \leq T$ such that $\gamma_t = \mathbb{1}_{t \leq T'}$ for a.e. t.*

To see the intuition, consider an arbitrary agent strategy γ. Construct another strategy γ' such that the total allocation towards the success arm until T is the same in γ and γ', i.e $T_1 := \int_0^T \gamma_t dt = \int_0^T \gamma'_t dt$. However, all of that allocation is frontloaded in γ' by setting $\gamma'_t = \mathbb{1}_{t \leq T_1}$. That is, the agent exclusively searches for a success up to T_1. Notice that the ex-ante probability of receiving an S signal under γ and γ' is identical but, in expectation, an S signal would arrive earlier in γ' compared to γ. Equally importantly, the ex-ante probability of receiving an F signal before T is the same in γ and γ' *but* this signal would arrive later, in expectation, under γ' as

compared to γ because all the search for failure is backloaded as much as feasible. Since $R^F < 1$, collecting a wage of 1 up front is always more desirable than collecting R^F. Therefore, γ' delivers a strictly higher value compared to γ. Results in optimal control theory deliver that a best response exists to such a problem. Therefore, a best response must have the structure that the agent searches for a success until some time T_1 and then switches to searching for failure.

But then, how would the principal best respond to such agent behavior? Notice that retaining the agent after T_1 offers no benefits. At best the principal can know with certainty that the project is of low quality. In this case, besides incurring the additional cost of experimentation, the principal will also have to pay the reward R^F. Moreover, *regardless* of the project quality, the principal cannot earn any positive revenue after T_1. Therefore, the principal must fire the agent at T_1.

We emphasize though that this does not establish that (γ', T_1) is an equilibrium as γ' was a best response to a deadline T and not T_1. However, what Proposition 2 delivers is that should an equilibrium exist, it cannot have the agent searching for failure until getting fired. But then, if the on-path behavior by the agent is the same as γ^1—the agent strategy of always searching for success—then the best that the principal can achieve is $V_S(\cdot)$ as discussed in Section 3.1.

Below, we summarize the discussion so far in a simple proposition—the highest payoff the principal can attain when $R^S \geq 1, R^F < 1$ is $V_S(\cdot)$.

Proposition 3. *Suppose $R^S \geq 1$ and $R^F < 1$. Then, $\Pi^*(\cdot|R^S, R^F) \leq V_S(\cdot) = \Pi^*(\cdot|1, 0)$.*

3.3. When $R^S, R^F \geq 1$.

Put together, the results until now imply that the only way the principal can possibly induce the agent to search for failure is by offering a reward R^F that is at least 1. Notice that both R^S and R^F are costs from the principal's perspective. If she performed experimentation herself, she would set both these rewards as low as possible. We will make use of this obvious observation and compute the principal's optimal experimentation policy as a single agent decision problem with $R^S = R^F = 1$. Thereafter, we will show that indeed this can be implemented as an equilibrium.

3.3.1. A Decision Maker Problem

A decision maker (DM) faces a project of unknown quality θ. The project can be either bad ($\theta = 0$) or good($\theta = 1$). At each instant, the DM has a unit resource. She allocates a fraction a_t towards experimentation and $(1 - a_t)$ to the safe arm that yields a flow

payoff of 0.[16] The flow cost of experimentation per unit resource allocated is c. Out of the resources allocated towards experimentation, the DM allocates a fraction γ_t to the success arm which produces an S signal at an instantaneous arrival rate of $\lambda_g \theta a_t \gamma_t$. The remaining $(1-\gamma_t)$ fraction of the allocated a_t is invested towards the failure arm which produces an F signal at an instantaneous arrival rate of $\lambda_b(1-\theta)a_t(1-\gamma_t)$. Either signal, S or F, reveals the project quality. The DM receives a value v^1 upon an S signal and a value of v^0 upon an F signal. In the context of our problem, $v^1 = \Gamma - R^S = \Gamma - 1$ and $v^0 = -R^F = -1$.

Let $P_t := \mathbb{E}_t\theta$ as before, and $\mathcal{Y} := [0,1]^2$. As before, a S(F) signal at time t conclusively establishes that $\theta = 1$ (0), i.e. $P_t = 1$ (0). In the absence of a signal, beliefs evolve according to (1). Let, $g(p) = pv^1 + (1-p)v^0$ and $\tau := \inf\{t \geq 0 : P_t \notin (0,1)\}$.

Definition 2. *A control $\alpha := (\alpha_t)_{t\geq 0} = ((a_t, \gamma_t))_{t\geq 0}$ is a U valued process such that $\alpha : \mathbb{R}_+ \to U$ is measurable. The space of admissible controls is denoted by $\mathcal{U}$.*

Let,

$$J(p,\alpha) := \mathbb{E}_\alpha \left[\int_0^\tau e^{-t} a_t(-c)\mathrm{d}t + e^{-\tau} g(P_\tau) \right]$$

be the DM's expected payoff by following a control α given that $P_0 = p$.

The DM solves the problem below.

$$V(p) := \sup_{\alpha \in \mathcal{U}} J(p,\alpha) \tag{DMP}$$

A control α^* is called an optimal control if $V(p) = J(p,\alpha^*)$. We will often call a control as policy.

In order to present our main result—the characterization of the optimal policy—we need to define four important policies. Notice that the policies below are Markov in beliefs.[17]

1. **Freezing policy:** We denote it by $\alpha^f = (1, \gamma^f)$. As the name suggests, here, regardless of the belief the DM chooses $\gamma_t = \gamma^f$ until the uncertainty is resolved by the arrival of an S or an F signal. In the appendix, we prove that the $V^f(\cdot) := J(\cdot, \alpha^f)$ is affine in p and is given by (6).
2. **S only policy:** We denote it by α^S. Formally, $\alpha^S := (a_t, \gamma_t) = (1,1)$ if $P_t \geq p_S$ and $(0,\cdot)$ otherwise. Recall that $J(\cdot, \alpha^S) = V_S(\cdot)$.
3. **FS policy:** This is denoted by $\alpha^{FS}_{p_1}$. First, $a_t = 1$ for all t and γ is described as follows:

[16] Notice that such a choice of allocating resources between experimentation and the outside option was absent in the baseline model. Therefore, what the decision maker can achieve in this problem is an upper bound to the principal's payoff. We will then show that we can implement the same as an equilibrium of the game.

[17] A control/policy $\alpha \in \mathcal{U}$ is Markov in belief, in short Markov, if $\exists$ a function $g : [0,1] \to U$ such that $\alpha_t = g(P_t)$ for all t.

- Look for F ($\gamma_t = 0$) if $P_t < p_1$
- Look for S ($\gamma_t = 1$) if $P_t > p_1$.
- Freeze ($\gamma_t = \gamma^f$) if $P_t = p_1$.

Call the above policy as $\alpha^{FS}_{p_1}$. Let $\underline{p} := \sup\{p : J(p, \alpha^{FS}_{p_1}) \leq 0\}$.[18] Notice that conditional on reaching p_1, the beliefs stay there until the uncertainty is resolved.

Of particular importance amongst such policies is the "FS with a switch at p^f", where p^f is given in (7). The intuition behind p^f is the following: Following a FS with a switch at p_1 policy, the DM receives the freezing value at p_1. Therefore, $J(p_1, \alpha^{FS}_{p_1}) = V^f(p_1)$. p^f is the unique belief where we also have $J'(p_1, \alpha^{FS}_{p_1}) = V^{f'}(p_1)$. Let $V_{FS}(\cdot) := J(\cdot, \alpha^{FS}_{p^f})$.

4. **SFS policy:** This is denoted by $\alpha^{SFS}_{p_1,p_2}$. First, $a_t = \mathbb{1}_{P_t > p_S}$. γ is described as follows:
   - Look for S ($\gamma_t = 1$) if $P_t \in [0, p_1) \cup (p_2, 1]$.
   - Freeze if $P_t = p_2$, ($\gamma_t = \gamma^f$).

Notation 2. *We will sometimes refer to the components of a specific policy using the relevant superscript and subscript. For example, the γ part of $\alpha^{FS}_{p^f}$ will be referred to as $\gamma^{FS}_{p^f}$*

Figure 2 plots the three policies—S only, SFS and FS, below as functions of beliefs. The arrows indicate the direction in which the beliefs move when choosing the associated action. "fr" stands for freezing.

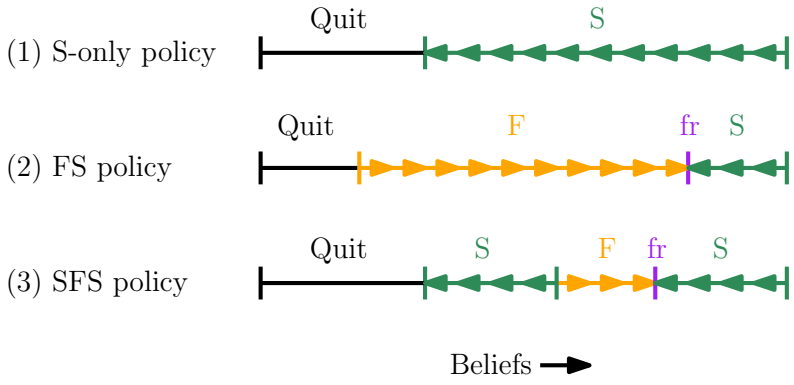


Figure 2: S, FS and SFS policies

Remark 1. *Notice that a SFS policy is either a S only policy or FS policy depending on the initial belief P_0.*

Below, we present a complete characterization of the optimal control. Unsurprisingly, there exists a Markov optimal policy. Therefore, we will describe an optimal

[18] Expressions of $J(p, \alpha_1)$ can be found in the appendix and are straightforward to compute, making $\underline{p}$ well defined.

policy as a function of the belief P_t.

Proposition 4. *An optimal policy to (DMP) exists and is Markov in beliefs. The optimal value function is given by,*

$$V(\cdot) = \max\{V_S(\cdot), V_{FS}(\cdot)\}$$

Moreover, if $V(p) = V_{FS}(p) \implies V(q) = V_{FS}(q)$ *for all* $q \geq p$.

For any optimal policy, $a(p) = \mathbb{1}_{V(p)>0}$. *An optimal policy* $\gamma^* : [0,1] \to [0,1]$ *can belong to one of three cases.*

1. ***S only policy:*** *If* $V_S(p^f) \geq V_{FS}(p^f)$, *then the optimal policy* $\alpha^*(\cdot) = \alpha^S$ *(fig.(3a)).*
2. ***FS policy:*** *If* $V_S(p^f) < V_{FS}(p^f)$ *and* $V_{FS}(p_S) \geq 0$, *then the optimal policy is* $\alpha^*(\cdot) = \alpha^{FS}_{p^f}$ *if* $P_0 \geq \underline{p}$ *and* $\alpha^*(\cdot) = (0, \cdot)$ *otherwise. That is, we follow* $\alpha^{FS}_{p^f}$ *if the initial prior is above* $\underline{p}$ *and quit otherwise (fig.(3b)).*
3. ***SFS policy:*** *If* $V_S(p^f) < V_{FS}(p^f)$ *and* $V_{FS}(p_S) < 0$, *then the optimal policy is* $\alpha^*(\cdot) = \alpha^{SFS}_{p_1,p^f}$ *where* $p_1 := \{p : V_S(p) = V_{FS}(p)\}$[19] *(fig.(3c)).*

Notice that all the all of the cases above are conditions on the primitives of the model, and therefore, we have a complete characterization of the optimal policy.[20]

Sketch of the proof: Typically, control problems like ours are solved by obtaining a solution to the HJB equation below and invoking verification theorems thereafter. One benefit this approach entails is that we do not need to prove the smoothness of $V(\cdot)$ explicitly.

HJB equation:

$$V(p) - \max_{a,\gamma} a\Big\{ -c + \lambda_b(1-p)(v^0 - V(p)) + \lambda_b p(1-p)V'(p) + \gamma H(p, V(p), V'(p))\Big\} = 0$$

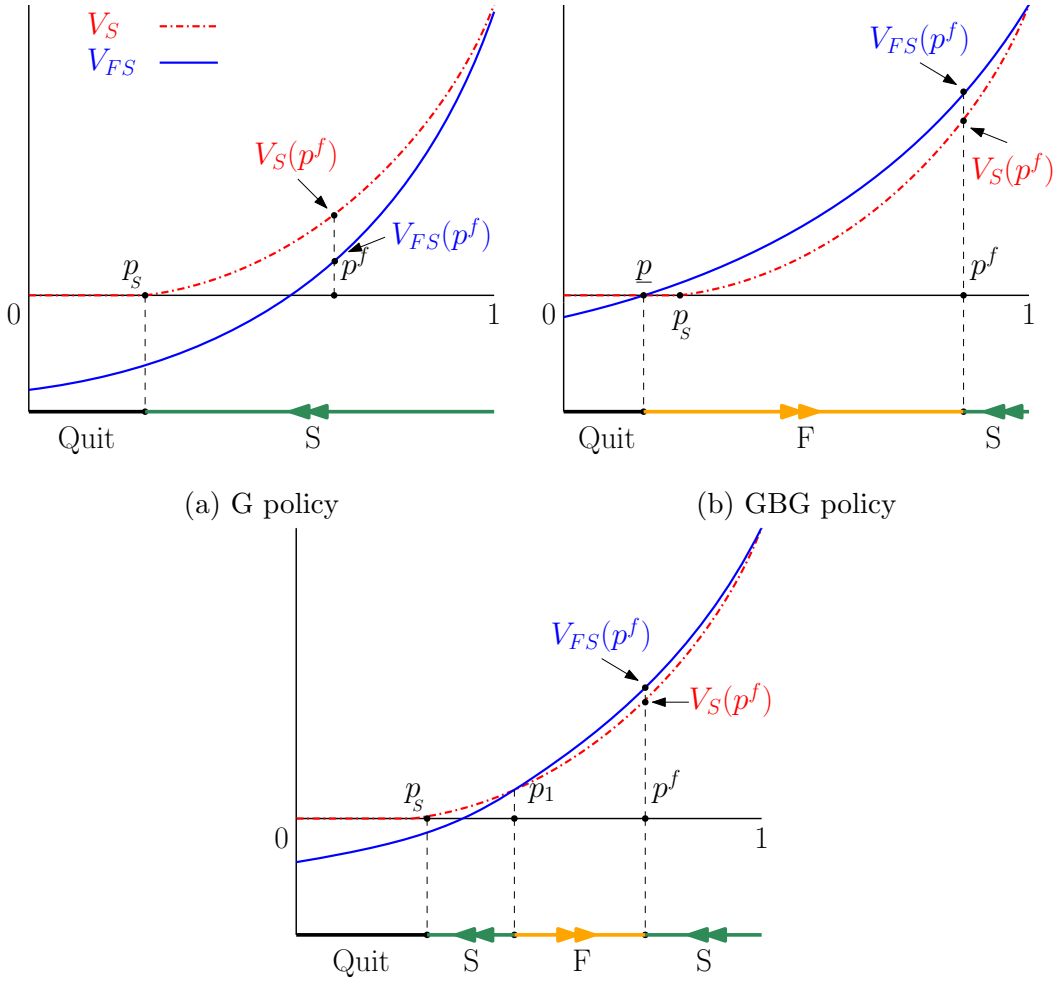
where

$$H(p, V(p), V'(p)) := \lambda_g p[v^1 - V(p)] - \lambda_b(1-p)[v^0 - V(p)] - [\lambda_b + \lambda_g]p(1-p)V'(p)$$

However, in our setting obtaining a candidate solution that satisfies the HJB equation globally, at least in the viscosity sense, is not straightforward. Therefore, we reason through the properties of the optimal policy. Of course, in order to do so, we invoke standard results from optimal control theory to establish the existence of an optimal policy, (a^*, γ^*) (Proposition 10).

[19] That such a p is unique follows from the fact that $V(p) = V_{FS}(p) > V_S(p) \implies V(q) = V_{FS}(q), \forall q \geq p$.

[20] There exist parameters for which each of the above may occur. For example, set $\Gamma = 11, \lambda_g = 1, c = 3$. Then, with $\lambda_b = 1$, we have the G only policy as optimal, with $\lambda_b = 10$, we have SFS policy as optimal. And for $\lambda_b = 25$ we have FS as the optimal policy.



(a) G policy

(b) GBG policy

(c) BG policy

Figure 3: Three possible optimal policies

The proof has three main parts. First, replicating the arguments in Strulovici and Szydlowski (2012) we obtain that $V(\cdot)$ is convex, and hence differentiable a.e. Typical dynamic programming arguments yield that $V(\cdot)$ satisfies the HJB equation at points of differentiability (Proposition 9). Linearity in a implies the obvious—it is optimal to set $a^* = 1$ when $V(\cdot) > 0$. Moreover, linearity in γ tells us that, wlog, $\gamma^* \in \{0, 1\}$ a.e. whenever $H(\cdot) \neq 0$. Therefore, $\gamma_t^* \in \{0, 1\}$ whenever $V(P_t)$ is differentiable (Lemma 11).[21] Lastly, if $H(p) = 0$ for some p, it is without loss to set $\gamma = \gamma^f$ which yields that $V(p) = V^f(p)$.

Second, we show that $\alpha_{p^f}^{FS}$ delivers a strictly a higher value than freezing at all points except p^f (Lemma 8). Therefore, if the optimal control involves freezing anywhere it can only be at p^f. Moreover, we show that, if the optimal policy entails freezing beliefs at some p for any interval of time $[t_1, t_2]$, it must mean that $V(p) = V^f(p) \implies p = p^f$.

[21] This argument even extends to the points where $V(\cdot)$ is not differentiable. Due to convexity, $V(\cdot)$ has one-sided derivatives everywhere and, therefore, appropriate one sided versions of HJB are used, see Lemma 11).

Lastly, results from stochastic optimal control theory yield that it is, in fact, without loss of generality, to restrict attention to Markov controls (Kurtz and Stockbridge (1998), Lemma 13). Therefore, we obtain Lemma 14—when following γ^* beliefs move (weakly) only in one direction absent conclusive news. That is, if $P_0 = p$ and $P_t = q > p$ for some $t > 0$, then they never come back to p. Therefore, only one of the following three possibilities remain: Beliefs keep moving down (α^S), beliefs move up and freeze at p^f, or beliefs move down and freeze at p^f.[22] The latter two combine to give the policy $\alpha^{FS}_{p^f}$ where the behavior differs only depending on whether P_0 is below p^f or above. As a result, the optimal value function is merely a comparison between the values by these two policies delivering Proposition 4.

3.4. Optimal Reward Structure

In Proposition 4 we obtained the optimal experimentation policy, α^* for the principal with $R^S = R^F = 1$ ignoring the agent incentives. Here, we will show that there is a rather simple way to implement the same. To this end, we present our main result below. In words, Proposition 5 says that there exists an equilibrium that delivers to the principal a value of $V(\cdot)$ obtained in Proposition 4 when $R^S = R^F = 1$. Notice that the optimal value the DM can achieve in (DMP) is decreasing in (R^S, R^F). Therefore, for any $(R^S, R^F) \geq (1, 1)$, $\Pi^*(\cdot|R^S, R^F) \leq V(\cdot)$.

Proposition 5. *Let $P_0 = p$. Then, $\exists\ R^F_* \in \{0, 1\}$ such that, $\Pi^*(p|1, R^F_*) = V(p)$.*[23]

Proof. As seen in Proposition 4, the optimal policy α^* to (DMP) is one of the following: (1) a S only policy α^S, (2) a FS policy $\alpha^{FS}_{p^f}$ or (3) a SFS policy $\alpha^{SFS}_{p_1,p^f}$, or

If $\alpha^* = \alpha^S$, then set $R^F_* = 0$, As seen in Section 3.2, with no rewards for failure, the agent only searches for success . Therefore, $\Pi^*(\cdot|1, 0) = V_S(\cdot) = V(\cdot)$.

Suppose $\alpha^* = \alpha^{SFS}_{p_1,p^f}$. Then we have two cases depending on whether $P_0 \geq p_1$, i.e. $V(p) = V_{FS}(p)$ or $P_0 < p_1$, i.e. $V(p) = V_S(p)$, where P_0 is the initial prior (Refer to Fig. (3c)).

If $P_0 \geq p_1$, the DM searches for failure until the beliefs reach p^f. At p^f, the DM freezes beliefs until a conclusive signal arrives. By setting $R^F_* = 1$ *and* $T = \infty$, the agent is indifferent across any γ_t. Therefore, in particular, the agent can choose $\gamma^{FS}_{p^f}$, yielding $\Pi(0, p, \gamma^{FS}_{p^f}, \infty|(1, 1)) = V_{FS}(p) = V(p)$. On the other hand, if $P_0 < p_1$, then choosing $R^F_* = 0$ delivers $\Pi^*(p|1, 0) = V_S(p) = V(p)$ as in the previous case.

Lastly, if $\alpha^* = \alpha^{FS}_{p^f}$, then, as in the previous case, setting $R^F_* = 1, T = \infty$, we can obtain $\Pi(0, p, \gamma^{FS}_{p^f}, \infty|(1, 1)) = V_{FS}(p) = V(p)$. Since, $\Pi^*(p|1, 1) \leq V(p)$, we obtain the desired equality $\Pi^*(p|1, 1) = V(p)$

[22] Technically, there is also the possibility of beliefs always moving up. However, that would mean the DM only looks for B which is obviously suboptimal.

[23] As will be clear in the proof, R^F_* can depend on the initial prior p.

□

Remark 2. *Notice that in the SFS case, R^F_* depends on P_0.*

Finally, as consequence, we obtain our main result—either the principal should reward success or failure equally, or do not reward failure at all. Even though the result is a corollary of Proposition 5 and Proposition 3, we state it as a Proposition.

Proposition 6. *$\Pi^*(\cdot|R^S, R^F) \leq \max\{\Pi^*(\cdot|1,0), \Pi^*(\cdot|1,1)\}$ for any $(R^S, R^F) \in [1,\infty) \times \mathbb{R}_+$.*

4. Some observations

4.1. Scale of innovation and rewarding for failure

As mentioned in the introduction, a key feature of the ideas that make the cut in Google X, according to Teller, is that they have a truly transformative potential, e.g. self-driving cars. In Teller's philosophy, the scale of profits is a lot higher in such ideas compared to incremental innovations that largely rely on a strong sales team to earn revenue. In our model, this is captured by Γ—the value of an S signal that establishes that $\theta = 1$. And indeed, for a range of non knife-edge parameters, we can confirm Teller's insight—when Γ is sufficiently high it may be optimal to reward failures, and not do so if Γ is low. We summarize this in Proposition 7.

Proposition 7. *$\exists\ \lambda_b, \lambda_g, c$ and $\Gamma_1 < \Gamma_2$ such that, $\alpha^* = \alpha^S$ when $\Gamma = \Gamma_1$ and $\alpha^* = \alpha^{FS}_{p^f}$ when $\Gamma = \Gamma_2$.*

To see the intuition, first notice that $V^f(1) < 0 \implies V^f(p^f) < 0$. Therefore, the optimal policy in this case must be α^S. Moreover, as (6) shows when $v^1 = \Gamma_1 - 1$ is sufficiently small, it is possible to satisfy Assumption 2 and yet have $V^f(1) < 0$.[24] More importantly, this is true *regardless* of how high λ_b is. Noticing this, we take both λ_b and Γ to infinity and notice that, when both λ_b and Γ are sufficiently high, we have that $J(p_S, \alpha^{FS}_{p^f}) > 0$.[25]

4.2. When to reward failure?

At this point, we know that if the optimal policy (α^*) in Proposition 4 is either $\alpha^{FS}_{p^f}$ or $\alpha^{SFS}_{p_1,p^f}$, we may end up rewarding the agent for failure. In fact, if $\alpha^* = \alpha^{FS}_{p^f}$, we

[24] The precision condition is $\lambda_g v^1 \in (c, (1+\lambda_g/\lambda_b)c)$.

[25] We would like to emphasize that both p_S and p^f change with Γ, and p^f changes with λ_b as well. Therefore, we first choose a Γ large enough so as to be able to generate $V_{FS}(p_S, \alpha^{FS}_{p^f}) > 0$. Then, we take $\lambda_b \to \infty$ to achieve this.

will necessarily do so.[26] However, we would like to draw attention to the case when $\alpha^* = \alpha^{SFS}_{p_1,p^f}$. Here we will reward the agent for failure iff $P_0 \geq p_1$, and not otherwise. That is, perhaps counterintuitively, the principal will reward the agent for failure if the initial prior is sufficiently high but not otherwise. The reasoning, in the context of our model, is not very hard to see—when the initial prior is low, the expected size of the pie ($P_0\Gamma$) is itself small. Therefore, promising a reward for failure may not be worthwhile. Practically, this suggests that we should be witnessing failures being rewarded more often for projects that were ex-ante more likely to succeed.

4.3. Relation between λ_b and rewarding for failure

An obvious comparative static one would guess is that if λ_b is very low, we should not see failure being rewarded. Similarly, if λ_b is very high we should see it being rewarded. We confirm the intuition below in Proposition 8.

Proposition 8. *Suppose* $v^0 + \frac{c}{1+\lambda_g} > 0$. *Then,* $\exists \lambda_b^1 < \lambda_b^2$ *such that* $\alpha^*(\lambda_b^1) = \alpha^S$ *and* $\alpha^*(\lambda_b^2) = \alpha^{FS}_{p^f}$.

First, if $\lambda_b \approx 0$ it is obvious that $\alpha^*(\lambda_b) = \alpha^S$ in Proposition 4 regardless of the value of Γ. Therefore, choosing $\Gamma = \Gamma_2$ as in Proposition 7, we can can vary λ_b from $\lambda_b^1 = 0$ to a sufficiently high value so that $\alpha^*(\lambda_b) = \alpha^{FS}_{p^f}$ to deliver the proposition. Therefore, we skip the proof.

As mentioned in the introduction the "bug bounty programs" are becoming increasingly popular. One reason could be that finding bugs or flaws in a code is easier, captured through higher λ_b, compared to finding similar flaws in, say, a manufacturing unit.

4.4. Searching for failure vs reporting failure

As mentioned before, much of the literature on rewarding failure has focused on incentivizing the agent to report failures. Suppose we enrich our model so that success is public but failure is not. So, the agent can hide a failure and continue receiving wages for longer. However, notice that the agent searches for failure only when $R^F = 1$. Therefore, even if he could conceal failure and receive wages longer, it does not help him attain a strictly higher payoff. Therefore, he has no incentive to withhold failures.

It is important to point out, however, that this insight relies on there being moral hazard only on the intensive margin—allocation choice between looking for success and failure—-and not on the extensive margin. If the effort was costly and the agent could search, then the agent may wish to conceal a failure and collect wages longer while shirking and saving effort costs.

[26] This is of course assuming $P_0 \geq \underline{p}$. To keep the discussion simple, we assume that to be the case.

5. Conclusion

In this paper, we studied a simple model of a principal-agent relationship with experimentation and limits to contractibility. The main focus of the paper was to determine whether and when the principal should reward the agent for reporting failure, and how the optimal reward scheme should be structured. Our main takeaway is that either the principal should offer no reward to the agent for failure, or she should offer the same reward for success or failure. Given that rewarding failure is costly, the sole reason for offering such a reward is to incentivize the agent to search for failure, thereby potentially saving future experimentation costs. Prior to this paper, most research that prescribed rewarding failure has focused on providing incentives to the agent to disclose failures. In contrast, we show that even when such concerns are absent, i.e. the signal is public, a fundamental source of conflict arises due to the agent's aversion to searching for failure because its arrival triggers his termination.

A key feature of our model—viewing experimentation as acquiring information from multiple sources—brings out novel dynamics. Our model predicts that rewarding for failure may be more common in experimentation environments where the informativeness of the failure arm is high. Our results may also provide an alternative explanation to why failures are not transmitted efficiently to management in organizations—it is not that the employees hide negative information, but rather that they choose not to acquire it when there is no reward for finding negative information. As mentioned before, this insight could, potentially, break down if we enriched the model to add moral hazard on the extensive margin which seems like a natural point of inquiry ahead.

A. Appendix

A.1. When $F < 1$ (Section 3.2).

Proof of Lemma 1. Suppose $T = \infty$, i.e. P never fires A in the absence of a conclusive signal. Notice that $U(t, p, T) \leq 1$. Choosing $\gamma_t = 1$ for all t delivers $U(t, p, T) = 1$. Moreover, any strategy involving $\gamma_t \neq 1$ for a.e. t results in a positive probability of an F signal, and therefore termination. Hence, $\gamma_t = 1$ for a.e. t is the unique best response to $T = \infty$. If $\gamma_t = 1$ for all t, by Proposition 1, it is optimal for the principal to fire the agent at $T_G(P_0)$.[27] Therefore, $T = \infty$ cannot be a best response for the principal.

Suppose $T < \infty$. Let $\gamma^1_t = 1 \ \forall t$, be a constant control where the agent searches for success at all times. Define,

$$\begin{aligned}\alpha :=& 1 - \exp\left(-\int_0^T \lambda_g \gamma_t \mathrm{d}t\right)\\ \beta :=& 1 - \exp\left(-\int_0^T \lambda_b (1-\gamma_t) \mathrm{d}t\right).\end{aligned}$$

α is the probability of receiving S before T conditional on $\theta = 1$, given γ. Similarly β is the probability of receiving F before T conditional on $\theta = 0$. Notice that $\beta > 0 \implies (1-p)\beta > 0$. Therefore, with positive probability, an F signal will arrive before T that would result in the agent's termination. Let $x(\beta)$ be the agent's ex-ante expected value conditional on an F signal arriving before T. Obviously, $x < (1 - e^{-T})$ since $R^F = 0$. Moreover, $x(\beta)$ is decreasing in β. Therefore,

$$\begin{aligned}S(0, p, \gamma^1, T) =& p[\alpha 1 + (1-\alpha)(1-e^{-T})] + (1-p)[\beta x + (1-\beta)(1-e^{-T})]\\ =& p\alpha e^{-T} + (1-e^{-T})(1-p) + (1-p)\beta(x - (1-e^{-T}))\end{aligned}$$

First, RHS is increasing in α. Second, since $(x-(1-e^{-T})) < 0$, increasing $\beta \implies \beta(x-(1-e^{-T}))$ decreases. Therefore, $U(\cdot)$ is decreasing in β. Therefore, the unique best response for any $T < \infty$ is to set $\gamma_t = 1$ for a.e. $t \leq T$. Therefore, $\Pi^*(\cdot|1, 0) \leq V_S(\cdot)$. Equality obtains if the principal sets $T = T_G(p)$, the time it takes for the beliefs to drift from $P_0 = p$ to p_S in the absence of an S signal (refer to Proposition 1).

□

Towards the proof of Proposition 2, we first observe that an optimal control exists in the (AP) with $T < \infty, R^S = 1, R^F < 1$. This follows from Corollary 1.4 Chapter VI from Bardi and Capuzzo-Dolcetta (2008). See Proposition 10 for details.

Lemma 4. *There exists a best response to (T, R^S, R^F). (Refer to (AP).)*

[27] To apply the result, set $v^1 = \Gamma - R = \Gamma - 1$.

Proof of Lemma 3. Suppose not. Let $T_1 := \int_0^T \gamma_t \mathrm{d}t$ Consider an alternative control $\gamma' := \mathbb{1}_{t \leq T_1}$. Notice that, for any t, $\int_0^t \gamma'_s \mathrm{d}s \geq \int_0^t \gamma_s \mathrm{d}s$. Moreover, since $\gamma'_t = 0$ whenever $t > T_1$,

$$\int_0^T (1-\gamma'_s)\mathrm{d}s = T - T_1 = \int_0^T (1-\gamma_s)\mathrm{d}s \implies \int_0^t (1-\gamma'_s)\mathrm{d}s \leq \int_0^t (1-\gamma_s)\mathrm{d}s \quad \forall t \in [0,T].$$

Also, the inequality is strict at least for some t by assumption.

For any control $\eta \in \mathcal{G}$, define, $\tau_B^\eta := \inf\{t \geq 0 : P_t^\eta = 0\} \wedge T$. Notice that τ_B^η is exponential distributed with $F_\eta(t) := \mathbb{P}(\tau_B^\eta \leq t | \theta = 0) = \frac{1-\exp(-\int_0^t \lambda_b(1-\eta_s)\mathrm{d}s}{\exp(-\lambda_b \int_0^T (1-\eta_s)\mathrm{d}s)}$ as its CDF for any $t < T$. Since, $\int_0^t (1-\gamma'_s)\mathrm{d}s \leq \int_0^t (1-\gamma_s)\mathrm{d}s$, $F_\gamma(t) \geq F_{\gamma'}(t)$ for all $t < T$, i.e. $\tau_B^{\gamma'}$ first order stochastically dominates τ_B^γ.

Lastly, notice that the probability of receiving an S signal using either γ or γ' before T is $x := p(1 - e^{-\lambda_g T_1})$. Therefore,

$$\begin{aligned} U(0,p,\gamma,T) =& p[x + (1-x)(1-e^{-T})] + (1-p)\mathbb{E}_\gamma\left[(1 - e^{-\tau_B^\gamma}(1-F))|\theta = 0\right] \\ <& p[x + (1-x)(1-e^{-T})] + (1-p)\mathbb{E}_{\gamma'}\left[(1 - e^{-\tau_B^{\gamma'}}(1-F))|\theta = 0\right] \\ =& U(0,p,\gamma',T) \end{aligned}$$

where the strictly inequality is due to the fact that $F_\gamma(\cdot) \neq F_{\gamma'}(\cdot)$ and $(1-e^{-t}(1-R^F))$ is a strictly increasing function. Therefore, γ' delivers a strictly higher value than γ, contradicting the optimality of γ. □

Proof of Proposition 2. Consider any equilibrium (γ, T) with the reward structure (R^S, R^F) with $R^S = 1$, $R^F < 1$, such that $T < \infty$. By Lemma 3, $\exists\ T_1 \leq T$ such that $\gamma_t = \mathbb{1}_{t \leq T_1}$ for a.e. t. Suppose $T_1 < T$. Then, principal can fire the agent at T_1 (conditional on signal). Since the agent only searches for a failure after T_1, the principal incurs the flow cost c of experimentation as well as a potential reward R^F for an F signal. By firing the agent at T_1, the principal can guarantee herself a continuation payoff of 0 at T_1 which is strictly larger in expectation compared to the continuation payoff at T_1 when following γ. Therefore, in equilibrium, $T_1 = T$. □

B. Proof of Proposition 4

We analyze (DMP) in this section to obtain the optimal control/policy.

Claim 1. *$V(\cdot)$ is convex.*

Proof. Let $h(a,\theta) := \mathbb{E}_{a,\theta}\left[\int_0^\tau e^{-t} a_t(-c)\mathrm{d}t + e^{-\tau} g(y_\tau)\right]$. Then, $J(p,a) = ph(a,1) + (1-p)h(a,0)$. Therefore, $J(p,a)$ is affine in p. Since the supremum of affine functions is convex, $V(p)$ is convex in p. □

Fact 1. *A convex function $f : [0,1] \to \mathbb{R}$ is continuous on $(0,1)$, and is differentiable almost everywhere. Moreover, one-sided derivatives $V'_+(p)$ and $V'_-(p)$ exist for all $p \in (0,1)$.*

Let

$$\begin{aligned}\mathcal{D} :=& \{p \in (0,1) : V(\cdot) \text{ is differentiable at } p\} \\ \mathcal{D}^C :=& (0,1) \backslash \mathcal{D}\end{aligned}$$

Claim 2. *$V(\cdot)$ is continuous at 0.*

Proof. Notice that $-c \leq J(\cdot, \alpha) \leq v^1$ for all α. Therefore, $\frac{\partial J(p,\alpha)}{\partial p} \leq M$ for some M. Moreover, $V(p) \geq 0$ is obvious as $\alpha = (0, \cdot)$ yields $J(, \cdot, \alpha) = 0$. Suppose $V(0+) = 2\epsilon > 0$. Therefore, $\exists \alpha_n, p_n$ such that $p_n \downarrow 0$ and $J(p_n, \alpha_n) \geq \epsilon$. Therefore, $\frac{\partial J(p_n,\alpha_n)}{\partial p} \geq \frac{\epsilon}{p_n} \to \infty$ as $n \to \infty$. A contradiction. Therefore, $V(0+) = V(0) = 0$.

□

Claim 3. *$\exists \underline{p} \in [0,1]$ such that $V(p) = 0$ iff $p \leq \underline{p}$.*

Proof. Define $\underline{p} := \sup\{p : V(q) = 0 \forall q \in [0,p]\}$. By definition, $V'_+(\underline{p}) \geq 0$. Consider any $q > \underline{p}$ such that $V(q) > 0$. By convexity, $V'_+(q) \geq V(q)/(q - \underline{p}) > 0$. Therefore, $V(p) > 0$ for all $p > \underline{p}$. Therefore, for any $p > q$, $V(p) \geq V(q) + V'_+(q)(p - q) > 0$. □

HJB Equation

Define,

$$H(p, V(p), V'(p)) := \lambda_g p[v^1 - V(p)] - \lambda_b(1-p)[v^0 - V(p)] - [\lambda_b + \lambda_g]p(1-p)V'(p) \tag{4}$$

The HJB equation for our problem is the following:

$$V(p) - \max_{a,\gamma} a\Big\{ -c + \lambda_b(1-p)(v^0 - V(p)) + \lambda_b p(1-p)V'(p) + \gamma H(p, V(p), V'(p))\Big\} = 0 \tag{5}$$

Let,

$$W(p, V(p), V'(p)) := \sup_{\gamma} \Big\{ -c + \lambda_b(1-p)(v^0 - V(p)) + \lambda_b p(1-p)V'(p) + \gamma H(p, V(p), V'(p))\Big\}$$

Proposition 9. *If $p \in \mathcal{D}$ then $V(p)$ satisfies (5).*

Proof. We will first show that if $V(\cdot)$ is differentiable at p, then

$$V(p) - \max_{a,\gamma} a\Big\{ -c + \lambda_b(1-p)(v^0 - V(p)) + \lambda_b p(1-p)V'(p) + \gamma H(p, V(p), V'(p))\Big\} \geq 0$$

Consider a constant control $\alpha = (a, \gamma)$. By the Dynamic Programming Principle, for a small $\mathrm{d}t > 0$,

$$\begin{aligned} V(p) \geq & a\Big\{(-c)\mathrm{d}t + \lambda_g \gamma p v^1 \mathrm{d}t + \lambda_b(1-\gamma)(1-p)v^0 \mathrm{d}t \\ & + (1-\mathrm{d}t)(1 - \lambda_g \gamma p \mathrm{d}t - \lambda_b(1-\gamma)(1-p)\mathrm{d}t)\Big[V(p) + (\lambda_b(1-\gamma) - \lambda_g \gamma)p(1-p)V'(p)\mathrm{d}t\Big]\Big\} \end{aligned}$$

Rearranging the equation and dividing by $\mathrm{d}t$, we obtain,

$$V(p) \geq a\Big\{ -c + \lambda_b(1-p)(v^0 - V(p)) + \lambda_b p(1-p)V'(p) \quad + \gamma H(p, V(p), V'(p))\Big\}$$

Since (a, γ) are arbitrary, we obtain

$$V(p) - \max_{a,\gamma} a\Big\{ -c + \lambda_b(1-p)(v^0 - V(p)) + \lambda_b p(1-p)V'(p) + \gamma H(p, V(p), V'(p))\Big\} \geq 0$$

For the reverse inequality, for any $\epsilon > 0$, $\exists \alpha$ such that $J(p, \alpha) \geq V(p) - \epsilon$. Let $\Delta := \sqrt{\epsilon}$. For a small ϵ, define $\tau' := \Delta \wedge \tau$. By the Dynamic programming principle,

$$\begin{aligned} V(p) - \epsilon \leq J(p, \alpha) &\leq \mathbb{E}\left[\int_0^{\tau'} (-c)a_t e^{-t}\mathrm{d}t + e^{-\tau'}V(P_{\tau'})\right] \\ &\leq \int_0^{\Delta} (-c)a_t e^{-t}\mathrm{d}t + p(1 - e^{-\int_0^{\Delta} \lambda_g a_s \gamma_s \mathrm{d}s})v^1 + (1-p)(1 - e^{-\int_0^{\Delta} \lambda_b a_s (1-\gamma_s)\mathrm{d}s})v^0 \\ &\quad + (1-\Delta)(p e^{-\int_0^{\Delta} \lambda_g a_s \gamma_s \mathrm{d}s} + (1-p)e^{-\int_0^{\Delta} \lambda_b a_s (1-\gamma_s)\mathrm{d}s})V(P_\Delta | \text{No news until } \Delta) \end{aligned}$$

Therefore, using first order approxmiations, we have,

$$\begin{aligned} V(p) - \epsilon \leq & (-c)\tilde{a}\Delta + p\lambda_g \tilde{\gamma}\tilde{a}\Delta v^1 + (1-p)\lambda_b(1-\tilde{a})(1-\tilde{\gamma})v^0 \\ & + (1-\Delta)\Big(1 - p\lambda_g \tilde{a}\tilde{\gamma}\Delta - (1-p)\lambda_b \tilde{a}(1-\tilde{g})\Delta\Big)\Big[V(p) - \tilde{a}(\lambda_b(1-\tilde{g}) - \lambda_g \tilde{\gamma})V'(p)\Big] \end{aligned}$$

where $\tilde{a} := \int_0^{\Delta} a_s \mathrm{d}s$, $\tilde{\gamma} := \frac{1}{\tilde{a}} \int_0^{\Delta} a_s \gamma_s \mathrm{d}s$.

Rearranging the above,

$$\begin{aligned} & V(p) - \tilde{a}\Big\{ -c + \lambda_b(1-p)(v^0 - V(p)) + \lambda_b p(1-p)V'(p) + \tilde{\gamma}H(p, V(p), V'(p))\Big\} \leq \frac{\epsilon}{\Delta} = \sqrt{\epsilon} \\ \implies & V(p) - \max_{a,\gamma} a\Big\{ -c + \lambda_b(1-p)(V(p) + v^0) + \lambda_b p(1-p)V'(p) + \gamma H(p, V(p), V'(p))\Big\} \leq \sqrt{\epsilon} \end{aligned}$$

Since ϵ is arbitrary, we obtain the desired inequality below.

$$V(p) - \max_{a,\gamma} a\Big\{ -c + \lambda_b(1-p)(v^0 - V(p)) + \lambda_b p(1-p)V'(p) + \gamma H(p, V(p), V'(p))\Big\} \leq 0$$

□

Proposition 10. *An optimal control exists, i.e.* $\exists \alpha^* = (a_t^*, \gamma_t^*)_{t \geq 0} \in \mathcal{A}$ *such that* $J(p, \alpha^*) = V(p)$.

Proof. We can view the control problem at hand as a deterministic control problem. To this end, define

$$A_t := \int_0^t a_s e^{-s} \mathrm{d}s, G_t := e^{-\lambda_g \int_0^t a_s \gamma_s \mathrm{d}s}, B_t := e^{-\lambda_b \int_0^t a_s (1-\gamma_s) \mathrm{d}s}.$$

Then, we have,

$$J(p,\alpha) = \int_0^\infty p\lambda_g a_t \gamma_t G_t \left[v^1 e^{-t} - cA_t\right] \mathrm{d}t + \int_0^\infty (1-p)\lambda_b a_t (1-\gamma_t) B_t \left[v^0 e^{-t} - cA_t\right] \mathrm{d}t$$
$$V(p) = \sup_{\alpha \in \mathcal{U}} J(p,\alpha)$$

Corollary 1.4, Chapter VI from Bardi and Capuzzo-Dolcetta (2008) establishes that an optimal control exists for the above problem. To formally apply the Corollary, we need to define a new state variable $x := (t, p, A, G, B)$. Then, $\dot{x} = f(x,u)$, where $u = (a,\gamma) \in \mathcal{Y}$. We can define

$$f(x,u) = \begin{pmatrix} 1 \\ a[\lambda_b(1-\gamma) - \lambda_g \gamma]p(1-p) \\ ae^{-t} \\ -\lambda_g a \gamma G \\ -\lambda_b a(1-\gamma)B \end{pmatrix}$$

Lastly, $l(x,\mathcal{Y}) := \{p\lambda_g a_t \gamma_t G_t[v^1 e^{-t} - cA_t] + (1-p)\lambda_b a_t (1-\gamma_t) B_t [v^0 e^{-t} - cA_t] : u \in \mathcal{Y}\}$. It is obvious that $f(x,\mathcal{Y}) \times l(x,\mathcal{Y})$ is convex, and hence the corollary applies. □

Define $V^f(p) := J(p, \alpha^f)$ where $\alpha_t^f = (1, \gamma^f)$ for all t.

Lemma 5. *At any $p > \underline{p}$ such that $p \in \mathcal{D}^C$, at least one of the following holds:*

1. $V(p) - \sup_{a,\gamma<\gamma^f} aW(p, V(p), V'_+(p)) = 0$.
2. $V(p) - \sup_{a,\gamma>\gamma^f} aW(p, V(p), V'_-(p)) = 0$.
3. $V(p) = V^f(p)$.

Proof. First of all, by standard dynamic programming argument using a constant control (as in Proposition 9), we have,

$$V(p) - \sup_{a,\gamma<\gamma^f} aW(p, V(p), V'_+(p)) \geq 0$$
$$V(p) - \sup_{a,\gamma<\gamma^f} aW(p, V(p), V'_-(p)) \geq 0$$

For the optimal control α^*, assume wlog that $a_t^* = 0 \implies \gamma_t^* = 0$. For any time t, define, $\tilde{a}_t := \frac{1}{t}\int_0^t a_s^* \mathrm{d}s$, $\tilde{\gamma}_t := \frac{\int_0^t a_s^* \gamma_s^* \mathrm{d}s}{t\tilde{a}_t}$.

Suppose $\exists$ a sequence $(h_n) \downarrow 0$ such that, in the absence of G or B, $P_n := P_{h_n}^{\alpha^*} > p$. Let $\tilde{\gamma}_n := \tilde{\gamma}_{h_n}$ and $\tilde{a}_n := \tilde{a}_{h_n}$. From (1), $P_n > p \implies \tilde{\gamma}_t < \gamma^f$.[28] Therefore, using first

[28] Beliefs will move up iff $\gamma_t < \gamma^f$ and $a_t > 0$.

order approxmiations, we have,

$$\begin{aligned}V(p) =&\tilde{a}_n\Big[-ch_n+\lambda_b(1-p)(1-\tilde{\gamma}_n)(v^0-V(p)h_n+\lambda_b p(1-p)V'_+(p)h_n+\\&\gamma\Big(\lambda_g p(v^1-V(p)h_n-\lambda_b(1-p)(v^0-V(p)-(\lambda_b+\lambda_g)p(1-p)V'_+(p)\Big)\Big]\\&+(1-h_n)V(p)\end{aligned}$$

Taking $n\to\infty$, possibly passing on to a subsequence where $(\tilde{a}_n,\tilde{\gamma}_n)\to(\tilde{a},\tilde{\gamma})$.

$$V(p)-\tilde{a}W(p,V(p),V'_+(p))=0 \implies V(p)-\sup_{\tilde{a},\tilde{\gamma}<\gamma^f}\tilde{a}W(p,V(p),V'_+(p))=0$$

For (2), suppose $\exists$ a sequence $(h_n)\downarrow 0$ such that, in the absence of G or B, $P_n:=P^{\alpha^*}_{h_n}<p$. Rest of the argument is identical as above to obtain

$$V(p)-\sup_{\tilde{a},\tilde{\gamma}>\gamma^f}W(p,V(p),V'_-(p))=0$$

However, if $\nexists$ a sequence as in (1) and (2) $\implies P_s=p$ for all $s\in[0,t]$ for some s. Therefore, either $a^*_s=0$ for all $s\in[0,T]$ or $\gamma^*_s=\gamma^f$ for a.e. s such that $a^*_s>0$. If $a^*_s=0$ for a.e. s, then $V(p)=0$, a contradiction as $p>\underline{p}$. Therefore, $\gamma^*_s=\gamma^f$ for a.e. s. It is straightforward to check, then, $V(p)=\tilde{a}V^f(p)$ where $\tilde{a}=\lim_{n\to\infty}\frac{\int_0^{\frac{1}{n}}a^*_s\mathrm{d}s}{\frac{1}{n}}$. Since $V(p)\geq V^f(p)$, we have, $V(p)=V^f(p)$.

□

Below, from Definition 3 up to the end of Lemma 8, we provide a control that delivers a strictly higher value than $V^f(\cdot)$ at all but one p, to be called p^f.

Definition 3. *The following control control $\alpha^{FS}_{p_1}=(a,\gamma)$ is called a "FS policy with a switch at p_1"*

1. *$a_t=1$ for all t.*
2. $$\gamma_t=\begin{cases}1 & \text{if } P_t>p_1\\ \gamma^f & \text{if } P_t=p_1\\ 0 & \text{if } P_t<p_1\end{cases}$$

Claim 4.

$$V^f(p)=\left(-c+\frac{\lambda_b\lambda_g}{\lambda_b\lambda_g+\lambda_b+\lambda_g}(c+v^0)\right)+\frac{\lambda_b\lambda_g[v^1-v^0]}{\lambda_b+\lambda_g+\lambda_b\lambda_g}p. \tag{6}$$

Proof. Using α, the beliefs remain at p until conclusive news arrives. Therefore,

$$\begin{aligned}V(p)=&-c\mathrm{d}t+\lambda_g p\gamma^f v^1\mathrm{d}t+\lambda_b(1-p)(1-\gamma^f)\mathrm{d}t+\\&+(1-\mathrm{d}t)(1-p\lambda_g\gamma^f\mathrm{d}t-(1-p)\lambda_b(1-\gamma^f)\mathrm{d}t)V(p)\end{aligned}$$

Rearranging the above using the fact that $\lambda_b(1-\gamma^f) = \lambda_g\gamma^f$, we obtain the above expression. □

Lemma 6.

$$J(p, \alpha^{FS}_{p_1}) = \begin{cases} \frac{\lambda_b(1-p)(v^0+c)}{\lambda_b+1} + C_0 p\left[\frac{p}{1-p}\right]^{\frac{1}{\lambda_b}} - c & \text{if } p < p_1, \\ \frac{\lambda_b\lambda_g}{\lambda_b\lambda_g+\lambda_b+\lambda_g}(c + p[v^1 - v^0] + v^0) - c & \text{if } p = p_1, \\ \frac{\lambda_g p(v^1+c)}{1+\lambda_g} + C_1(1-p)\left[\frac{1-p}{p}\right]^{\frac{1}{\lambda_g}} - c & \text{if } p > p_1 \end{cases}$$

where C_0 and C_1 are determined using continuity of $J(\cdot, \alpha^{FS}_{p_1})$ at p_1.

Proof. The proof is straightforward given that following $\alpha^{FS}_{p_1}$, for any belief $p < p_1$, there is exclusive search for failure until beliefs hit p_1.

In the absence of a signal, we have,

$$\dot{Z}_t = \lambda_b \implies Z_t = Z_0 + \lambda_b t,$$

where $Z_t := \log\left(\frac{P_t}{1-P_t}\right)$. Let $Z^1 = \log\left(\frac{p_1}{1-p_1}\right)$, we can define $t_1(p) := (Z^1 - Z_0)/\lambda_b$. This is the time it takes for the beliefs to reach p_1 starting at $P_0 < p_1$ in the absence of a signal. Therefore, letting $P_0 = p < p_1$,

$$\begin{aligned} f_0(p) := J(p, \alpha^{FS}_{p_1}) = &\int_0^{t_1} (1-p)\lambda_b e^{-\lambda_b t}[e^{-t}v^0 - (1-e^{-t})c]\mathrm{d}t \\ &+ e^{-\lambda_b t_1}\left[(1-e^{-t_1})(-c) + e^{-t}V^f(p_1)\right] \end{aligned}$$

Since $t_1(\cdot)$ is differentiable, so is $f_0(\cdot)$. Therefore, it satisfies the differential equation below.

$$\begin{aligned} f_0(p) &= -c + \lambda_b(1-p)[v^0 - f_0(p)] + \lambda_b p(1-p)f_0'(p) \\ \implies f_0(p) &= -c + \frac{\lambda_b(1-p)(v^0+c)}{\lambda_b+1} + C_0 p\left[\frac{p}{1-p}\right]^{\frac{1}{\lambda_b}} \end{aligned}$$

C_0 is obtained by using the fact that the DM switches to γ^f, i.e. freezing at p_1. The value by freezing is $V^f(p_1)$.

An analogous argument gives yields $f_1(p) := J(p, \alpha^{FS}_{p_1})$ when $p > p_1$. For the sake of completeness, we provide the differential equation that $f_1(p)$ satisfies:

$$f_1(p) = -c + \lambda_g p[v^1 - f_G(p)] - \lambda_g p(1-p)f_1'(p)$$

Lastly, $J(p_1, \alpha^{FS}_{p_1}) = V^f(p_1)$.

□

Claim 5. *$\exists! p_1$ such that $J'_-(p_1, \alpha^{FS}_{p_1}) = V^{f\prime}(p_1)$.*

Proof. For left differentiability, we need, $f_B(p_1) = V^f(p_1)$ and $f'_B(p_1) = V^{f'}(p_1)$. Notice that $V^f(\cdot)$ is affine. Let $K := V^{f'}(p) = \frac{\lambda_b\lambda_g}{\lambda_b\lambda_g+\lambda_b+\lambda_g}(v^1-v^0)$. Let $V^f(p) = \beta + Kp$ From (6), we get that

$$\beta := \left(-c + \frac{\lambda_b\lambda_g}{\lambda_b\lambda_g + \lambda_b + \lambda_g}(c + v^0)\right).$$

Therefore, we need,

$$\begin{aligned}\beta + Kp_1 =& - c + \lambda_b(1-p_1)[v^0 - (b + Kp_1)] + \lambda_b p_1(1-p_1)K \\ \implies p_1 =& \frac{-c - \beta + \lambda_b(v^0 - \beta)}{K + \lambda_b(v^0 - \beta)}\end{aligned}$$

Substituting the values for β and K, we get,

$$p_1 = \frac{\lambda_b(c + v^0)}{\lambda_g(v^1 + c) + \lambda_b(v^0 + c)}$$

□

Claim 6. *$\exists! p_2$ such that $J'_+(p_2, \alpha^{FS}_{p_2}) = V^{f'}(p_2)$.*

Proof. The proof is identical to the previous Claim. Repeating the steps above, we obtain,

$$\begin{aligned}p_2 =& \frac{-c - \beta}{K - \lambda_g(v^1 - \beta - c)} \\ \implies p_2 =& \frac{\lambda_b(c + v^0)}{\lambda_g(v^1 + c) + \lambda_b(v^0 + c)}\end{aligned}$$

□

Since p_1 and p_2 obtained in Claim 5 and 6 coincide, we obtain the following lemma.

Lemma 7. *$\exists! p^f$ such that $J(p, \alpha^{FS}_{p^f})$ is differentiable at p^f with $J'(p^f, \alpha^{FS}_{p^f}) = V^{f'}(p)$. Also, p^f is the following:*

$$p^f := \frac{\lambda_b(v^0 + c)}{\lambda_g(v^1 + c) + \lambda_b(v^0 + c)}, \tag{7}$$

Lemma 8. *If $c > R^F$ and $p \neq p^f \implies V(p) > V^f(p)$.*

Proof.

$$J(p, \alpha^{FS}_{p^f}) =: V_{FS}(p) = \begin{cases} \frac{\lambda_b(1-p)(v^0+c)}{\lambda_b+1} + C_0 p\left[\frac{p}{1-p}\right]^{\frac{1}{\lambda_b}} - c & \text{if } p < p^f, \\ \frac{\lambda_b\lambda_g}{\lambda_b\lambda_g+\lambda_b+\lambda_g}(c + p[v^1 - v^0] + v^0) - c & \text{if } p = p^f, \\ \frac{\lambda_g p(v^1+c)}{1+\lambda_g} + C_1(1-p)\left[\frac{1-p}{p}\right]^{\frac{1}{\lambda_g}} - c & \text{if } p > p^f \end{cases} \tag{8}$$

where p^f is given in (7) and the constants C_0 and C_1 given below are calculated by using continuity at p^f.

$$C_0 = \left[\frac{\lambda_g(v^1+c)}{\lambda_b(v^0+c)}\right]^{\frac{1}{\lambda_b}} \left[\frac{\lambda_b}{1+\lambda_b}\right] \left[\frac{\lambda_g\lambda_b}{\lambda_b+\lambda_g+\lambda_b\lambda_g}\right] (v^1+c),$$
$$C_1 = \left[\frac{\lambda_b(v^0+c)}{\lambda_g(v^1+c)}\right]^{\frac{1}{\lambda_g}} \left[\frac{\lambda_g}{1+\lambda_g}\right] \left[\frac{\lambda_g\lambda_b}{\lambda_b+\lambda_g+\lambda_b\lambda_g}\right] (v^0+c).$$

Also,

$$V''_{FS}(p) = \begin{cases} C_0 p \left[\frac{p}{1-p}\right]^{\frac{1}{\lambda_b}} \frac{(1+\lambda_b)}{\lambda_b^2(p(1-p)^2)} & \text{if } p < p^f, \\ C_1 p \left[\frac{1-p}{p}\right]^{\frac{1}{\lambda_g}} \frac{(1+\lambda_g)}{\lambda_g^2(p^2(1-p))} & \text{if } p > p^f. \end{cases}$$

Note that $V_{FS}(\cdot)$ is strictly convex if $C_0 > 0$ and $C_1 > 0$, that is if $v^1 + c$ and $v^0 + c$ are both strictly positive. Since $c > R^F$, these conditions hold and hence $V_{FS}(\cdot)$ is strictly convex. By construction, $V_{FS}(p^f) = V^f(p^f)$ and $V'_{FS}(p^f) = V^{f\prime}(p^f)$. Therefore $V_{FS}(p) > V^f(p)$ when $p \neq p^f$. Therefore the optimal value function $V(p) > V^f(p)$ when $p \neq p^f$. □

Notice that $c > 1$ by Assumption 3. Therefore, $c > R^F = 1$ is always satisfied.

Lemma 9. $P_t > \underline{p} \implies a_t^* = 1$.

Proof. If $P_t \in \mathcal{D}$, $V(P_t)$ satisfies (5). If $W(P_t, V(P_t), V'(P_t)) \leq 0$, then $V(p) = 0$, a contradiction. Therefore, $W(P_t, V(P_t), V'(P_t)) > 0 \implies a_t^* = 1$.

If $P_t \in \mathcal{D}^C$ then $V(p)$ satisfies one of the 3 cases in Lemma 5. Suppose (1), i.e.

$$V(p) - \sup_{a,\gamma<\gamma^f} aW(p, V(p), V'_+(p)) = 0$$

Obviously, if $W(p, V(p), V'_+(p)) \leq 0$, then $V(p) = 0$, a contradiction as $p > \underline{p}$. Therefore, $W(p, V(p), V'_+(p)) > 0$ and $a^*(p) = 1$. Similar argument holds for (2). If (3), then a constant control $\alpha = (1, \gamma^f)$ yields $J(p, \alpha) = V^f(p) = V(p)$. □

As consequence of Lemma 9, the HJB equation when $p > \underline{p}$ reduces to the following:

$$V(p) - \Big\{ -c + \lambda_b(1-p)(v^0 - V(p)) + \lambda_b p(1-p)V'(p) + \max_{\gamma\in[0,1]} \gamma H(p, V(p), V'(p)) \Big\} = 0 \tag{9}$$

The cases in Lemma 5 will no longer feature dependence on a.

Define,

$$F(x, y, z, \Gamma) := y - \Big\{ -c + \lambda_b(1-x)(v^0 - y) + \lambda_b x(1-x)z + \sup_{\gamma\in\Gamma} \gamma H(x, y, z) \Big\}$$

Lemma 10. *Suppose $p \in \mathcal{D}^C$ and $V(p) > V^f(p)$. The following holds:*

1. $F(p, V(p), V'_+(p), [0, \gamma^f)) = 0 \implies H(p, V(p), V'_+(p)) < 0.$
2. $F(p, V(p), V'_-(p), (\gamma^f, 1]) = 0 \implies H(p, V(p), V'_-(p)) > 0.$

Proof. Consider (1). Suppose $H(p, V(p), V'_+(p)) \geq 0$. Then, $\sup_{\gamma \in [0, \gamma^f)} \gamma H(p, V(p), V'_+(p)) = \gamma^f H(p, V(p), V'_+(p))$. Therefore, $V(p) = V^f(p)$, a contradiction. Same holds for (2). □

Lemma 11. $P_t > \underline{p} \implies$ *wlog* $\gamma_t^* \in \{0, \gamma^f, 1\}$.

Proof. When $P_t \in \mathcal{D}$, $V(P_t)$ satisfies (9). Therefore, we have,

$$\gamma_t^* = \begin{cases} 1 & \text{if } H(P_t, V(P_t), V'(P_t)) > 0 \\ [0,1] & \text{if } H(P_t, V(P_t), V'(P_t)) = 0 \\ 0 & \text{if } H(P_t, V(P_t), V'(P_t)) < 0 \end{cases}$$

In particular, when $H(P_t, V(P_t), V'(P_t)) = 0$, we can set $\gamma_t^* = \gamma^f$.

On the other hand, if $P_t \in \mathcal{D}^C$ and $V(P_t) > V^f(P_t)$, then at least one of the two holds from Lemma 10

1. $F(P_t, V(P_t), V'_+(P_t), [0, \gamma^f)) = 0$ and $H(P_t, V(P_t), V'_+(P_t)) < 0$.
2. $F(P_t, V(P_t), V'_-(P_t), (\gamma^f, 1]) = 0$ and $H(P_t, V(P_t), V'_-(P_t)) > 0$.

Since an optimal control exists, if (1) holds, $\gamma_t^* = 0$, and if (1) fails but (2) holds, $\gamma_t^* = 1$. Lastly, if $V(P_t) = V^f(P_t)$, then $\gamma_t^* = \gamma^f$ delivers the requisite value.

□

As a consequence of Lemma 8, Lemma 10 and 11, we obtain the following corollary.

Corollary 1. $P_t > \underline{p}$ *and* $P_t \neq p^f \implies \gamma_t^* \in \{0, 1\}$.

Remark 3. *Notice that* $\dot{P}_t > 0$ *if* $\gamma_t^* = 0$ *and* $\dot{P}_t < 0$ *if* $\gamma_t^* = 1$.

Lemma 12. *If* $P_s = p$ *for all* $s \in [0, t]$ *for some* $t > 0$ *then* $\gamma_s^* = \gamma^f$ *for a.e.* $s \in [0, t]$.

Proof. Suppose $\gamma_s^* \in \{0, 1\}$ for a.e. $s \in [0, t]$.[29] Define, $A := \{t : \gamma_t^* = 0\}$. For $P_s = p$ for all $s \leq t$, we require, for any $0 \leq t_1 \leq t_2 \leq t$,

$$\int_{t_1}^{t_2} \mathbb{1}_A \mathrm{d}t = \int_{t_1}^{t_2} (1 - \mathbb{1}_A) \mathrm{d}t$$

In other words, we need $\ell(A \cap I) = \ell(I \backslash A) \implies \ell(A \cap I) = \frac{1}{2}\ell(I)$ for all intervals $I = [t_1, t_2]$.[30] Therefore,

$$\lim_{\epsilon \to 0} \frac{\ell(A \cap [s - \epsilon, s + \epsilon]}{2\epsilon} = \frac{1}{2}.$$

[29] We can simply ignore the times when $\gamma_s^* = \gamma^f$ and consider a smaller interval.
[30] $\ell(A)$ is the Lebesgue measure of set A.

On the other hand, by the Lebesgue density theorem, for a.e. $x \in S$,

$$\lim_{\epsilon \to 0} \frac{\ell(A \cap (x-\epsilon, x+\epsilon))}{2\epsilon} = 1$$

A contradiction. Therefore, no such A exists, i.e. $\gamma_s^* = \gamma^f$ for a.e. $s \in [0,t]$. □

Claim 7. *Suppose $P_0 = p$. If $\gamma_s^* = \gamma^f$ for a.e. $s \in [0,t]$ then $V(p) = V^f(p)$.*

Proof. Since the beliefs do not move between $[0,t]$, by Lemma 12, $\gamma_t = \gamma^f$ for a.e. $t \in [0,T]$. Therefore,

$$\begin{aligned} V(p) =& \lambda_g \gamma^f p v^1 \mathrm{d}t - c\mathrm{d}t + \lambda_b(1-p)(1-\gamma^f)v^1 \mathrm{d}t \\ &+ (1-\mathrm{d}t)(1-\lambda_g\gamma^f p \mathrm{d}t - \lambda_b(1-\gamma^f)(1-p)\mathrm{d}t)V(p) \\ \implies V(p) =& V^f(p) \end{aligned}$$

since $\lambda_g \gamma^f - \lambda_b(1-\gamma^f) = 0$. □

Corollary 2. *$V(p) = V^f(p) \implies \gamma_t^* = \gamma^f$ wlog for a.e. t and $p = p^f$.*

Definition 4. *A control α is called a Markov control if there exists a measurable function $g : [0,1] \to \mathcal{Y}$ such that $\alpha_t = g(P_t)$ for all $t \geq 0$. Let the space of Markovian controls be denoted by $\mathcal{M}$.*

Lemma 13. *Given an $\alpha \in \mathcal{U}$, $\exists \alpha' \in \mathcal{M}$ such that $J(\cdot, \alpha) = J(\cdot, \alpha')$.*

Proof. The stochastic control problem in (DMP) is the same as the first-passage problem in Kurtz and Stockbridge (1998). Theorem 5.5 in Kurtz and Stockbridge (1998) proves that for any admissible control α, there exists a Markovian control α' such that $J(\cdot, \alpha) = J(\cdot, \alpha')$. The paper formulates the stochastic control problem as a controlled martingale problem where the martingale is often defined through the infinitesimal generator.

To see the exact mapping, use $E = [0,1]$, $\Delta = 0$. For any $f \in \mathcal{C}^1([0,1])$ and a control $\alpha \in \mathcal{U}$, the infinitesimal generator for P_t is a map $A : \mathcal{D}(A) := \mathcal{C}^1([0,1]) \to \mathcal{C}([0,1] \times \mathcal{Y})$ as below.

$$\begin{aligned} Af(p) :=& \lim_{t \downarrow 0} \frac{\mathbb{E}[f(P_t^p)] - f(p)}{t} \\ =& a\Big[\lambda_g \gamma p [f(1) - f(p)] + \lambda_b(1-\gamma)(1-p)[f(0) - f(p)] \\ &+ [\lambda_b(1-\gamma) - \lambda_g \gamma] p(1-p) f'(p)\Big] \end{aligned}$$

Notice that $|Af(x,u)| \leq (\lambda_b + \lambda_g)(||f|| + ||f'||)$. Hence conditions (i)-(vi) apply enabling us to invoke Theorem 5.5. □

Remark 4. *To be precise, Theorem 5.5 in Kurtz and Stockbridge (1998) gives a payoff equivalent "relaxed" Markovian control, i.e. $\alpha : [0,1] \to \Delta(\mathcal{Y})$. However, given the linearity in the law of motion and the arrival rates of news, such mixing can be done away with at no additional cost.*

Therefore, it is without loss to restrict attention to Markovian controls. Moreover, due to Lemma 11, we can restrict attention to $\{0, \gamma^f, 1\}$-valued Markovian controls for γ^*. We will use γ_t^* and $\gamma^*(p)$ interchangeably at the risk of abusing notation.

Lemma 14. ***Beliefs move only in one direction:*** *Suppose $P_0 = p$ and $P_s = q < p$ for some s. Then, $P_u \notin [p, q)$ for any $u > s$. Similarly, if $P_s = q > p$ then $P_u \notin (q, p]$ for any $u > s$.*

Proof. Notice that γ_t^* is uniquely pinned down by P_t by Lemma 13. Suppose, $P_s = q < p$ and $P_u = r \in (q, p)$ for some $u > s$. Let, wlog, $s = \inf\{t \geq 0 : P_t^p = q\}$. Since $\min\{\lambda_b, \lambda_g\} \leq |\dot{P}_t| \leq \max\{\lambda_b, \lambda_g\}$, $s > \epsilon$ for some $\epsilon > 0$. Choose a δ small enough such that $P_{s-\eta}^p = q + \delta$ for some $\eta > 0$ and $P_t^p < r$ for all $t \in [s - \eta, s]$. Such a δ exists because $\dot{P}_t$ is bounded. If $P_u^p = r$ for some $u > s \implies P_{s+\eta_1}^p = P_{\eta_1}^q = q + \delta$ for some $\eta_1 = \{\inf t \geq 0 : P_{s+t}^q = q + \delta\}$. Obviously, $\eta_1 > 0$. However, by the Markov property, $P_{s+\eta_1+\eta}^p = P_\eta^{q+\delta} = q$. Moreover, if δ is sufficiently small, $P_t^p < r$ for all $t \in [s.s + \eta_1 + \eta]$. Repeating the argument, $P_t^p < r$ for all $t \in [s + \eta_1 + \eta, s + K(\eta_1 + \eta)]$ for any $K \in \mathbb{N}$. Therefore, $P_u = r$ is not possible for any $u \in \mathbb{R}$. A contradiction.

Similar argument holds for the reverse inequality. □

Lemma 15. *If $P_0 = p$ and $P_t = q < (>) p$. Then, $\gamma_s^* = 1(0)$ for a.e. $s \in [0, t]$.*

Proof. Let us argue when $q < p$. The reverse argument is identical. Suppose $\gamma_s^* \neq 1$ for a.e. $s \in [0, t]$. Let $A := \{s : \gamma_s^* = 0\}$. By hypothesis, $\ell([0, t]) > \ell(A) > 0$. By Lemma 14, P_s is decreasing in s. Therefore, for any interval $[t_1, t_2] \subset [0, t]$, $\int_{t_1}^{t_2} \gamma_s^* ds < 0$. That is, $\ell([t_1, t_2] \backslash A) > \ell(A \cap [t_1, t_2])$. Therefore, $\ell(A \cap [t_1, t_2]) < \frac{1}{2}\ell([t_1, t_2])$.

Once again, by the Lebesgue density theorem,

$$\lim_{\epsilon \downarrow 0} \frac{\ell(A \cap [s - \epsilon, s + \epsilon])}{2\epsilon} = 1$$

for a.e. $s \in A$. Therefore, $\ell(A \cap [s - \epsilon, s + \epsilon) < \frac{1}{2}\ell([s - \epsilon, s + \epsilon]$ is not possible.

□

Corollary 3. *In the optimal policy, only one of the following happens (in the absence of conclusive news):*

1. *"Always look for S": $\gamma_t^* = 1$ for all t. Beliefs move down until p_S.*
2. *"Look for F and freeze": $\gamma_t^* = 0$ if $P_t < p^f$ and $\gamma_t^* = \gamma^f$ otherwise. Beliefs move up and freeze at p^f (Corollary 2).*
3. *"Look for S and freeze": $\gamma_t^* = 1$ if $P_t > p^f$ and $\gamma_t^* = \gamma^f$ otherwise. Beliefs move down and freeze at p^f.*
4. *"Always look for F": $\gamma_t^* = 0$ for all t. Beliefs move up.*

Claim 8. *"Always look for F" is never optimal.*

Proof. This is obvious as looking for F can only produce $v^0 \leq 0$ while incurring the cost. So $J(p, \gamma_B) < 0 \leq V(p)$ where γ_B is the constant control of $\gamma = 0$, i.e. looking for failure. □

Recall that $V_{FS}(p)$ is the value delivered by the "FS policy with a switch at p^f (8), while $V_S(p)$ is the value function of the policy of using only the success arm and stopping at the optimal stopping cutoff as in Keller et al. (2005), refer to Proposition 1 for details).

Lemma 16. $V(\cdot) = \max\{V_S(\cdot), V_{FS}(\cdot)\}$.

Proof. By Corollary 3, the optimal policy must either be either looking for S forever or a FS policy with a switch at p^f. The former delivers the value of $V_S(\cdot)$ while the latter delivers $V_{FS}(\cdot)$. Moreover, if $V(p) = 0$ it is optimal to quit and take the outside option. □

Lemma 17. *If $V_S(q) = V_{FS}(q)$ for some $q \in [p^f, 1)$ then $V_S(q) = V_{FS}(q)$ for all $q \in [p^f, 1]$.*

Proof. Suppose, $V_S(q) = V_{FS}(q)$ for some $q \in [p^f, 1)$. On $(p^f, 1)$, both $V_S(\cdot)$ and $V_{FS}(\cdot)$ satisfy the following differential equation.

$$\begin{aligned} f(p) =& -c + \lambda_g p(v^1 - f(p)) - \lambda_g p(1-p) f'(p) \\ f'(p) =& \frac{c - \lambda_g p v^1 + (1 + \lambda_g p) f(p)}{\lambda_g p (1-p)} \end{aligned}$$

Standard results in the theory of first order linear differential equations imply that if f, g satisfy the above and $f(x) = g(x)$ for some $x \in (p^f, 1)$, then $f(\cdot) = g(\cdot)$ on $(p^f, 1)$.

□

Lemma 18. $V(p) = V_{FS}(p) > 0 \implies V(q) = V_{FS}(q)$ *for all* $q \geq p$.

Proof. Suppose $V(p) = V_{FS}(p)$ for some p and for some $q > p$ $V(q) = V_S(q) > V_{FS}(q)$. By Lemma 17, if $V_S(\cdot)$ and $V_{FS}(\cdot)$ intersect, they must intersect before p^f. Therefore, $q < p^f$.

First, it is without loss to follow the control $\alpha^{FS}_{p^f}$ starting from p to deliver $V(p)$. Let T be the time it takes for the beliefs to reach q starting from p if the agent follows $a_t = 1$ and $\gamma_t = 0$. Consider the following control α' such that $a'_t = 1$ whenever $P_t^{p,\alpha'} \geq p_S$ and

$$\gamma'_t = \begin{cases} 0 & \text{if } t \leq T \\ 1 & \text{if } t > T \text{ and } P_t \geq p_S \end{cases}$$

Therefore,

$$\begin{aligned} V(p) = J(p, \alpha_{p^f}^{FS}) &= \int_0^T (1-p)\lambda_b e^{-\lambda_b t}(1-e^{-t})(-c)\mathrm{d}t + [p + (1-p)e^{-\lambda_b T}]e^{-T}V_{FS}(q) \\ &< \int_0^T (1-p)(1-e^{-t})\lambda_b e^{-\lambda_b t}(-c)\mathrm{d}t + [p + (1-p)e^{-\lambda_b T}]e^{-T}V_G(q) \\ &= J(p, \alpha') \end{aligned}$$

A contradiction. □

Below, we prove Proposition 4 using the series of results so far.

Proof of Proposition 4. By Lemma 16, we have $V(\cdot) = \max\{V_S(\cdot), V_{FS}(\cdot)\}$. By Lemma 18, if $0 < V(p) = V_{FS}(p) \implies V(q) = V_{FS}(q)$ for all $q \geq p$. Therefore, we have only three possibilities:

1. $V(\cdot) = V_S(\cdot)$. Here, the optimal policy is α^S.
2. $$V(p) = \begin{cases} V_S(p) & \text{if } p \leq p_1 \\ V_{FS}(p) & \text{if } p > p_1 \end{cases}$$
   Here, the optimal policy is α_{p_1,p^f}^{SFS}.
3. $V(\cdot) = V_{FS}(\cdot)$. Here, the optimal policy is $\alpha_{p^f}^{FS}$.

□

C. Proofs from Section 4

Proof of Proposition 7. First, fix a λ_g, c such that $v^0 + \frac{c}{1+\lambda_g} > 0$.

The objective is to produce two values of $v^1 := \Gamma - 1$, say v_1^1, v_2^1, such that, ceteris paribus, when $v^1 = v_1^1$, $\alpha^*(v_1^1) = \alpha^S$ and $\alpha^*(v_2^1) = \alpha_{p^f}^{FS}$.[31] To get α^S to be optimal, a sufficient condition is $V^f(1) < 0$. Notice that,

$$V^f(1) = \frac{\lambda_g v^1 - (1 + \frac{\lambda_g}{\lambda_b})c}{1 + \frac{\lambda_g}{\lambda_b} + \lambda_g}$$

Choose v_1^1 so that $\lambda_g v_1^1 \in \left(c, (1 + \frac{\lambda_g}{\lambda_b})c\right)$, we can satisfy Assumption 2 and have $v^f(1) < 0$. Therefore, for any such v_1^1, $\alpha^*(v_1^1) = \alpha^S$.

For v_2^1, notice that

$$J(p, \alpha_{p^f}^{FS}) \to v^0 + p\left[\frac{\lambda_g v^1 - c - (1+\lambda_g)v^0}{1+\lambda_g}\right] =: J_\infty(p)$$

[31] In this context, $\alpha^*(v)$ denotes the optimal policy in Proposition 4 when $v^1 = v$.

pointwise as $\lambda_b \to \infty$. We would like to emphasize that p^f changes with λ_b above, and therefore so does $J(\cdot, \alpha^{FS}_{p^f})$. Notice that

$$J_\infty(p_S) = v^0 + \frac{c}{\lambda_g(1+\lambda_g)}\left[\lambda_g - (c + (1+\lambda_g)v^0)/v^1\right].$$

Again, we emphasize that p_S changes with v_1. Therefore, in the absence of uniform convergence, we need to have the correct order of limits in order to obtain the right optimal policy. Therefore, we first choose v^1 large enough so that $J_\infty(p_S) > 0$. Let this by v_2^1. Fixing v_2^1 also fixed p_S, and therefore, we can invoke the pointwise convergence to argue that $\exists \lambda_b$ such that $J(p_S, \alpha^{FS}_{p^f}) > 0$ as $J(p_S, \alpha^{FS}_{p^f}) \to J_\infty(p_S)$.

Moreover, for any finite λ_b, $c, (1 + \frac{\lambda_g}{\lambda_b})c)$ is non-empty. Therefore, we have found v_1^1, v_2^1, and λ_b such that $\alpha^*(v_1^1) = \alpha^S$ and $\alpha^*(v_2^1) = \alpha^{FS}_{p^f}$. Lastly, $v_k^1 + 1 = \Gamma_k$, $k \in \{1, 2\}$ give Γ_1, Γ_2 stated in the statement of Proposition 7.

□

References

Bardi, M. and I. Capuzzo-Dolcetta (2008). *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*. Springer Science & Business Media.

Bergemann, D. and U. Hege (2005). The financing of innovation: Learning and stopping. *RAND Journal of Economics*, 719–752.

Blackwell, D. (1953). Equivalent comparison of experiments. *Annals of Mathematics and Statistics 32*, 265–272.

Chade, H. and N. Kovrijnykh (2016). Delegated information acquisition with moral hazard. *Journal of Economic Theory 162*, 55 – 92.

Che, Y.-K. and K. Mierendorff (2016). Optimal sequential decision with limited attention. *unpublished, Columbia University*.

Damiano, E., H. Li, and W. Suen (2017). Learning while experimenting. *University of British Columbia working paper*.

Fudenberg, D., P. Strack, and T. Strzalecki (2017). Stochastic choice and optimal sequential sampling.

Garfagnini, U. (2011). Delegated experimentation.

Guo, Y. (2016). Dynamic delegation of experimentation. *American Economic Review 106*(8), 1969–2008.

Halac, M., N. Kartik, and Q. Liu (2016). Optimal contracts for experimentation. *The Review of Economic Studies 83*(3), 1040–1091.

Hidir, S. (2017). Contracting for experimentation and the value of bad news. *o Unpublished*.

Hrner, J. and L. Samuelson (2013). Incentives for experimenting agents. *The RAND Journal of Economics 44*(4), 632–663.

Keller, G. and S. Rady (2015). Breakdowns. *Theoretical Economics 10*(1), 175–202.

Keller, G., S. Rady, and M. Cripps (2005). Strategic experimentation with exponential bandits. *Econometrica 73*(1), 39–68.

Kurtz, T. G. and R. H. Stockbridge (1998). Existence of markov controls and characterization of optimal markov controls. *SIAM Journal on Control and Optimization 36*(2), 609–653.

Kuvalekar, A. V. and E. Lipnowski (2018). Job insecurity. *SSRN working paper*.

Levitt, S. D. and C. M. Snyder (1997). Is no news bad news? information transmission and the role of "early warning" in the principal-agent model. *The RAND Journal of Economics 28*(4), 641–661.

Liang, A. and X. Mu (2018). Overabundant information and learning traps. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pp. 71–72. ACM.

Liang, A., X. Mu, and V. Syrgkanis (2017). Optimal learning from multiple information sources. *arXiv preprint arXiv:1703.06367*.

Manso, G. (2011). Motivating innovation. *The Journal of Finance 66*(5), 1823–1860.

Mayskaya, T. (2017). Dynamic choice of information sources.

Strulovici, B. and M. Szydlowski (2012). On the smoothness of value functions. Technical report, Discussion Paper, Center for Mathematical Studies in Economics and Management Science.