Endogenous Experimentation in Organizations*

Germán Gieczewski[†] **Svetlana Kosterina**[‡]

March 2020

Abstract

We study policy experimentation in organizations with endogenous membership. An organization initiates a policy experiment and then decides when to stop it based on its results. As information arrives, agents update their beliefs, and become more pessimistic whenever they observe bad outcomes. At the same time, the organization's membership adjusts endogenously: unsuccessful experiments drive out conservative members, leaving the organization with a radical median voter. We show that, under mild conditions, the latter effect dominates. As a result, policy experiments, once begun, continue for too long. In fact, the organization may experiment forever in the face of mounting negative evidence. This result provides a rationale for obstinate behavior by organizations, and contrasts with models of collective experimentation with fixed membership, in which under-experimentation is the typical outcome.

Keywords: experimentation, dynamics, median voter, endogenous population

1 Introduction

Organizations frequently face opportunities to experiment with promising but untested policies. Two conclusions follow from our understanding of experimentation

*We would like to thank Daron Acemoglu, Avidit Acharya, Alessandro Bonatti, Glenn Ellison, Robert Gibbons, Faruk Gul, Matias Iaryczower, Navin Kartik, Wolfgang Pesendorfer, Caroline Thomas, Juuso Toikka, the audiences at the 2018 American Political Science Association Conference (Boston), the 2018 Columbia Political Economy Conference, the 2019 Canadian Economic Theory Conference (Montreal) and the 2020 Utah Winter Organizational and Political Economics Conference, as well as seminar participants at MIT, Princeton, Indiana University Kelley School of Business, University of Calgary and UCEMA for helpful comments. All remaining errors are our own.

†Department of Politics, Princeton University.

‡Department of Politics, Princeton University.

to date. First, experimentation should respond to information. That is, when a policy experiment performs badly, agents should become more pessimistic about it, and if enough negative information accumulates, they should abandon it. Second, when experimentation is collective, the temptation to free-ride and fears that information will be misused by other agents lower incentives to experiment. Thus organizations should experiment too little. However, history is littered with examples of organizations that have stubbornly persisted with unsuccessful policies to the bitter end. This presents a puzzle for our understanding of decision-making in organizations.

Consider, for example, the case of Theranos, a Silicon Valley start-up founded by Elizabeth Holmes in 2003. Theranos sought to produce a portable machine capable of running hundreds of medical tests on a single drop of blood. If successful, Theranos would have revolutionized medicine, but its vision was exceedingly difficult to realize. Over the course of ten years, the firm invested over a hundred million dollars into trying to attain Holmes's vision, while devoting little effort to developing a more incremental improvement over existing technologies as a fall-back plan. Theranos eventually launched in 2013 with a mixture of inaccurate and fraudulent tests, and the ensuing scandal irreversibly damaged the company.

Up to the point where Theranos began to engage in outright fraud, a pattern repeated itself. The company would bring in high-profile hires and create enthusiasm with its promises, but once inside the organization, employees and board members would gradually become disillusioned by the lack of progress.[1] As a result, many left the company,[2] with those who were more pessimistic about Theranos's prospects being more likely to leave than those who saw Holmes as a visionary. While the board came close to removing Holmes as CEO early on, she managed to retain control for many years after, because too many of the people who had lost faith in her leadership had left the organization before they could have a majority.

The takeaway from this example is that Theranos experimented for too long in spite of not succeeding precisely because the pessimistic members of the organization kept leaving. Motivated by this and similar examples, we propose an explanation for

[1]For instance, Theranos's lead scientist, Ian Gibbons, told his wife that nothing at Theranos was working, years after joining the company. See Carreyrou (2018).

[2]For example, board member Avie Tevanian, while mulling over a decision to buy more shares of the company at a low price, was asked by a friend: "Given everything you now know about this company, do you really want to own more of it?" See Carreyrou (2018).

obstinate behavior by organizations that rests on three key premises. First, agents disagree about the merits of different policies, that is, they have heterogeneous prior beliefs. Second, the membership of organizations is fluid: agents are free to enter and leave in response to information. Third, the organization's policies are responsive to the opinions of its members.

We operationalize these assumptions in the following model. An organization chooses between a safe policy and a risky policy in each period. The safe policy yields a flow payoff known by everyone, while the risky policy yields an uncertain flow payoff. There is a continuum of agents. In every period, each agent decides whether to work for the organization or independently. The payoff of working for the organization depends on its policy. The outside option yields a guaranteed flow payoff.

All agents want to maximize their returns but hold heterogeneous prior beliefs about the quality of the risky policy. As long as agents work for the organization, they remain voting members of the organization and vote on the policy it pursues. We assume that the median voter—that is, the member with the median prior belief—chooses the organization's policy. Whenever the risky policy is used, the results are publicly observed.

We show that experimentation in organizations is inefficient in two ways that are novel to the literature. First, there is over-experimentation relative to the social optimum.[3] Over-experimentation takes a particularly stark form: the organization experiments forever regardless of the outcome. Second, the organization may experiment more if the risky policy is bad than if it is good. In other words, the organization's policy may respond to information in a perverse way.

Our main result provides a simple necessary and sufficient condition under which perpetual experimentation is the unique equilibrium outcome. The condition requires that, at each history, the pivotal agent prefers perpetual experimentation to no experimentation. We then analyze the comparative statics. We show that perpetual experimentation is more likely when the outside option is more attractive, the organization's safe policy is less attractive, agents are more patient, and the distribution of prior beliefs contains more optimists in the MLRP sense.

[3] In particular, we show that there is too much experimentation from the point of view of all agents.

Two forces affect the amount of experimentation in our model. On the one hand, the median member of an organization is reluctant to experiment today if she anticipates losing control of the organization tomorrow as a result. On the other hand, if no successes are observed, as time passes, only the most optimistic members remain in the organization, and these are precisely the members who want to continue experimenting the most. The first force makes under-experimentation more likely, while the second pushes the organization to over-experiment. Ex-ante, it is not obvious which force will dominate. We show that the second force often dominates. This underpins the main result of our paper.

We next show that our result of perpetual experimentation is robust. Our baseline model features perfectly informative good news. We show that perpetual experimentation also obtains under bad news and imperfectly informative good news. In addition, a novel result arises when news are imperfectly informative: for appropriately chosen parameter values, there is an equilibrium in which the organization stops experimenting with a strictly positive probability *only if enough successes are observed*. Hence, counterintuitively, the organization is more likely to experiment forever if the technology is bad than if it is good.[4] The implication is that self-selection of agents into organizations may not only induce excessive experimentation overall—it may also cause organizations to actively radicalize in the face of failure. Conversely, success may make organizations more conservative and prone to backing away from the very strategies that brought them success.

Finally, we show that our result is robust to general voting rules; settings in which the members' flow payoffs, or the learning rate, depend on the size of the organization; an alternative specification in which agents have common values; and a specification in which the size of the organization is fixed rather than variable and the organization hires agents based on ability.

The rest of the paper proceeds as follows. Section 2 discusses the applications of the model. Section 3 reviews the related literature. Section 4 introduces the baseline model. Section 5 analyzes the set of the equilibria in the baseline model. Section 6 considers other learning processes. Section 7 develops other extensions of the model, such as general voting rules and settings where the members' flow payoffs depend on

[4]Approximately, for this to happen, it is sufficient for the distribution of prior beliefs to be single-dipped enough over some interval.

the size of the organization.

2 Applications

Our model has a variety of applications besides the one discussed in the Introduction. In this Section, we discuss how our assumptions map to several settings such as cooperatives, non-profits, activist organizations, firms and political parties.

We first present another example of experimentation in a firm that fits our model. Consider the decline of Blockbuster, the once-ubiquitous video rental chain that went bankrupt in 2010. In the early 2000s, Blockbuster began to face competition from companies like Netflix and Redbox, which offered DVD rentals over mail or movie downloads over the Internet. In the face of this competition, Blockbuster's strategy of sticking to their traditional retail store business model became a risky experiment with uncertain outcomes.[5] In the beginning the managerial staff at Blockbuster was optimistic about their business model and skeptical of their competitors'.[6]

By 2004, when the CEO came to see Netflix as a threat and decided to compete with it, Blockbuster's majority owner had lost faith in the company and sold its stake, a large share of which was bought by an activist investor who was skeptical of Blockbuster's efforts to expand into the digital market.[7] Eventually, in 2007, Blockbuster launched a major expansion into the online market and attempted a partial merger with Netflix. The board responded by firing the CEO, and their chosen replacement reversed course and turned his focus to brick-and-mortar stores.[8] Blockbuster went bankrupt in 2010.[9]

---

[5]In this example, every alternative Blockbuster had access to was risky. Our assumption that the organization has access to one risky policy and one safe policy is for simplicity; the analysis can be extended to a case with multiple risky policies.

[6]John Antioco, the CEO of Blockbuster at the time, "didn't believe that technology would threaten the company as fast as critics thought" when he decided to join in 1998, and this belief was one of his reasons for joining (Antioco 2011).

[7]Antioco (2011) said of the investor and his allied board members: "Mostly, though, they questioned our strategy, which focused on growing an online business and finding new ways to satisfy customers [...]."

[8]The new CEO said in 2008 of the online market: "Should we put shareholder money at risk in a market that's at best five years away from being commercial? I don't think so," and of their competitors, "Neither Redbox nor Netflix are even on the radar screen in terms of competition." See `https://www.fool.com/investing/general/2008/12/10/blockbuster-ceo-has-answers.aspx`.

[9]In a similar vein, other once-dominant companies such as Kodak and Sears, when faced with

Next, consider experimentation in a cooperative. Agents are individual producers who own factors of production. In a dairy cooperative, for example, each member owns a cow. The agent can manufacture and sell his own dairy products independently or he can join the cooperative. If he joins, his milk will be processed at the cooperative's plants, which benefit from economies of scale. The cooperative can choose from a range of dairy production policies, some of which are riskier than others. For instance, it can limit itself to selling fresh milk and yogurt, or it can develop a new line of premium cheeses that may or may not become profitable. Dairy farmers have different beliefs about the market viability of the latter strategy. Should this strategy be used, only the more optimistic farmers will choose to join or remain in the cooperative. Moreover, cooperatives typically allow their members to elect directors.

In the case of activist organizations, agents are citizens seeking to change the government's policy or the behavior of multinational corporations. Agents with environmental concerns can act independently by writing to their elected representatives, or they can join an organization, such as Greenpeace, that has access to strategies not available to a citizen acting alone, such as lobbying, demonstrations, or direct action—for instance, confronting whaling ships. While all members of the organization want to bring about a policy change, their beliefs as to the best means of achieving this goal differ. Some support safe strategies, such as lobbying, while others prefer riskier ones, such as direct action.

An organization that employs direct action will drive away its moderate members, increasingly so if its attempts are unsuccessful. The resulting self-selection can sustain a base of support for extremist strategies. Our model can thus explain the behavior of fringe environmental groups, such as Extinction Rebellion, Animal Liberation Front and Earth Liberation Front that engage in ecoterrorism and economic sabotage in spite of the apparent ineffectiveness of their approach.[10] The same logic applies to other forms of activism, as well as to charitable organizations choosing between more or less widely understood poverty alleviation tactics, for example, cash

growing competition from new technologies such as digital cameras and online shopping respectively, failed to adapt and instead clung to outdated business models that were seen by outsiders as increasingly unprofitable (Gavetti et al. 2004; Colvin and Wahba 2019).

[10]See, for example, `https://www.theguardian.com/commentisfree/2019/apr/19/extinction-rebellion-climate-change-protests-london`.

transfers as opposed to microcredit.[11]

Our model is also relevant to the functioning of political parties. Here agents are potential politicians or active party members, and the party faces a choice between a widely understood mainstream platform—for example, social democracy—and an extremist one which may be vindicated or else fade into irrelevance. A communist platform that claims the collapse of capitalism is imminent is an example of the latter. Again, the selection of extremists into extremist parties, which intensifies when such parties are unsuccessful, explains their rigidity in the face of setbacks. For example, the French Communist Party declined from a base of electoral support of roughly 20% in the postwar period to less than 3% in the late 2010s.[12] Despite this dramatic decline, partly caused by the demise of the Soviet Union, they have preserved the main tenets of their platform, such as the claim that the capitalist system is on the verge of collapse.

3 Related Literature

This paper is related to the literature on strategic experimentation with multiple agents (Keller et al. 2005, Keller and Rady 2010, Keller and Rady 2015, Strulovici 2010), as well as the literature on dynamic decision-making in clubs (Acemoglu et al. (2008, 2012, 2015), Roberts 2015, Bai and Lagunoff 2011, Gieczewski 2019).

In Keller, Rady and Cripps (2005), multiple agents with common priors control two-armed bandits of the same type which may have breakthroughs at different times. In their model, there is under-experimentation due to free-riding. In contrast, we study an organization making a single collective decision in each period about whether to experiment. While the organization experiments, members can exit and collect an outside option, but at the cost of their voting rights. The selection of optimists into the organization causes excessive experimentation.[13]

[11]Note that in these examples agents should be modeled as having common values, since agents benefit from a change in public policy regardless of how much their actions contributed to it. Although we write our main model for the case of private values, we show in Section 7 that our main results survive in the common values setting. Another way to accommodate common values in our model is to endow agents with expressive payoffs, whereby agents benefit not just from a policy change but also from having participated in the efforts that brought it about.

[12]See, for example, Bell (2003), Bréchon (2011) and Damiani and De Luca (2016).

[13]While there is free-riding in the sense that outsiders benefit from the option value of experimen-

In Strulovici (2010) a community of agents decides by voting whether to collectively experiment with a risky technology. Agents have common priors, but experimentation gradually reveals some of them to be *winners* and others to be *losers* from the risky technology. In equilibrium, there is too little experimentation because agents fear being trapped into using a technology that turns out to be bad for them.

A similar motive to under-experiment is present in our model. Indeed, consider an agent who would prefer to experiment today but not tomorrow. If she anticipates that learning will result in an extreme optimist coming to power tomorrow, then she may choose not to experiment today, lest she be forced to over-experiment or switch to her inefficient outside option. However, there are two important differences between our model and Strulovici's. First, in our model, agents can switch to an outside option at the cost of their voting rights. This novel assumption is what allows for the organization to be captured by extremists. Second, in Strulovici's model, learning exacerbates the conflict between agents, while in our model learning helps agents converge to a common belief.

The literature on decision-making in clubs studies dynamic policy-making in a setting where control of the club depends on policy choices but there is no uncertainty about the consequences of policies. Instead, different agents prefer different policies. The present paper shares with this strand of literature (Acemoglu et al. (2008, 2012, 2015), Roberts 2015, Bai and Lagunoff 2011) the feature that the policy chosen by the pivotal decision-maker today affects the identities of future decision-makers, leading agents to fear that myopically attractive policies may lead to a future loss of control. Most closely related is Gieczewski (2019), which, like this paper, studies a setting in which agents can choose to join an organization or stay out and are only able to influence the policy if they are members. The present paper differs in considering agents with heterogeneous beliefs rather than preferences and in allowing for learning from new information.[14]

tation, it is not socially costly because we assume the learning rate to be independent of the size of the organization. See Section 7 for an extension with an endogenous learning rate.

[14]Because disagreements over policy arise only from differences in prior beliefs, which vanish as information accumulates, we might expect that in the long run the organization's policy will converge to the one desired by most agents. However, this is not what happens: rather, learning can lead to a capture of the organization by extremists.

4 The Model

Time $t \in [0, \infty)$ is continuous. There is an organization that has access to a *risky* policy and a *safe* policy. The risky policy is either *good* or *bad* and its type is persistent. We use the notation $\theta = G, B$ for each respective scenario.

The world is populated by a continuum of agents, represented by a continuous density f over $[0, 1]$. The position of an agent in the interval $[0, 1]$ indicates her beliefs: an agent $x \in [0, 1]$ has a prior belief that the risky policy is good with probability x. All agents discount the future at rate γ.

At every instant, each agent chooses whether to be a member of the organization. Agents can enter and leave the organization at no cost.[15] Agents who choose not to be members work independently and obtain a guaranteed flow payoff s. The flow payoffs of members depend on the organization's policy.

Whenever the organization uses the safe policy ($\pi_t = 0$), all members receive a guaranteed flow payoff r. When the risky policy is used ($\pi_t = 1$), its payoffs depend on the state of the world. If the risky policy is good, it succeeds according to a Poisson process with rate b. If the risky policy is bad, it never succeeds. Each time the risky policy succeeds, all members receive a lump-sum unit payoff. At all other times, the members receive zero while the risky policy is used.

We assume that $0 < s < r < b$. This implies that the organization's safe policy is always preferable to working independently. Moreover, the risky policy would be the best choice were it known to be good, but the bad risky policy is the worst of all options.

When the risky policy is used, its successes are observed by everyone, and agents update their beliefs based on this information. By Bayes' rule, the posterior belief of an agent with prior x who has seen k successes after experimenting for a length of time τ is

$$\frac{x}{x + (1 - x)L(k, \tau)}$$

where $L(k, \tau) = \mathbb{1}_{k=0} e^{b\tau}$. Since $L(k, \tau)$ serves as a sufficient statistic for the information observed so far, suppressing the dependence of $L(k, \tau)$ on k and τ, we take

[15] Our main results survive if we assume that agents cannot reenter after exiting.

$L = L(k, \tau)$ to be the state variable in our model and hereafter define $p(L, x)$ as the posterior of an agent with prior x given that the state variable is L.

We assume that the organization is using the risky policy in the beginning of the game ($\pi_0 = 1$),[16] and at every instant $t > 0$ the median member of the organization chooses whether the organization should continue to experiment at that instant.[17]

We let $m(L, \pi)$ denote the median member given the state variable L and the incumbent policy π. The formal definition of equilibrium is given in the Appendix. Here we provide an informal definition.

Definition 1. An equilibrium satisfies the following:

(i) Agents choose whether to be members based on their flow payoffs.

(ii) If $m(L, \pi)$ prefers policy π' to the alternative, then policy π' is chosen.

(iii) If $m(L, \pi)$ gets the same continuation utility under either policy choice, but would get strictly higher utility from 'locking in' policy π' for any small length of time $\epsilon > 0$, then π' is chosen.

The reason that agents make membership decisions based on their flow payoffs is that there is a continuum of agents, so an agent obtains no value from her ability to vote.

Part (ii) of the definition rules out equilibria in which agents join the organization despite disliking its policy, because they expect other like-minded members to join at the same time and immediately change the policy. Part (iii) rules out unattractive equilibria in which weakly dominated policies are chosen. For instance, even under common knowledge that the risky policy is good, there would be an equilibrium in which all decision-makers choose the safe policy because any deviation to the risky policy would be reversed immediately.

The equilibrium we define can be obtained as a limit of the equilibria of a discrete-time game in which membership and policy decisions are made at times

[16] If the organization starts with the safe policy at $t = 0$, two outcomes are possible. If the initial median prefers the equilibrium continuation resulting from switching to the risky policy, she will switch immediately, and the solution is the same as when starting with the risky policy. If not, the safe policy is used forever.

[17] For the median to be well-defined, the set of members must be Lebesgue-measurable. This will not be an issue, since in equilibrium the set of members is always an interval.

$t \in \{0, \epsilon, 2\epsilon, \ldots\}$ with $\epsilon > 0$ small. In this game, at each time t in $\{0, \epsilon, 2\epsilon, \ldots\}$, first the incumbent median chooses a policy π_t for time $[t, t+\epsilon)$, and then all agents choose whether to be members. The agents who choose to be members at time t – and hence accrue the flow payoffs generated by policy π_t – are the incumbent members at time $t + \epsilon$. The median of this set of members then chooses $\pi_{t+\epsilon}$.

Note that our definition is a special case of Markov Perfect Equilibrium, as we only allow the strategies to condition on the information about the risky policy revealed so far and on the existing policy (which determines the identity of the current median voter).

5 Equilibria in the Baseline Model

In this section we characterize the equilibria of the model described above. The presentation of the results is structured as follows. We first explain who the members of the organization are depending on what has happened in the game so far. We then make several observations which allow us to reduce the problem of equilibrium characterization to finding an optimal stopping time. Next, we state our first main result, which shows that the organization may experiment forever and provides simple sufficient conditions for this to happen (Propositions 1 and 9). Finally, in Proposition 10 we characterize the equilibria in the alternative case where experimentation cannot go on forever.

We start with three useful observations. First, because the bad risky policy never succeeds, the posterior belief of every agent with a positive prior jumps to 1 if a success is observed. Since $b > r$, s, if a success is ever observed, the risky policy is always used thereafter, and all agents enter the organization and remain members forever.

Second, recall that, whenever the risky policy is being used, the set of members is the set of agents for whom $p(L, x)b \geq s$. It is clear that, for any $L > 0$ (that is, if no successes have been observed), $p(L, x)$ is increasing in x. That is, agents who are more optimistic at the outset remain more optimistic after observing additional information. Hence the set of members is an interval of the form $[y_t, 1]$.

Third, since $r > s$, whenever the safe policy is used, all agents choose to join the organization, and the population median becomes the pivotal decision-maker. Observe that the population median is more pessimistic than the median of any interval of the form $[y, 1]$ with $y > 0$. In particular, she is more pessimistic than the median voter of the organization before a switch to the safe policy. Thus if the median of the organization preferred to switch to the safe policy, so does the population median. Because no further learning happens when the safe policy is used, a switch to the safe policy is permanent.

The above observations imply that an equilibrium path must have the following structure. The risky policy is used until some time $T \in [0, \infty]$. If it succeeds by then, it is used forever. Otherwise, the organization switches to the safe policy at time T.[18] While no successes are observed, agents become more pessimistic over time and the organization becomes smaller. As soon as a success occurs or the organization switches to the safe policy, all agents join and remain members of the organization forever, and no further learning occurs.

Proposition 1 states our first main result. The result provides a simple condition that is sufficient for over-experimentation to arise in equilibrium. More specifically, if this condition is satisfied, then the organization uses the risky policy forever regardless of its results.

To state Proposition 1, we will need the following definitions. We let $V(x)$ denote the continuation utility of an agent with posterior belief x at time t, provided that she expects experimentation to continue for all $s \geq t$. We let m_t denote the median voter at time t provided that the organization has experimented unsuccessfully up to time t, and we let $p_t(m_t)$ denote m_t's posterior in this case.

Proposition 1. *If $V(p_t(m_t)) > \frac{r}{\gamma}$ for all t, there is an essentially unique*[19] *equilibrium. In this equilibrium, the organization experiments forever. If $\inf_{t \geq 0} V(p_t(m_t)) < \frac{r}{\gamma}$, there is no equilibrium in which the organization experiments forever.*

The intuition behind Proposition 1 is illustrated in Figure 1. As the organization experiments unsuccessfully on the equilibrium path, all agents become more

[18] If $T = \infty$, the risky policy is used forever.

[19] If $V(p_t(m_0)) = \frac{r}{\gamma}$ for some t, and the organization were to stop at time t, after all agents enter, the population median m_0 would be indifferent between switching back to the risky policy or not. Hence there are multiple equilibria in this case, but they only differ in their off-path behavior.

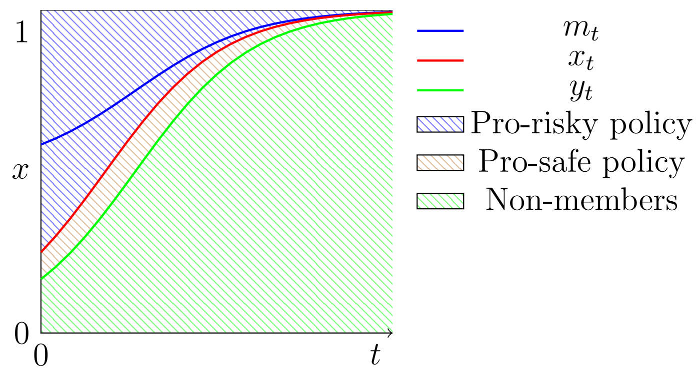

Figure 1: Median voter, indifferent voter, and marginal member on the equilibrium path

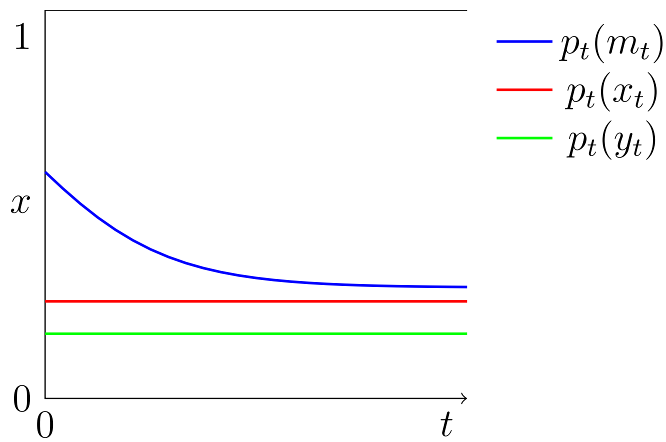

Figure 2: Posterior beliefs on the equilibrium path

pessimistic. That is, $p_t(x)$ is decreasing in t for fixed x. Letting x_t denote the agent indifferent about continuing experimentation at time t, so that $V(p_t(x_t)) = \frac{r}{\gamma}$, this implies that x_t must be increasing in t. Thus there is a shrinking mass of agents in favor of the risky policy (the agents shaded in blue in Figure 1) and a growing mass of agents against it (the agents shaded in red and green). For high t, almost all agents agree that experimentation should be stopped.

However, growing pessimism induces members to leave. Hence the marginal member becomes more extreme, and so does the median member. If $m_t \geq x_t$ for all t, that is, if the prior of the median is always higher than the prior of the indifferent agent, then the agents in favor of the risky policy always retain a majority within the organization, due to most of their opposition forfeiting their voting rights.

Figure 2 shows the same result in the space of posterior beliefs. The accumulation of negative information puts downward pressure on $p_t(m_t)$ as t grows, but selection forces prevent it from converging to zero. Instead, $p_t(m_t)$ converges to a belief strictly between 0 and 1, which is above the critical value $p_t(x_t)$ in this example. Hence the median voter always remains optimistic enough to continue experimenting.

To establish whether this equilibrium entails over-experimentation, we need a definition of over-experimentation in a setting with heterogeneous priors. We will use the following notion. Consider an alternative model in which an agent with initial belief x controls the policy at all times. It is well-known that whenever $0 < x < 1$, the agent would experiment until some finite time depending on x. We say that an equilibrium of our model features over-experimentation from x's point of view if experimentation continues for longer than that. By this definition, when the condition in Proposition 1 is satisfied, there is over-experimentation from the point of view of *all* agents except those with prior belief exactly equal to 1.

The level of experimentation in equilibrium is determined by the interaction of two opposing forces, in addition to the usual incentives present in the canonical single-agent bandit problem. When the pivotal agent decides whether to stop experimenting at time t, she takes into account the difference in the expected flow payoffs generated by the safe policy and the risky one, as well as the option value of experimenting further. However, because the identity of the median voter changes over time, the pivotal agent knows that if she chooses to continue experimenting, the organization will stop at a time chosen by some other agent, which she likely considers suboptimal. This force encourages her to stop experimentation while the decision is still in her hands, leading to under-experimentation. It is similar to the force behind the under-experimentation result in Strulovici (2010) in that, in both cases, agents prefer a sub-optimal amount of experimentation because they expect a loss of control over future decisions if they allow experimentation to continue. It is also closely related to the concerns about slippery slopes faced by agents in the clubs literature (see, for example, Bai and Lagunoff (2011) and Acemoglu et. al. (2015)).

The second force stems from the endogeneity of the median voter's position in the distribution. As discussed above, the more pessimistic a fixed observer becomes about the risky policy, the more extreme the median voter is. This effect is so strong that, as time passes, the posterior belief of the median after observing no successes does not converge to zero, and the median voter may choose to continue experimenting when no successes have been observed for an arbitrarily long time.

Next, we explain why the equilibrium with perpetual experimentation is unique. The key here is that if an agent prefers to experiment forever rather than not at all, then she also prefers to experiment for any finite amount of time T rather than not at all. Thus if the median conjectured that the organization will experiment for some finite time T instead of forever, the median still would not want to stop experimentation.

The central result that we use in the proof here is that the value function $W_T(x)$ of an agent with prior x who expects experimentation to continue for time T is single-peaked in T. That is, there is a time T^* such that if the agent had control over the policy at all times, the agent would experiment for time T^*, and the farther away the actual length of experimentation is from T^*, the less happy the agent is. Since $W_0(x) = \frac{r}{\gamma}$ and $W_T(x)$ is increasing in T before the peak, the agent prefers to experiment for time T before the peak rather than not at all. Moreover, since $\lim_{T\to\infty} W_T(x) = V(x)$ and $V(x) > \frac{r}{\gamma}$ by our hypothesis, the fact that $W_T(x)$ is decreasing in T after the peak implies that the agent also prefers to experiment for time T after the peak rather than not at all. This establishes our result.

For certain families of belief densities we can obtain a closed-form expression for the condition in Proposition 1. For instance, if the density f is non-decreasing, then the value function V in Proposition 1 satisfies the following:[20]

$$\gamma \inf_{t\geq 0} V(p_t(m_t)) = \gamma V\left(\frac{2s}{b+s}\right) = \frac{2bs}{b+s} + \left(\frac{1}{2}\right)^{\frac{\gamma}{b}} \frac{s(b-s)}{b+s}\frac{b}{\gamma+b} \tag{1}$$

The value function of the pivotal agent has a simple interpretation. Under a non-decreasing density f, as the organization experiments unsuccessfully, the posterior belief of the median converges from above to $\frac{2s}{b+s}$. The first term on the right-hand side, then, represents the agent's expected flow payoff from experimentation: it is the product of the probability $\frac{2s}{b+s}$ that the asymptotic median assigns to the risky policy being good, and the expected flow payoff b from the good risky policy. The second term is the option value derived from the agent's ability to leave the organization when she becomes pessimistic enough, and to return if there is a success.

[20]See Proposition 9 in the Appendix for more details.

Our next major result concerns the comparative statics of our model. We show that perpetual experimentation is more likely when the payoffs from the risky policy and from the outside option are high, the payoff from the organization's safe policy is low, the agents are patient, and there are many optimists.

Proposition 2. *If there is an equilibrium with perpetual experimentation under parameters (b, r, s, γ, f), then the same holds for any set of parameters (b, r', s', γ', f') such that $r' \leq r$, $s' \geq s$, $\gamma' \leq \gamma$ and f' MLRP-dominates f.*[21]

The intuition behind this result is as follows. An increase in the payoff r from the safe policy makes the safe policy more attractive and has no effect on the expected payoff of perpetual experimentation. An increase in patience is equivalent to an increase in the rate of learning from experimentation, and a higher learning rate allows the agents to make better entry-exit decisions.

An increase in the number of optimists leaves the value function and the marginal member unchanged but results in a more optimistic median who is more likely to support experimentation.[22] An increase in s has two effects that favor experimentation. First, it increases the payoff from perpetual experimentation because the agent expects to quit the organization and collect the outside option payoff with some probability. Second, it induces agents to quit, which leaves the organization with a more radical median voter.

Finally, an increase in b has three effects. On the one hand, it makes experimentation more attractive, both by directly increasing the payoff from the good risky policy and by increasing the learning rate. At the same time, it induces more agents to stay in the organization, which makes the median more pessimistic. Hence the overall effect is ambiguous. However, for a well-behaved family of densities, the first two effects dominate, so that perpetual experimentation is also more likely for high b.[23]

If there does not exist an equilibrium with perpetual experimentation, there

[21] We say that g MLRP-dominates f if $x \mapsto \frac{g(x)}{f(x)}$ is non-decreasing for $x \in [0, 1)$.

[22] For example, if f is uniform, m_t is the midpoint between y_t and 1, while if f is increasing, then m_t is closer to 1 than y_t.

[23] Specifically, for each $\alpha > 0$, let $f_\alpha(x)$ denote a density in the power-law family given by $f_\alpha(x) = (\alpha + 1)(1 - x)^\alpha$. If $f = f_\alpha$ for some α, then $\inf_t V(p_t(m_t))$ is increasing in b. See Proposition 9 in the Appendix for details.

may be multiple equilibria featuring different levels of experimentation, supported by different off-path behavior.[24] To state the results, we let $\hat{T}$ denote the time such that the initial median is indifferent between switching to the safe policy and continuing to experiment if she anticipates that, should she continue, the organization will stop experimentation at time $\hat{T}$.

Remark. *Any equilibrium stopping time must lie in* $\left[0, \hat{T}\right]$.

Note that experimentation must stop by time $\hat{T}$, as otherwise the initial median would switch to the safe policy immediately. Proposition 3 speaks to the extent of experimentation when there does not exist an equilibrium with perpetual experimentation.

Proposition 3.

1. *There are parameters under which*[25] *all* $t \in \left[0, \hat{T}\right]$ *are equilibrium stopping times.*
2. *If for all* $t \in [0, \bar{t}]$, m_t *prefers to experiment forever rather than not at all, then in any equilibrium the organization experiments up to at least* $\bar{t}$.

It can be shown that, because W_T is single-peaked in T, the initial median's ideal stopping time lies between 0 and $\hat{T}$. Then the first part of the Proposition implies that both over and under-experimentation are possible depending on which equilibrium is played. Under-experimentation obtains if an early median voter expects that, should she continue experimenting, the next stopping time will be too far in the future.

The second part of the Proposition obtains for the following reason. If all pivotal agents up to some time $\bar{t}$ prefer experimenting forever to not at all, then, because W_T is single-peaked, these agents will never stop experimenting. Therefore, the equilibrium stopping time must be at least $\bar{t}$. This means that our result of perpetual experimentation survives in an approximate form even when, for instance,

[24] More generally, a pure strategy equilibrium can be described by a sequence $t_0 < t_1 < t_2 < \dots$ of stopping times as follows. For any $t \in (t_{n-1}, t_n]$, if the risky policy was used in the period $[0, t]$ and no successes were observed, the organization continues using it until time t_n. If the risky policy has not succeeded by t_n, the organization switches to the safe policy at t_n.

[25] The condition on the parameters roughly amounts to requiring that $p_t(m_t)$ does not decrease too steeply in t. For example, $p_t(m_t)$ being constant in t would be sufficient. See the Appendix for details.

the support of the prior distribution is truncated away from 1.[26]

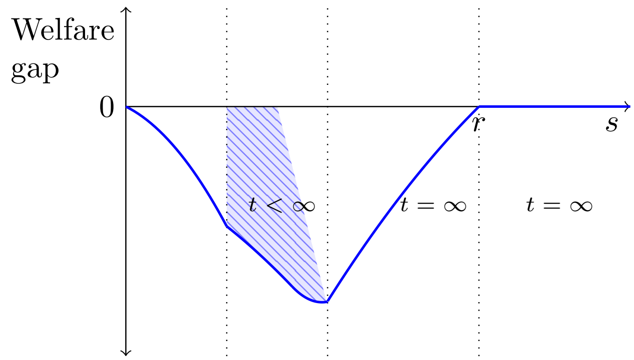

Figure 3: Welfare gap between the equilibrium and the socially optimal stopping time

Figure 3 illustrates the welfare effect of varying the quality of the outside option. The blue curve is the difference between the initial median's equilibrium utility and her utility from stopping experimentation at her most preferred time. The shaded blue region represents the range of welfare outcomes that obtain when multiple equilibria exist. The reason this region occupies only part of the space between the two dotted lines is that, as s increases, some equilibria disappear, so the range of obtainable welfare outcomes shrinks.

In this example, the welfare gap in the worst equilibrium, compared to the initial median's optimal stopping time, is U-shaped. The reason is as follows. When $s \geq r$, the organization experiments forever. Here perpetual experimentation is optimal from the point of view of all agents. This is because, since the organization's safe policy is no better than the outside option, agents who lose faith in the risky policy prefer to switch to the outside option rather than use the organization's safe policy.

For $s < r$ relatively close to r, the organization still experiments forever, but this is now inefficient from the point of view of all agents, with the size of the welfare loss increasing as the gap between r and s grows. The inefficiency comes from the fact that pessimistic agents are denied access to the organization's safe policy.

For even lower values of s, perpetual experimentation is impossible because too few agents leave the organization. Here a range of outcomes is possible. These

[26]Formally, suppose that for some density f there is perpetual experimentation, and let f_y be f truncated at y, that is, $f_y(x) \propto f(x)\mathbb{1}_{x\leq y}$. Denote by $\bar{t}_y$ the minimal equilibrium stopping time as a function of y. Then $\bar{t}_y \to \infty$ as $y \to 1$.

outcomes may include the initial median's preferred stopping time. Finally, when s is very close to 0, we can show that the equilibrium is again unique. In this equilibrium, few agents ever leave, but the median still becomes more optimistic over time as bad news arrive. This means that, generally, the stopping decision is still made by an agent more optimistic than the initial median. Moreover, the gap between the initial median and the median who stops is growing in the number of people who leave, which itself is increasing in s.

6 Other Learning Processes

The baseline model presented above has two salient features. First, experimentation has a low probability of generating a success, which increases agents' posterior beliefs substantially, and a high probability of generating no successes, which lowers their posteriors slightly. In other words, the baseline model is a model of good news. Second, because the risky policy can only succeed when it is good, good news are perfectly informative.

In this Section, we relax these assumptions, presenting variants of the model which allow for imperfectly informative good news and for bad news. In the first case, we show that our finding of over-experimentation is robust to imperfectly informative news. We also show that the organization may respond perversely to information, becoming more reluctant to experiment after a success. In the case of perfectly informative bad news, in contrast, there is typically under-experimentation.

6.1 A Model of Imperfectly Informative (Good) News

We first treat the case of imperfectly informative news, which allows for much richer dynamics than the baseline model: agents' beliefs, rather than decreasing monotonically or else jumping permanently to 1, can change in both directions as successes and failures arrive. For brevity, we consider the case of good news, but similar results can be obtained for imperfectly informative bad news.

The model is the same as in Section 4 except for the payoffs generated by the risky policy. If the risky policy is good, it generates successes according to a Poisson process with rate b. If it is bad, it generates successes according to a Poisson process with rate c. We now assume that $b > r > s > c > 0$.

As before, the effect of past information on the agents' beliefs can be aggregated into a one-dimensional sufficient statistic. Suppose the risky policy has been used for a length of time t and k successes have occurred during that time. Define

$$L(k,t) = \left(\frac{c}{b}\right)^k e^{(b-c)t}$$

Then the posterior of an agent with prior x at time t after observing the organization use the risky policy for a length of time t and achieve k successes is

$$\frac{x}{x + (1-x)L(k,t)}$$

We again suppress the dependence of $L(k,t)$ on k and t and use L to denote our sufficient statistic. Note that high L indicates bad news about the risky policy.

We let $V_x(L)$ denote the value function of an agent with prior x given that the state is L and the organization experiments forever in the continuation. In addition, we denote x's ex ante utility in the same situation by $V(x) = V_x(1)$. The next Proposition shows that, as in Section 5, experimentation can continue forever regardless of how badly the risky policy performs.

Proposition 4. *If $V_{m(L)}(L) > \frac{r}{\gamma}$ for all L, then there is a unique equilibrium. In this equilibrium, the organization experiments forever.*

We can show that there exist parameters such that the inequality $V_{m(L)}(L) > \frac{r}{\gamma}$ is satisfied. For instance, if f is non-decreasing, then the posterior of the median converges to $\frac{2(s-c)}{(b-c)+(s-c)}$, so $\inf_t V(p_t(m_t)) = V\left(\frac{2(s-c)}{(b-c)+(s-c)}\right)$. Then it is enough to verify that $V\left(\frac{2(s-c)}{(b-c)+(s-c)}\right) \geq \frac{r}{\gamma}$.[27]

The following result illustrates the novel outcomes that can arise under imperfectly informative news.

Proposition 5. *There exist parameters such that there is an equilibrium in which the organization experiments more when the risky policy is bad than when it is good.*

[27]However, in this case it is not possible to give an exact expression for V, owing to the complicated behavior of L over time.

The intuition for the result in Proposition 5 is as follows. We first show that, for an appropriately chosen density f, an equilibrium of the following form exists: whenever $L = L^*$, the organization stops experimenting with probability ϵ, and at all other times the organization continues experimenting for sure. For this to work, f must be such that the median is most pessimistic when $L = L^*$.[28] Moreover, r must be such that $V_{m(L^*)}(L^*) = \frac{r}{\gamma}$, so that the median is indifferent about stopping experimentation at L^*, while other agents prefer to continue experimenting when they are pivotal.

The striking feature of this equilibrium is that stopping only happens for an intermediate value of L. In particular, if L^* is smaller than the initial L, the only way experimentation will stop is *if it succeeds* enough times for L to decrease all the way to L^*, which is more likely to happen when the risky policy is good.

6.2 A Model of Bad News

In this section we consider the same model as in Section 4, except that the risky policy now generates different flow payoffs. In particular, if the risky policy is good, it generates a guaranteed flow payoff b. If it is bad, it generates a guaranteed flow payoff b but also experiences failures, which arrive according to a Poisson process with rate b. Each failure lowers the payoffs of all members by 1. Thus, as in the baseline model, the expected flow payoff from using the risky policy is b when it is good and 0 when it is bad. The learning process, however, is different.

The dynamics of organizations under bad news differ substantially from those in the baseline model. As is usual in models of bad news, as long as no failures are observed, all agents become more optimistic about the risky technology, so the organization expands over time instead of shrinking. This gradual expansion continues either forever or until some time T unless a failure occurs, in which case the organization switches to the safe technology and all agents previously outside the organization become members. Interestingly, the switch to the safe technology must happen upon observing a failure but may happen even if no failures are observed.

[28]This occurs, for instance, if f is very high in a small neighborhood of $y(L^*)$. Then, when $L > L^*$, all the pessimists to the left of $y(L^*)$ leave, so that $m(L)$ is more optimistic, while when $L < L^*$, pessimists become members, yielding a lower $m(L)$. This construction is not possible for some densities of prior beliefs. In particular, if f is uniform or follows a power law distribution, then $p(L, m(L))$ is decreasing in L (Lemma 17).

As before, m_t is the median member at time t provided that the risky policy has been used up to time t and no failures have been observed. $p_t(m_t)$ is the median's posterior belief at time t, and $V(p_t(m_t))$ is her continuation value when experimentation is expected to continue forever unless there is a failure. Let us say that a pivotal agent m_t is an *optimist* if $V(p_t(m_t)) > \frac{r}{\gamma}$ and a *pessimist* if $V(p_t(m_t)) < \frac{r}{\gamma}$. Let us use $\underline{t}$ to denote the earliest time when a pessimist is pivotal. Proposition 6 provides a characterization of the equilibrium in this variant of the model.

Proposition 6.

1. *If all pivotal agents are optimists, then there is a unique equilibrium. In it, the organization experiments forever.*

2. *If some pivotal agents are pessimists, then in any equilibrium the organization stops experimenting at a finite time $T < \underline{t}$.*[29]

Proposition 6 shows that perpetual experimentation is the unique equilibrium outcome if all pivotal agents prefer it to the safe policy. If, however, some pivotal agents are pessimistic enough to stop experimentation, the organization will always switch to the safe policy even *before* any of these pessimists become pivotal. Note that, even when perpetual experimentation arises in the bad news setting, it does not constitute over-experimentation, as it is possible only when all agents agree that perpetual experimentation is optimal.

To understand these results, it is instructive to consider the associated single-agent bandit problem. In a model of bad news, the agent switches to the safe policy permanently upon observing a failure, and becomes more optimistic over time if she pursues the risky policy and observes no failures. The more optimistic the agent becomes, the more she wants to continue using the risky policy. Hence the agent will either want to experiment forever or not at all.

Two implications follow from this observation. First, pessimists always switch to the safe policy when they are pivotal: by assumption, they prefer no experimentation to perpetual experimentation, and thus also to any other continuation. Second, optimists have stronger incentives to experiment if they expect experimentation to continue in the future: only then can they collect the option value of learning about

[29] This is true so long as $\underline{t} > 0$. If $\underline{t} = 0$, then $T = 0$.

the policy. For them, current and future experimentation are strategic complements.

These implications lead to the organization stopping experimentation strictly before $\underline{t}$, as follows. Optimists are willing to experiment if they expect perpetual experimentation in the continuation. However, agents who are pivotal shortly before $\underline{t}$ know that any experimentation they attempt will be short-lived. Thus they will choose to stop experimenting even if they are optimists. In turn, their expected behavior may induce even earlier pivotal agents to switch to the safe policy as well.

To summarize, in a bad news setting, over-experimentation is never possible from the point of view of any pivotal agent, while under-experimentation is possible, and always obtains when experimentation does not continue forever. These results stand in stark contrast to those of our previous models. The results depend on a special feature of the perfectly informative bad news learning process: bad news create common knowledge that the risky policy is bad. Because of this, there is no room for organizational capture by optimists who would disagree with the majority.

7 Other Extensions

7.1 General Voting Rules

We assume throughout the paper that the median member of the organization is pivotal. This assumption is not essential to our analysis: our results can be extended to other voting rules under which the agent at the q-th percentile is pivotal.

It is instructive to consider how the results change as we vary q. Letting q_t represent the pivotal agent at time t, it is clear that q_t and $p_t(q_t)$ are increasing in q for all t. To illustrate further, assume that f is uniform. Then, as $t \to \infty$, the posterior belief of the pivotal agent converges to $\frac{s}{qs+(1-q)b}$ rather than $\frac{2s}{b+s}$. It follows that more stringent supermajority requirements are functionally equivalent to more optimistic leadership of the organization, and make it easier to sustain an equilibrium with excessive experimentation.

7.2 Size-Dependent Payoffs

In some settings the payoffs that an organization generates may depend on its size. In this section we discuss how different operationalizations of this assumption affect our results. We show that our main result is robust to this extension, and discuss how different kinds of size-dependent payoffs may exacerbate or prevent over-experimentation.

We consider three types of size-dependent payoffs. For the first two, we suppose that when the set of members of the organization has measure μ, the safe policy yields a flow payoff $rg(\mu)$, the good risky policy yields instantaneous payoffs of size $g(\mu)$ generated at rate b, and the bad risky policy yields zero payoffs. We assume that $g(1) = 1$, so that b, r and 0 are the expected flow payoffs from the good risky policy, the safe policy and the bad risky policy respectively when all agents are in the organization. For the first type of payoffs we consider, $g(\mu)$ is increasing in μ, so there are economies of scale. For the second type, $g(\mu)$ is decreasing in μ, so there is a congestion effect.

In general, the effect of size-dependent payoffs on the level of experimentation is ambiguous because of two countervailing effects. On the one hand, when there is a congestion effect, as the organization shrinks, higher flow payoffs increase the benefits from experimentation, which makes experimentation more attractive.[30] We call this the *payoff effect*. On the other hand, because increasing flow payoffs provide incentives for agents to stay in the organization, the organization shrinks at a lower speed, which causes the median voter in control of the organization to be more pessimistic about the risky policy. We call this the *control effect*. When there are economies of scale, these effects are reversed.

When there are economies of scale, the set of members may not be uniquely determined as a function of the state at time t. This is because the more members there are, the higher payoffs are, so the membership stage may have multiple equilibria. We will assume, for simplicity, that the set of members is uniquely determined.[31] It is sufficient to assume that g does not increase too fast.

[30]While the safe policy could also yield high payoffs when the organization is small, all agents will enter as soon as the safe policy is implemented, so these high payoffs can never be captured.

[31]Formally, we require that the equation $\frac{y_t}{y_t+(1-y_t)e^{bt}} = \frac{s}{g(1-F(y_t))b}$ has a unique fixed point for all $t \geq 0$.

The following Proposition presents our first result.

Proposition 7. *Suppose that* $f = f_\alpha$.[32] *Let* $\overline{g} = \lim_{\mu \to 0} g(\mu)$, *and let* $V_{g,t}(p_t(m_t))$ *denote the utility of the pivotal agent at time* t *if she expects experimentation to continue forever. If*

$$\lambda^{\frac{\gamma}{b}} s \frac{b}{\gamma + b} + \frac{s}{\lambda} \frac{\gamma}{\gamma + b} > b \frac{b}{\gamma + b} + s \frac{\gamma}{\gamma + b}$$

then $\lim_{t \to \infty} V_{g,t}(p_t(m_t))$ *is strictly increasing in* $\overline{g}$ *for all* $\overline{g} \in [s, \infty)$. *In this case, perpetual experimentation obtains for a greater set of parameter values with a congestion effect and for a smaller set of parameter values with economies of scale, relative to the baseline model.*

Conversely, if the reverse inequality holds strictly, then $\lim_{t \to \infty} V_{g,t}(p_t(m_t))$ *is strictly decreasing in* $\overline{g}$ *for all* $\overline{g} \in [s, \infty)$.

The intuition for the Proposition is as follows. By the same argument as in the baseline model, Proposition 1 holds: a sufficient condition to obtain experimentation forever is that $V_{g,t}(p_t(m_t)) \geq \frac{r}{\gamma}$ for all t. While it is difficult to calculate $V_{g,t}(p_t(m_t))$ explicitly for all t, calculating its limit as $t \to \infty$ is tractable and often allows us to determine whether the needed condition holds for all t. We show that the limit depends only on $\overline{g}$ rather than the entire function g. Moreover, it is a hyperbola in $\overline{g}$, so it is either increasing or decreasing in $\overline{g}$ everywhere. In the first case, size-dependent payoffs affect the equilibrium mainly through the *payoff effect*, so experimentation is more attractive with a congestion effect and less so with economies of scale. In the second case, the *control effect* dominates, and the comparative statics are reversed. These statements are precise as $t \to \infty$ (that is, conditional on the risky policy having been used for a long time). We can show that when congestion effects make experimentation more likely in the limit, they do so for all t.[33]

The inequality in the Proposition determines which case we are in. Because $b > \lambda^{\frac{\gamma}{b}} s$ and $\frac{s}{\lambda} > s$, if b is large enough relative to γ, then over-experimentation is easier to obtain with economies of scale than in the baseline model, and easier to obtain in the baseline model than with a congestion effect. The opposite happens

[32] Recall that $f_\alpha(x)$ is a density with support $[0, 1]$ such that $f_\alpha(x) = (\alpha+1)(1-x)^\alpha$ for $x \in [0, 1]$.

[33] It can be shown that when congestion effects make experimentation less likely in the limit, they may not do so for all t.

if γ is large relative to b. The reason is that, under economies of scale, the pivotal decision-maker is very optimistic about the risky policy but expects to receive a low payoff from the first success. If $\frac{b}{\gamma}$ is large, so that successes arrive at a high rate or the agent is very patient, the first success is expected to be one of many, while if $\frac{b}{\gamma}$ is small, further successes are expected to be heavily discounted. Conversely, with a congestion effect, for large t the pivotal decision-maker is almost certain that the risky policy is bad but believes that, with a low probability, it will net a very large payoff before she leaves.

The third way to operationalize size-dependent payoffs that we consider deals with changes to the learning rate rather than to flow payoffs. Here we suppose that when the organization is of size μ, the good risky policy generates successes at a rate $b\mu$. Each success pays a total of 1, which is split evenly among members, so that each member gets $\frac{1}{\mu}$. All other payoffs are the same as in the baseline model. An example that fits this setting is a group of researchers trying to find a breakthrough. If there are fewer researchers, breakthroughs are just as valuable but happen less often. When f is uniform, and using V to denote the continuation utility under perpetual experimentation, we have

$$\gamma \inf V_t(p_t(m_t)) = \gamma \lim_{t\to\infty} V_t\left(\frac{2s}{b+s}\right) = \frac{2bs}{b+s}$$

In other words, the asymptotic median's expected payoff from experimentation comes only from the flow payoff of the risky policy; the option value of experimentation vanishes as the learning rate converges to zero. It follows that perpetual experimentation is less likely to obtain here than in the baseline model, but is still the unique outcome if $\frac{2bs}{b+s} > \frac{r}{\gamma}$. Note that, in this case, as agents who join the organization increase the learning rate, they confer a positive externality on outsiders which is not internalized. Hence, there is free-riding as in Keller et al. (2005). It is simultaneously possible that too few agents partake in experimentation—given that the organization's policy is risky—and that the risky policy is used for too long.

7.3 Organizations of Fixed Size

For the sake of simplicity and clarity, our main model assumes that the organization allows agents to enter and exit freely and adjusts in size to accommodate them.

While free exit is a reasonable assumption in all of our applications, the assumptions of free entry and flexible size are often violated: in the short run, organizations may need to maintain a stable size to sustain their operations. In this Section, we discuss a variant of our model incorporating these concerns.

Assume now that agents differ in two dimensions: their prior belief $x \in [0,1]$ and their ability $z \in [\underline{z}, \overline{z}]$. Suppose that the density of agents at each pair (x, z) is of the form $f(x)h(z)$, where f is a probability density function and h is a degenerate density such that, for any $z > \underline{z}$, $\int_{\underline{z}}^{z} h(\tilde{z})d\tilde{z} = \infty$ but $\int_{z}^{\overline{z}} h(\tilde{z})d\tilde{z} < \infty$. In other words, prior beliefs and ability are independently distributed, and for each belief x there is a deep pool of agents, if candidates of low enough ability are considered.

Assume that the organization must maintain a fixed size μ; that it observes only ability, and not beliefs, from its hiring process; and that it benefits from hiring high-ability agents. Suppose that agents are compensated equally for their ability inside or outside the organization (that is, their propensity to be members is independent of ability). Then, in equilibrium, at time t only candidates with prior belief $x \geq y_t$ are willing to work at the organization. Here y_t is given by $p_t(y_t) = \frac{s}{b}$, as in the baseline model. The organization hires all candidates of ability at least z_t, where z_t is chosen so that $(1 - F(y_t))(1 - H(z_t)) = \mu$.

Since x and z are independently distributed, the median belief within the organization at time t is still $p_t(m_t)$. From this fact we can derive an analog of Proposition 1 and show that perpetual experimentation can also obtain in this case.[34] Moreover, over-experimentation becomes even more likely if prior beliefs and ability are positively correlated, or if the organization is able to observe ability to some extent and prefers optimistic agents.

7.4 Common Values

Although our model features agents with private values, our results can be extended to a model with common values, which is more appropriate for some of our

[34] In fact, perpetual experimentation is easier to obtain in this case. Letting z_0 be the ability threshold when everyone wants to be a member, that is, when there has been a success or the safe policy is being used, we can show that the incentives to advocate for experimentation are the same as in the baseline model for agents (x, z) with $z \geq z_0$. However, agents (x, z) with $z < z_0$ have a dominant strategy to advocate for experimentation, because they know that a switch to the safe policy would see them fired immediately.

applications, such as environmental organizations or civil rights activism.

We discuss this extension in the context of our example of civil rights activism. Instead of each agent x generating some private income flow, she now makes a flow *contribution* to a rate of change in the relevant laws which can be attributed to the activism of agent x. The mapping from membership and policy decisions to the outcomes is the same as in Section 4, but now agents care only about the overall rate of changes in the law, and not about their own contribution.

Formally, we let $U_x(\sigma_y, \sigma)$ denote the private utility of agent x when she plays the equilibrium strategy of agent y and the equilibrium path is dictated by the strategy profile σ. Then in the private values case x's equilibrium utility is $U_x(\sigma_x, \sigma)$, while in the common values case it is $\int_0^1 U_x(\sigma_y, \sigma) f(y) dy$. Note that, even though all agents share the objective of maximizing the aggregate rate of legal change, their utility functions still differ due to differences in prior beliefs. However, it is still optimal for them to make membership decisions that maximize their flow contributions at each point in time, just as in Section 4.[35]

Let us conjecture a strategy profile in which the organization experiments forever, and let $\tilde{V}_t(x)$ denote the continuation utility at time t of an agent who has posterior belief x at time t under this strategy profile.[36] Then Proposition 1 holds with the same proof, replacing $V(p_t(m_t))$ in the original statement with $\tilde{V}_t(p_t(m_t))$. Moreover, the following lower bound for the value function holds:

Proposition 8. *For any $x \in [0,1]$ and any $t \geq 0$,*

$$V(x) \geq \tilde{V}_t(x) \geq \min \left\{ x\frac{b}{\gamma}, x\frac{b}{\gamma} + (1-x)\frac{s}{\gamma} - x\frac{b-s}{\gamma+b} \right\}$$

Proposition 8 can be used to obtain a sufficient condition for perpetual experimentation. For instance, when the density of the prior beliefs f is uniform,

[35] Agents are now indifferent about their membership decisions: the membership status of a set of agents of measure zero has no impact on anyone's payoffs. However, it is natural to assume that each agent joins when doing so would be optimal if her behavior had a positive weight in her own utility function. For instance, this is the case in a model with a finite number of agents.

[36] t matters in this case, in contrast to the model in Section 4, because the membership strategies of other agents, which depend on t rather than x, enter the agent's utility function.

experimentation continues forever as long as

$$\min\left\{\frac{2bs}{b+s}, \frac{2bs}{b+s} + \frac{(b-s)s}{b+s}\left(1 - 2\frac{\gamma}{\gamma+b}\right)\right\} \geq r$$

In other words, over-experimentation can still occur in equilibrium for reasonable parameter values.

The common values setting differs from the baseline model in two important ways. First, the fact that $\tilde{V}_t(x) \leq V(x)$ means that agents' payoffs from experimentation are always weakly lower in the common values case than in the private values case. As a result, over-experimentation occurs for a smaller set of parameter values in the common values case. The reason is that, under common values, an agent considers the entry and exit decisions of other agents suboptimal, and her payoff is affected by these decisions as long as experimentation continues. In contrast, in the private values case, agents' payoffs depend only on their own entry and exit decisions, which are chosen optimally given their beliefs.

Second, in the private values case, experimentation can continue forever even if agents are impatient, as long as the density f does not decrease too quickly near 1 and other parameters are chosen appropriately (for example, s is close to r). This occurs because the pivotal agent is optimistic enough that the expected flow payoff from experimentation is higher than r, even without taking the option value into account. In contrast, in the common values case, the expected flow payoff from experimentation goes to s as $t \to \infty$ if there are no successes, no matter how optimistic the agent is. Indeed, here agents care about the contributions of all players, and they understand that for large t most players will become outsiders and generate s, regardless of the quality of the policy. Thus perpetual experimentation is only possible if agents are patient enough.

References

Acemoglu, Daron, Georgy Egorov, and Kostantin Sonin, "Coalition Formation in Non-Democracies," *Review of Economic Studies*, 2008, *75* (4), 987–1009.

_ , _ , and _ , "Dynamics and Stability of Constitutions, Coalitions and Clubs," *American Economic Review*, 2012, *102* (4), 1446–1476.

_ , _ , and _ , "Political Economy in a Changing World," *Journal of Political Economy*, 2015, *123* (5), 1038–1086.

Antioco, John, "How I Did It: Blockbuster's Former CEO on Sparring with an Activist Shareholder," , Harvard Business Review April 2011. `https://hbr.org/2011/04/how-i-did-it-blockbusters-former-ceo-on-sparring-with-an-activist-shareholder`, Last accessed on 2019-12-10.

Bai, Jinhui and Roger Lagunoff, "On the Faustian Dynamics of Policy and Political Power," *Review of Economic Studies*, 2011, *78* (1), 17–48.

Bell, David S, "The French Communist Party: from revolution to reform," in Jocelyn Evans, ed., *The French party system*, Manchester University Press, 2003, chapter 2.

Bréchon, Pierre, *Les partis politiques français*, La Documentation française, 2011.

Carreyrou, John, *Bad Blood: Secrets and Lies in a Silicon Valley Startup*, Penguin Random House, 2018.

Colvin, Geoff and Phil Wahba, "Sears' Seven Decades of Self-Destruction," May 2019. `https://fortune.com/longform/sears-self-destruction/`, Last accessed on 2019-12-10.

Damiani, Marco and Marino De Luca, "From the Communist Party to the Front de gauche. The French radical left from 1989 to 2014," *Communist and Post-Communist Studies*, 2016, *49* (4), 313–321.

Durrett, Rick, *Probability: Theory and Examples*, 4 ed., Cambridge University Press, 2010.

Gavetti, Giovanni M., Rebecca Henderson, and Simona Giorgi, "Kodak and The Digital Revolution (A)," Harvard Business School Case 705-448, Harvard Business School Publishing 2004.

Gieczewski, Germán, "Policy Persistence and Drift in Organizations," 2019.

Keller, Godfrey and Sven Rady, "Strategic experimentation with Poisson bandits," *Theoretical Economics*, 2010, *5* (2), 275–311.

_ and _ , "Breakdowns," *Theoretical Economics*, 2015, *10* (1), 175–202.

_ , _ , and Martin Cripps, "Strategic Experimentation with Exponential Bandits," *Econometrica*, 2005, *73* (1), 39–68.

Roberts, Kevin, "Dynamic voting in clubs," *Research in Economics*, 2015, *69* (3), 320–335.

Strulovici, Bruno, "Learning While Voting: Determinants of Collective Experimentation," *Econometrica*, 2010, *78* (3), 933–971.

Appendix (For Online Publication)

Definition of Equilibrium

We let π_{t^-} and π_{t^+} denote the left and right limits of the policy path at time t respectively, whenever the limits are well-defined. We require that π_t, the current policy at time t, should be optimal from the point of view of the decision-maker who is pivotal given the incumbent policy π_{t^-}. Similarly, π_{t^+} should be optimal from the point of view of the decision-maker who is pivotal given π_t. That is, for the policy to change from π to π' along the path of play, the decision-maker induced by π must be in favor of the change.

We define a membership function β so that $\beta(x, L, \pi) = 1$ if agent x chooses to be a member of the organization given information L and policy π, and $\beta(x, L, \pi) = 0$ otherwise. We define a policy function α so that $\alpha(L, \pi)$ is the set of policies the median voter, $m(L, \pi)$, considers optimal.[37] Our notion of strategy profile summarizes the above requirements:

Definition 2. A strategy profile is given by a membership function $\beta : [0,1] \times \mathbb{R}_+ \times \{0,1\} \to \{0,1\}$, a policy function $\alpha : \mathbb{R}_+ \times \{0,1\} \to \{\{0\},\{1\},\{0,1\}\}$, and a stochastic path of play consisting of information and policy paths $(L_t, \pi_t)_t$ satisfying the following:

(a) Conditional on the policy type θ, $(L_t, \pi_t)_{t\geq 0}$ is a progressively measurable Markov process with paths that have left and right limits at every $t \geq 0$ satisfying $(L_0, \pi_0) = (1, 1)$.

(b) Letting $\left(\tilde{k}_\tau\right)_\tau$ denote a Poisson process with rate b or 0 if $\theta = G$ or B respectively, letting $\left(\tilde{L}_\tau\right)_\tau$ be given by $\tilde{L}_\tau = L\left(\tilde{k}_\tau, \tau\right)$, and letting $n(t) = \int_0^t \pi_s ds$ denote the amount of experimentation up to time t, we have $L_t = \tilde{L}_{n(t)}$.

(c) $\pi_t \in \alpha(L_t, \pi_{t^-})$ for all $t \geq 0$.

(d) $\pi_{t^+} \in \alpha(L_t, \pi_t)$ for all $t \geq 0$.

[37] $\alpha(L, \pi)$ can take the values $\{0\}$, $\{1\}$ and $\{0,1\}$. Defining $\alpha(L, \pi)$ in this way is convenient because some paths of play cannot be easily described in terms of the instantaneous switching probabilities of individual agents.

Before we provide a definition of equilibrium, a short digression on continuation utilities after deviations is required. We define $V_x(L,\pi)$ as the continuation utility of an agent with prior belief x given information L and incumbent policy π. In other words, $V_x(L,\pi)$ is the utility agent x expects to get starting at time t_0 when the state follows the process $(L_t,\pi_t)_{t\geq t_0}$ given that $(L_{t_0},\pi_{t_0})=(L,\pi)$. In state (L,π), the median $m(L,\pi)$ can choose between the continuations starting in states $(L,1)$ and $(L,0)$. In the well-behaved case where these continuations are different, it is natural to define the set of the optimal policies $\alpha(L,\pi)$ as the set of policies π' that maximize the median's continuation payoff $V_{m(L,\pi)}(L,\pi')$.

However, if the continuations are identical,[38] applying this definition would imply that $\alpha(L,\pi)=\{0,1\}$ because the choice made by $m(L,\pi)$ has no impact on the continuation. This allows for unattractive equilibria in which weakly dominated policies may be chosen: even under common knowledge that the risky policy is good, there is an equilibrium in which all decision-makers choose the safe policy because any deviation to the risky policy would be reversed immediately.

To eliminate these equilibria, our definition considers short-lived deviations optimal if they would be profitable when extended for a short amount of time. To formalize this, we define $\overline{V}_x(L,\pi,\epsilon)$ as x's continuation utility under the following assumptions: the state is (L,π) at time t_0, the policy π is *locked in* for a length of time $\epsilon>0$ irrespective of the equilibrium path of play, and the equilibrium path of play continues after time $t_0+\epsilon$.

Definition 3. An equilibrium σ is a strategy profile such that:

(i) $\beta(x,L,1)=1$ if $p(L,x)b>s$, $\beta(x,L,1)=0$ if $p(L,x)b<s$ and $\beta(x,L,0)=1$ if $r>s$.

(ii) If $V_{m(L,\pi)}(L,\pi')>V_{m(L,\pi)}(L,1-\pi')$, then $\alpha(L,\pi)=\{\pi'\}$.

(iii) If $V_{m(L,\pi)}(L,1)=V_{m(L,\pi)}(L,0)$ but $\overline{V}_{m(L,\pi)}(L,\pi',\epsilon)-\overline{V}_{m(L,\pi)}(L,1-\pi',\epsilon)>0$ for all $\epsilon>0$ small enough, then $\alpha(L,\pi)=\{\pi'\}$.

Part (i) of the definition of equilibrium says that agents make membership decisions that maximize their flow payoffs. Part (ii) says that the pivotal agent chooses

[38] This would happen, for example, if future decision-makers coming immediately after $m(L,\pi)$ are expected to choose the same policy π' independently of the choice made by $m(L,\pi)$.

her preferred policy based on her expected utility, assuming that the equilibrium strategies are played in the continuation. Part (iii) is a tie-breaking rule which enforces optimal behavior even when the agent's policy choice only affects the path of play for an infinitesimal amount of time.

Proofs

Lemma 1. *Suppose that the initial distribution of priors is f_α for some $\alpha \geq 0$, as in Proposition 9. The posterior belief of the median member of the organization at time t, provided that experimentation has continued from time 0 to time t and no successes have been observed, is*

$$p_t(m_t) = \frac{s + (1-\lambda)(b-s)e^{-bt}}{\lambda(b-s) + s + (1-\lambda)(b-s)e^{-bt}}$$

Proof of Lemma 1. The posterior belief of agent x at time t is given by $p_t(x) = \frac{xe^{-bt}}{xe^{-bt}+1-x}$. Using the fact that $p_t(y_t) = \frac{s}{b}$ for the marginal member y_t, we set $\frac{y_t e^{-bt}}{y_t e^{-bt}+1-y_t} = \frac{s}{b}$. Solving for y_t, we obtain

$$y_t = \frac{\frac{s}{b}}{\frac{s}{b} + \left(1 - \frac{s}{b}\right) e^{-bt}} = \frac{s}{s + (b-s)e^{-bt}}$$

The median m_t must satisfy the condition $2\int_{m_t}^1 f_\alpha(x)dx = \int_{y_t}^1 f_\alpha(x)dx$, so that $2(1 - m_t)^{\alpha+1} = (1-y_t)^{\alpha+1}$. Hence $1 - m_t = \lambda(1-y_t)$, which implies that

$$m_t = 1 - \lambda + \lambda y_t = 1 - \lambda + \lambda \frac{s}{s + (b-s)e^{-bt}} = \frac{s + (1-\lambda)(b-s)e^{-bt}}{s + (b-s)e^{-bt}}$$

Substituting the above expression into the formula for $p_t(x)$, we obtain

$$p_t(m_t) \quad = \frac{s + (1-\lambda)(b-s)e^{-bt}}{\lambda(b-s) + s + (1-\lambda)(b-s)e^{-bt}}$$

In particular, if $\alpha = 0$, then f is uniform and we have $p_t(m_t) = \frac{2s+(b-s)e^{-bt}}{b+s+(b-s)e^{-bt}}$. ■

We now provide a formula for $V(x)$.

Lemma 2. *Let $t(x)$ denote the time it will take for an agent's posterior belief to go* *from x to $\frac{s}{b}$ (provided that no successes are observed during this time), at which time she would leave the organization. Then*

$$V(x) = xb\frac{1}{\gamma} + (1-x)e^{-\gamma t(x)}\frac{s}{\gamma} - x(b-s)\frac{e^{-(\gamma+b)t(x)}}{\gamma+b}$$

Proof of Lemma 2.

Let $P_t = x(1-e^{-bt})$ denote the probability that an agent with prior belief x assigns to having a success by time t. Then

$$V(x) = x\int_0^{t(x)} be^{-\gamma\tau}d\tau + \int_{t(x)}^{\infty} \left(P_\tau b + (1-P_\tau)s\right)e^{-\gamma\tau}d\tau$$

The first term is the payoff from time 0 to time $t(x)$, when the agent stays in the organization. The second term is the payoff after time $t(x)$, when the agent leaves the organization and obtains the flow payoff s thereafter, unless the risky technology has had a success (in which case the agent returns to the organization and receives a guaranteed expected flow payoff b). We have

$$\begin{aligned}
V(x) &= x\int_0^{t(x)} be^{-\gamma\tau}d\tau + \int_{t(x)}^{\infty} se^{-\gamma\tau}d\tau + \int_{t(x)}^{\infty} P_\tau(b-s)e^{-\gamma\tau}d\tau \\
&= xb\frac{1-e^{-\gamma t(x)}}{\gamma} + e^{-\gamma t(x)}\frac{s}{\gamma} + x(b-s)\left(\frac{e^{-\gamma t(x)}}{\gamma} - \frac{e^{-(\gamma+b)t(x)}}{\gamma+b}\right) \\
&= xb\frac{1}{\gamma} + (1-x)e^{-\gamma t(x)}\frac{s}{\gamma} - x(b-s)\frac{e^{-(\gamma+b)t(x)}}{\gamma+b}.
\end{aligned}$$

■

Lemma 3. *Let $t^y(x)$ denote he time it takes for an agent's posterior belief to go from x to y. Then*

$$t^y(x) = -\frac{\ln\left(\frac{y}{1-y}\frac{1-x}{x}\right)}{b} \qquad t(x) = -\frac{\ln\left(\frac{s(1-x)}{(b-s)x}\right)}{b}$$

If $x = \frac{2s}{b+s}$, then $e^{-bt(x)} = \frac{1}{2}$. If $x = \frac{s}{\lambda b + (1-\lambda)s}$, then $e^{-bt(x)} = \lambda$.

Proof of Lemma 3.

We solve $p_t(x) = \frac{xe^{-bt}}{xe^{-bt}+1-x} = y$ for t. Then we obtain $e^{-bt^y(x)} = \frac{y}{1-y}\frac{1-x}{x}$ or, equivalently, $t^y(x) = -\frac{\ln\left(\frac{y}{1-y}\frac{1-x}{x}\right)}{b}$.

In particular, $t(x) = t^{\frac{s}{b}}(x) = -\frac{\ln\left(\frac{s(1-x)}{(b-s)x}\right)}{b}$. Substituting $x = \frac{2s}{b+s}$ into $e^{-bt(x)} = \frac{s(1-x)}{(b-s)x}$ and simplifying, we obtain $e^{-bt} = \frac{1}{2}$. Substituting $x = \frac{s}{\lambda b+(1-\lambda)s}$ into $e^{-bt(x)} = \frac{s(1-x)}{(b-s)x}$ and simplifying, we obtain $e^{-bt(x)} = \lambda$. ■

Lemma 4. *Let $V_x(L, \pi)$ denote the value function of an agent with prior x when the initial state is (L, π). Then for all (L, π), $x \mapsto V_x(L, \pi)$ is strictly increasing for all agents x that are in the organization while it experiments, at a set of times of positive measure with a positive probability (on the equilibrium path).*

Proof of Lemma 4.

Consider two agents $x' > x$. Let $V_{x'}^x(L, \pi)$ denote the payoff to agent x' from copying the equilibrium strategy of agent x. When x and x' are outside the organization, their flow payoffs are equal to s and do not depend on their priors.

When x' is in the organization, if the organization is using the risky policy, at a continuation where the state variable is $\tilde{L}$, x''s expected flow payoff is $p\left(\tilde{L}, x'\right) b$. Because $x' > x$, we have $p\left(\tilde{L}, x'\right) > p\left(\tilde{L}, x\right)$ and thus $p\left(\tilde{L}, x'\right) b > p\left(\tilde{L}, x\right) b$, so x''s flow payoff is higher than x's when x and x' are members.

We then have $V_{x'}^x(L, \pi) > V_x(L, \pi)$ if x is in the organization while it experiments, at a set of times of positive measure with a positive probability. Because $V_{x'}(L, \pi) \geq V_{x'}^x(L, \pi)$, we have $V_{x'}(L, \pi) > V_x(L, \pi)$, as required. ■

Corollary 1. *If the organization experiments forever on the equilibrium path, then $V_x(L, \pi) = V_{p(L,x)}(1, \pi)$. Moreover, $V_{p(L,x)}(1, \pi)$ is increasing in $p(L, x)$.*

Lemma 5. *For any policy path $(\pi_t)_t$ with left and right-limits everywhere, there is another policy path $(\hat{\pi}_t)_t$ such that $\hat{\pi}_0 = \pi_0$, $(\hat{\pi}_t)_t$ is càdlàg for all $t > 0$, and $(\hat{\pi}_t)_t$ is equal to $(\pi_t)_t$ almost everywhere.*[39]

Proof of Lemma 5.

Define $\hat{\pi}_0 = \pi_0$ and $\hat{\pi}_t = \pi_{t+}$ for all $t > 0$. Let $\mathcal{T} = \mathbb{R}_{\geq 0} \setminus \{t \geq 0 : \pi_{t-} = \pi_t = \pi_{t+}\}$. Because $(\pi_t)_t$ has left and right-limits everywhere, $\mathcal{T}$ must be countable—

[39] Hence it is payoff-equivalent to π_t and generates the same learning path $(L_t)_t$.

otherwise $\mathcal{T}$ would have an accumulation point t_0, and either the left-limit or right-limit of $(\pi_t)_t$ at t_0 would not be well-defined. Then, since $\hat{\pi}_t = \pi_t$ for all $t \notin \mathcal{T}$, $(\hat{\pi}_t)_t$ and $(\pi_t)_t$ only differ on a countable set. Moreover, it is straightforward to show that, for all $t > 0$, $\hat{\pi}_{t-} = \pi_{t-}$ and $\hat{\pi}_{t+} = \pi_{t+} = \hat{\pi}$, so $(\hat{\pi}_t)_t$ is càdlàg. ■

Corollary 2. *For any strategy profile $(\beta, \alpha, (L_t, \pi_t)_t | (L, \pi, \theta))$ the stochastic process $(L_t, \hat{\pi}_t)_t | (L, \pi, \theta)$ (where $(\hat{\pi}_t)_t$ is as in Lemma 5) has càdlàg paths, satisfies Conditions (a) and (b), and induces a path of play that yields the same payoffs as the strategy.*

Lemma 6 (Recursive Decomposition). *Let $\Theta \subseteq \mathbb{R}^n$ be a closed set, let $(\theta_t)_t$ be a right-continuous progressively measurable Markov process with support contained in Θ, let f be a bounded function, and let*

$$U(\theta_0) = \int_0^\infty e^{-\gamma t} E_{\theta_0}[f(\theta_t)]dt$$

Let A be a closed subset of Θ and define a stochastic process $(s_t)_t$ with a co-domain $(A \cup \{\varnothing\})$ as follows: $s_t = \theta$ if there exists $t' \leq t$ such that $\theta_{t'} = \theta$ and $\theta_{t''} \notin A$ for all $t'' < t'$. If this is not true for any $\theta \in A$, then $s_t = \varnothing$.[40] *Then*

$$U(\theta_0) = \int_0^\infty e^{-\gamma t} E_{\theta_0}\left[f(\theta_t)\mathbb{1}_{\{s_t=\varnothing\}}\right] dt + \int_A U(\theta) dP_s$$

where P_s is defined as follows: P_{s_t} is the probability measure on $A \cup \varnothing$ induced by s_t, and $P_s = \gamma \int_0^\infty e^{-\gamma t} P_{s_t} dt$.

Proof of Lemma 6.

$$\begin{aligned} U(\theta_0) = \int_0^\infty & e^{-\gamma t} E_{\theta_0}[f(\theta_t)]dt = \\ & \int_0^\infty e^{-\gamma t} E_{\theta_0}\left[f(\theta_t)\mathbb{1}_{\{s_t=\varnothing\}}\right] dt + \int_0^\infty e^{-\gamma t} E_{\theta_0}\left[f(\theta_t)\mathbb{1}_{\{s_t\in A\}}\right] dt \end{aligned}$$

So it remains to show that

$$\int_A U(\theta) dP_s = \int_0^\infty e^{-\gamma t} E_{\theta_0}\left[f(\theta_t)\mathbb{1}_{\{s_t\in A\}}\right] dt$$

Define a random variable z with a co-domain $(A \times [0, \infty)) \cup \{\varnothing\}$ as follows: $z = (\theta, t)$

[40]In other words, s_t takes the value of the first $\theta \in A$ that $(\theta_t)_t$ hits.

if $\theta_t = \theta \in A$ and $\theta_{t'} \notin A$ for all $t' < t$. If this is not true for any $\theta \in A$ and $t \geq 0$, then $z = \varnothing$.[41] Let P_z be the probability measure on $(A \times [0,\infty)) \cup \{\varnothing\}$ induced by z. Let $\theta(z)$ and $t(z)$ be the random variables equal to the first and second coordinates of z, conditional on $z \neq \varnothing$. Note that $s_t = \theta$ if and only if $z = (\theta, t')$ for some $t' \leq t$. Then we can write

$$\int_0^\infty e^{-\gamma t} E_{\theta_0}\left[f(\theta_t)\mathbb{1}_{\{s_t \in A\}}\right] dt = \int_0^\infty e^{-\gamma t} E_{\theta_0}\left[f(\theta_t)\mathbb{1}_{\{z \in A\times[0,\infty)\}}\mathbb{1}_{\{t \geq t(z)\}}\right] dt =$$
$$= \int_{A\times[0,\infty)} \left(\int_{t(z)}^\infty e^{-\gamma t} E_{\theta_0}[f(\theta_t)|z] dt\right) dP_z = \int_{A\times[0,\infty)} e^{-\gamma t(z)} U(\theta(z)) dP_z =$$
$$= \int_{A\times[0,\infty)} \left(\int_{t(z)}^\infty \gamma e^{-\gamma t} dt\right) U(\theta(z)) dP_z = \int_0^\infty \int_{A\times[0,\infty)} \gamma e^{-\gamma t}\mathbb{1}_{\{t \geq t(z)\}} U(\theta(z)) dP_z dt =$$
$$= \int_0^\infty \gamma e^{-\gamma t}\left(\int_{A\times[0,\infty)} \mathbb{1}_{\{t \geq t(z)\}} U(\theta(z)) dP_z\right) dt = \int_0^\infty \gamma e^{-\gamma t}\left(\int_A U(\theta) dP_{s_t}\right) dt = \int_A U(\theta) dP_s$$

as desired. ■

Let $W_T(x)$ denote the continuation value of an agent with current belief x in an equilibrium in which the organization stops experimenting after a length of time T.[42] Let $T^* = \operatorname{argmax}_T W_T(x)$ denote the optimal amount of time that an agent with prior x would want to experiment for if she was always in control of the organization.

Lemma 7. *(i) $T \mapsto W_T(x)$ is differentiable for all $T \in (0,\infty)$ and right-differentiable at $T = 0$.*

(ii) $W_0(x) = \frac{r}{\gamma}$ and $\left.\frac{\partial_+ W_T(x)}{\partial T}\right|_{T=0} = \max\{xb, s\} - r + \frac{xb(b-r)}{\gamma}$.

(iii) $T \mapsto W_T(x)$ is strictly increasing for $T \in [0, T^*]$ and strictly decreasing for $T > T^*$.

(iv) If $V(x) > \frac{r}{\gamma}$, then $W_T(x) > \frac{r}{\gamma}$ for all $T > 0$.

Proof of Lemma 7.

[41] In other words, z takes the value of the first $\theta \in A$ that $(\theta_t)_t$ hits, and the time when it hits.
[42] This is defined for $T \in [0,\infty]$, where $W_\infty(x) = V(x)$.

Fix $T_0 \geq 0$ and $\epsilon > 0$. By Lemma 6, we have

$$\begin{aligned} W_{T_0+\epsilon}(x) - W_{T_0}(x) &= e^{-rT_0} Q_{T_0}(x) \left(W_\epsilon(p_{T_0}(x)) - W_0(p_{T_0}(x)) \right) \\ &= e^{-rT_0} Q_{T_0}(x) \left(W_\epsilon(p_{T_0}(x)) - \frac{r}{\gamma} \right) \end{aligned}$$

where $Q_T(x)$ is the probability that there is no success up to time T, based on the prior x, and $p_T(x)$ is the posterior belief of an agent with prior x in this case. Then

$$\begin{aligned} \left. \frac{\partial W_T(x)}{\partial T} \right|_{T=T_0} &= \lim_{\epsilon \searrow 0} \frac{W_{T_0+\epsilon}(x) - W_{T_0}(x)}{\epsilon} = \\ &= \lim_{\epsilon \searrow 0} e^{-rT_0} Q_{T_0}(x) \frac{W_\epsilon(p_{T_0}(x)) - \frac{r}{\gamma}}{\epsilon} = e^{-rT_0} Q_{T_0}(x) \left. \frac{\partial W_T(p_{T_0}(x))}{\partial T} \right|_{T=0} \end{aligned}$$

The case with $\epsilon < 0$ is analogous. Then it is enough to prove that $T \mapsto W_T(x)$ is right-differentiable at $T = 0$. This can be done by calculating $W_T(x)$ explicitly for small $T > 0$. We do this for $x > \frac{s}{b}$. In this case, for T sufficiently small, x is in the organization because the experiment with the risky policy is going to stop before she wants to leave. Using that $1 - Q_T(x) = x(1 - e^{-bT})$,

$$\begin{aligned} W_T(x) &= \int_0^T x b e^{-\gamma t} dt + \int_T^\infty e^{-\gamma t} (Q_T(x) r + (1 - Q_T(x)) b) dt = \\ &= x b \frac{1 - e^{-\gamma T}}{\gamma} + \frac{e^{-\gamma T}}{\gamma} \left(r + x \left(1 - e^{-bT} \right) (b - r) \right) \end{aligned}$$

This implies that $\left. \frac{\partial_+ W_T(x)}{\partial T} \right|_{T=0} = xb - r + \frac{xb(b-r)}{\gamma}$.

Similarly, it can be shown that if $x \leq \frac{s}{b}$, then $\left. \frac{\partial_+ W_T(x)}{\partial T} \right|_{T=0} = s - r + \frac{xb(b-r)}{\gamma}$. This proves (i) and (ii).

For (iii), note that, by our previous result, $\left. \frac{\partial W_T(x)}{\partial T} \right|_{T=T_0}$ is positive (negative) whenever $\left. \frac{\partial W_T(p_{T_0}(x))}{\partial T} \right|_{T=0}$ is positive (negative). In addition, it follows from our calculations that $y \mapsto \left. \frac{\partial W_T(y)}{\partial T} \right|_{T=0}$ is increasing and $T_0 \mapsto p_{T_0}(x)$ is decreasing. Moreover, for large T_0, $p_{T_0}(x)$ is close to zero, so $\left. \frac{\partial W_T(p_{T_0}(x))}{\partial T} \right|_{T=0}$ is negative. It follows that $T \mapsto W_T(x)$ is single-peaked. If $\left. \frac{\partial W_T(x)}{\partial T} \right|_{T=0} > 0$, then the peak is the unique T^*

satisfying $\left.\frac{\partial W_T(x)}{\partial T}\right|_{T=T^*} = 0$. If $\left.\frac{\partial W_T(x)}{\partial T}\right|_{T=0} \leq 0$, then $T^* = 0$.

Hence if $0 < T \leq T^*$, then $W_T(x) > W_0(x) = \frac{r}{\gamma}$ because in this case $T \mapsto W_T(x)$ is increasing by (iii), and if $T > T^*$, then $W_T(x) \geq \lim_{T\to\infty} W_T(x) = V(x)$ because in this case $T \mapsto W_T(x)$ is decreasing by (iii). This proves (iv). ■

Lemma 8. *Let $m(L)$ and $\tilde{m}(L)$ denote the median voter when the state variable is $(L,1)$ and the density is f and $\tilde{f}$ respectively. Suppose that $\tilde{f}$ MLRP-dominates f. Then $\tilde{m}(L) \geq m(L)$ for all L.*

Proof of lemma 8.

Let $y(L)$ denote the indifferent agent given information L under either density (note that $y(L)$ is given by the condition $p(L, y(L)) = \frac{s}{b}$, which is independent of the density). By definition, we have $\int_{y(L)}^{m(L)} f(x)dx = \int_{m(L)}^{1} f(x)dx$. Suppose that $\tilde{m}(L) < m(L)$. This is equivalent to

$$\int_{y(L)}^{m(L)} s(x)f(x)dx = \int_{y(L)}^{m(L)} \tilde{f}(x)dx > \int_{m(L)}^{1} \tilde{f}(x)dx = \int_{m(L)}^{1} s(x)f(x)dx$$

where $s(x) = \frac{\tilde{f}(x)}{f(x)}$. Since $\tilde{f}$ MLRP-dominates f, $s(x)$ is weakly increasing. Thus

$$\int_{y(L)}^{m(L)} s(m(L))f(x)dx \geq \int_{y(L)}^{m(L)} s(x)f(x)dx > \int_{m(L)}^{1} s(x)f(x)dx \geq \int_{m(L)}^{1} s(m(L))f(x)dx$$

which is a contradiction. ■

Lemma 9. *Let $m(L)$ and $\tilde{m}(L)$ denote the median voters when the state variable is L under the uniform density and a non-decreasing density f respectively. Suppose that $y(L) \to 1$ as $L \to \infty$. Then $\frac{1-\tilde{m}(L)}{1-m(L)} \to 1$ as $L \to \infty$.*

Proof of Lemma 9.

Given a state variable L and the marginal member $y(L)$ corresponding to it, let $f_{0L} = f(y(L))$ and $f_1 = f(1)$. By Lemma 8, we have $m(L) \leq \tilde{m}(L) \leq \hat{m}(L)$ where $\hat{m}(L)$ is the median corresponding to a density $\hat{f}$ such that $\hat{f}(x) = f_{0L}$ for $x \in [y(L), \hat{m}(L)]$ and $\hat{f}(x) = f_1$ for $x \in [\hat{m}(L), 1]$.

By construction, because $\hat{m}(L)$ is the median, we have $f_{0L}(\hat{m}(L) - y(L)) =$

$f_1(1-\hat{m}(L))$, so $\hat{m}(L) = \frac{f_{0L}y(L)+f_1}{f_{0L}+f_1}$. Thus $1-\hat{m}(L) = \frac{f_{0L}(1-y(L))}{f_{0L}+f_1}$ and, because $m(L) = \frac{y(L)+1}{2}$ so that $1-m(L) = \frac{1-y(L)}{2}$, we have $\frac{1-\hat{m}(L)}{1-m(L)} = \frac{2f_{0L}}{f_{0L}+f_1}$.

Since f is increasing, using the fact that $f(x) \to \sup_{y\in[0,1)} f(y)$ as $x \to 1$, we find that $f(x) \to f(1)$ as $x \to 1$. Then, as $t \to \infty$, we have $y(L) \to 1$, $f_{0L} = f(y(L)) \to f_1$ and $\frac{1-\hat{m}(L)}{1-m(L)} \to 1$. ■

Lemma 10. *Let x_t, $\tilde{x}_t$ be two time-indexed sequences of agents such that $x_t \leq \tilde{x}_t$ for all t and $x_t \to 1$ as $t \to \infty$. If $\frac{1-x_t}{1-\tilde{x}_t} \to 1$, then $\frac{p_t(\tilde{x}_t)}{p_t(x_t)} \to 1$.*

Proof of Lemma 10.

Using the formula for the posterior beliefs, we have

$$\frac{p_t(\tilde{x}_t)}{p_t(x_t)} = \frac{\tilde{x}_t}{\tilde{x}_t + (1-\tilde{x}_t)L_t}\frac{x_t + (1-x_t)L_t}{x_t} = \frac{\tilde{x}_t}{x_t}\frac{x_t + (1-x_t)L_t}{\tilde{x}_t + (1-\tilde{x}_t)L_t}.$$

Since $x_t \to 1$ and $\tilde{x}_t \geq x_t$ for all t, $\tilde{x}_t \to 1$, whence $\frac{\tilde{x}_t}{x_t} \to 1$. In addition, since $\frac{1-x_t}{1-\tilde{x}_t} \to 1$, $\frac{(1-x_t)L_t}{(1-\tilde{x}_t)L_t} \to 1$. As a result, for all t,

$$\min\left\{\frac{x_t}{\tilde{x}_t}, \frac{(1-x_t)L_t}{(1-\tilde{x}_t)L_t}\right\} \leq \frac{x_t + (1-x_t)L_t}{\tilde{x}_t + (1-\tilde{x}_t)L_t} \leq \max\left\{\frac{x_t}{\tilde{x}_t}, \frac{(1-x_t)L_t}{(1-\tilde{x}_t)L_t}\right\}$$

so $\frac{x_t+(1-x_t)L_t}{\tilde{x}_t+(1-\tilde{x}_t)L_t} \to 1$, which concludes the proof. ■

Lemma 11 shows that agents strictly prefer the risky policy after a success.

Lemma 11. *In any equilibrium, $\alpha(0,1) = \alpha(0,0) = 1$.*

Proof of Lemma 11.

Observe that $L_{t_0} = 0$ implies $L_t = 0$ for all $t \geq t_0$ no matter what policy path is followed, and hence $p(L_t, x) = 1$ for all t and x. For the rest of the argument, we can then write $V(0,\pi)$ instead of $V_x(0,\pi)$. By Lemma 6, there is $\rho \in [0,1]$ such that

$$V(0,0) = \rho\frac{r}{\gamma} + (1-\rho)V(0,1) \tag{2}$$

It follows that there exist $\eta \in [0,1]$ and $\eta' \in [0,1]$ such that $\eta \geq \eta'$[43] and

$$V(0,0) = \eta\frac{r}{\gamma} + (1-\eta)\frac{b}{\gamma} \qquad V(0,1) = \eta'\frac{r}{\gamma} + (1-\eta')\frac{b}{\gamma}$$

η and η' are the discounted fractions of the expected time that the organization spends on the safe policy, when starting in states $(0,0)$ and $(0,1)$, respectively.

Observe that if $\eta > \eta'$, then $V(0,0) < V(0,1)$. In particular, $V_{m(0,\pi)}(0,1) > V_{m(0,\pi)}(0,0)$ for all π, which implies that $\alpha(0,\pi) = 1$ for all π by Condition (ii). If $\eta = \eta'$, then $V(0,0) = V(0,1)$. Because $\overline{V}(0,0,\epsilon) < \overline{V}(0,1,\epsilon)$ for any $\epsilon > 0$, by Condition (iii), in this case we must also have $\alpha(0,\pi) = 1$ for all π. ■

Lemma 12. *For any state (L,π), there is a CDF G with support*[44] *contained in $[0,\infty]$ such that*

$$V_x(L,\pi) = \int_0^\infty W_T(p(L,x))dG(T)$$

for all $x \in [0,1]$, where $W_T(y)$ is as defined in Lemma 7.

Similarly, for any state (L,π) and any $\epsilon > 0$, there is a distribution G_ϵ with support contained in $[0,\infty]$ such that

$$\overline{V}_x(L,\pi,\epsilon) = \int_0^\infty W_T(p(L,x))dG_\epsilon(T)$$

for any $x \in [0,1]$.

Proof of Lemma 12.

We prove the first statement. The proof of the second statement is analogous.

Note that we can, without loss of generality, assume that the distribution over future states (L,π) induced by the continuation starting in state (L,π) satisfies the following: the policy is equal to 1 in the beginning and, if it ever changes from 1 to

[43] We have $\eta \geq \eta'$ for the following reason. $V(0,1) = \eta'\frac{r}{\gamma} + (1-\eta')\frac{b}{\gamma}$ and (2) imply that $\eta\frac{r}{\gamma} + (1-\eta)\frac{b}{\gamma} = V(0,0) = \rho\frac{r}{\gamma} + (1-\rho)V(0,1) = (\rho + (1-\rho)\eta')\frac{r}{\gamma} + (1-\rho)(1-\eta')\frac{b}{\gamma}$. Then $\eta = \rho + (1-\rho)\eta'$, which implies that $\eta \geq \eta'$, as required.

[44] G is a degenerate CDF that can take the value ∞ with positive probability. Equivalently, G satisfies all the standard conditions for the definition of a CDF, except that $\lim_{T\to\infty} G(T) \leq 1$ instead of $\lim_{T\to\infty} G(T) = 1$. This is needed to allow for the case where experimentation continues forever.

0, it never changes back to 1. Indeed, suppose that π switches from 1 to 0 at time t and switches back at a random time $t+\nu$, where ν is distributed according to some CDF H. Let $p = \int_0^\infty e^{-\gamma\nu} dH(\nu)$. Then a continuation path on which the policy only switches to 0 at time t with probability $1-p$ and never returns to 1 after switching induces the same discounted distribution over future states.

Under the above assumption and given that the policy always remains at 1 after a success by Lemma 11, the path of play can be described as follows: experimentation continues uninterrupted until a success or a permanent stop. Then we can let G be the CDF of the stopping time, conditional on no success being observed. ■

Lemma 13 shows that switches to the safe policy are permanent.

Lemma 13. *In any equilibrium, for any L, if $0 \in \alpha(L,1)$, then $\alpha(L,0) = 0$.*

Proof of Lemma 13.

If $L=0$, then $\alpha(0,\pi) = 1$ for all π by Lemma 11, so the statement is vacuously true. Suppose then that $L > 0$. Suppose for the sake of contradiction that the statement is false.

Observe that for all L there is $\rho_L \in [0,1]$ independent of x such that

$$V_x(L,0) = \rho_L \frac{r}{\gamma} + (1-\rho_L)V_x(L,1) \tag{3}$$

for all x. This follows from Lemma 6, with the added observation that ρ_L (equivalently, P_s in the notation of Lemma 6) is independent of x in this case because the stochastic process governing (L,π) is independent of x if $\pi = 0$.[45] We now consider three cases.

Case 1: Suppose that $\rho_L > 0$, and that the expected amount of experimentation under the continuation starting in state $(L,1)$ is positive. By Lemma 4, $V_x(L,1)$ is strictly increasing in x. Then equation (3) implies that $V_x(L,1) - V_x(L,0)$ is strictly increasing in x. Since $m(L,1) > m(L,0)$, we have $V_{m(L,1)}(L,1) - V_{m(L,1)}(L,0) >$

[45] If (L_t,π_t) has càdlàg paths, this follows from Lemma 6. If not, then Lemma 6 cannot be applied because the stochastic process in question is not necessarily right-continuous. However, we can use Corollary 2 of Lemma 5 to obtain a payoff-equivalent path of play with càdlàg paths and then apply Lemma 6 to it.

$V_{m(L,0)}(L,1)-V_{m(L,0)}(L,0)$. Since $1 \in \alpha(L,0)$ implies that $V_{m(L,0)}(L,1)-V_{m(L,0)}(L,0) \geq 0$, we have $V_{m(L,1)}(L,1) - V_{m(L,1)}(L,0) > 0$, and thus $\alpha(L,1) = 1$, a contradiction.

Case 2: Suppose that $\rho_L = 0$. We make two observations. First, $V_x(L,0) = V_x(L,1)$ for all x. Second, the expected amount of experimentation under the continuation starting in state $(L,1)$ is positive. Indeed, $\rho_L = 0$ implies that, conditional on the state at t being $(L_t, \pi_t) = (L,0)$, we have $\inf\{t' > t : \pi_{t'} = 1\} = t$ a.s. Since Condition (d) requires that π_{t+} exists and the result that $\inf\{t' > t : \pi_{t'} = 1\} = t$ a.s. rules out that $\pi_{t+} = 0$ with a positive probability, it must be that $\pi_{t+} = 1$ a.s. In turn, this implies that $\inf\{t' > t : \pi_{t'} = 0\} > t$ a.s. Then $E\left[\inf\{t' > t : \pi_{t'} = 0\}\right] - t > 0$.

By definition, we have

$$\overline{V}_x(L,0,\epsilon) = \rho_\epsilon \frac{r}{\gamma} + (1-\rho_\epsilon)V_x(L,1) \tag{4}$$

for $\rho_\epsilon = 1 - e^{-\gamma\epsilon}$.

In the following argument, for convenience, we subtract $\frac{r}{\gamma}$ from every value function.[46] Note that, by definition, because we lock policy 1 in for time ϵ, G_ϵ satisfies $1 - G_\epsilon(T) = \min\left\{\frac{1-G(T)}{1-G(\epsilon)}, 1\right\}$. Then for $T \in [0,\epsilon]$, $1 - G_\epsilon(T) = 1$ and for $T > \epsilon$, $1 - G_\epsilon(T) = \frac{1-G(T)}{1-G(\epsilon)}$. Hence for $\epsilon > 0$ sufficiently small we have

$$\begin{aligned}
\overline{V}_x(L,1,\epsilon) &= \int_0^\infty W_T(p(L,x))dG_\epsilon(T) = \\
&= \int_0^\epsilon W_T(p(L,x))dG_\epsilon(T) + \int_\epsilon^\infty W_T(p(L,x))dG_\epsilon(T) = \\
&= 0 + \frac{1}{1-G(\epsilon)}\int_\epsilon^\infty W_T(p(L,x))dG(T) = \\
&= \frac{V_x(L,1)}{1-G(\epsilon)} - \frac{1}{1-G(\epsilon)}\int_0^\epsilon W_T(p(L,x))dG(T) = \frac{V_x(L,1)}{1-G(\epsilon)} + G(\epsilon)\mathcal{O}(\epsilon)
\end{aligned}$$

The first part of the third equality follows from the fact that $G_\epsilon(T) = 0$ for all $T \in [0,\epsilon]$ and the second part of the third equality follows from the fact that $G_\epsilon(T) = \frac{G(T)-G(\epsilon)}{1-G(\epsilon)}$ for $T > \epsilon$. The last equality follows from the fact that $\lim_{\epsilon\to 0} G(\epsilon) = 0$ since

[46] That is, we let $\mathring{V}_x(L,\pi) = V_x(L,\pi) - \frac{r}{\gamma}$, $\mathring{W}_T(x) = W_T(x) - \frac{r}{\gamma}$, $\mathring{\overline{V}}_x(L,\pi,\epsilon) = \overline{V}_x(L,\pi,\epsilon) - \frac{r}{\gamma}$. For the rest of this proof, we work with the normalized functions $\mathring{V}$, $\mathring{W}$, $\mathring{\overline{V}}$, but drop the operator $^\circ$ to simplify notation.

$\inf\{t' > t : \pi_{t'} = 1\} = t$ a.s. and from the fact that $\left.\frac{\partial_+ W_T(x)}{\partial T}\right|_{T=0} = \max\{xb, s\} - r + \frac{xb(b-r)}{\gamma}$ by part (ii) of lemma 7.[47]

Suppose for the sake of contradiction that $V_{m(L,0)}(L,1) < 0$. Note that (4) then implies that $V_{m(L,0)}(L,1) < \overline{V}_{m(L,0)}(L,0,\epsilon)$. Therefore, we have $\overline{V}_{m(L,0)}(L,1,\epsilon) \leq V_{m(L,0)}(L,1) < \overline{V}_{m(L,0)}(L,0,\epsilon)$ for all $\epsilon > 0$ sufficiently small and hence $\alpha(L,0) = 0$, a contradiction. Hence $V_{m(L,0)}(L,1) \geq 0$. It follows that, because $m(L,1) > m(L,0)$ and, by Lemma 4, $x \mapsto V_x(L,1)$ is strictly increasing, we have $V_{m(L,1)}(L,1) > 0$. Then $\overline{V}_x(L,1,\epsilon) = \frac{V_x(L,1)}{1-G(\epsilon)} + G(\epsilon)\mathcal{O}(\epsilon)$ implies that $\overline{V}_{m(L,1)}(L,1,\epsilon) \geq V_{m(L,1)}(L,1)$.[48] Moreover, because $V_{m(L,1)}(L,1) > 0$, we have $V_{m(L,1)}(L,1) > \overline{V}_{m(L,1)}(L,0,\epsilon)$, as $\overline{V}_{m(L,1)}(L,0,\epsilon)$ is a convex combination of $V_{m(L,1)}(L,1)$ and 0.[49] Then $\overline{V}_{m(L,1)}(L,1,\epsilon) \geq V_{m(L,1)}(L,1) > \overline{V}_{m(L,1)}(L,0,\epsilon)$ for all $\epsilon > 0$ sufficiently small. By Condition (iii), this implies that $\alpha(L,1) = 1$, a contradiction.

Case 3: Suppose that the expected amount of experimentation starting in state $(L,1)$ is zero. In this case $V_x(L,0) = V_x(L,1) = \frac{r}{\gamma}$ for all x, and $\overline{V}_x(L,0,\epsilon) = \frac{r}{\gamma}$ for all x and $\epsilon > 0$. Again, we subtract $\frac{r}{\gamma}$ from every value function for simplicity.

By definition, for all $\epsilon > 0$ the path starting in state $(L,1,\epsilon)$ has a positive expected amount of experimentation. Moreover, G_ϵ defined in Lemma 12 is FOSD-decreasing in ϵ (that is, if $\epsilon' < \epsilon$, then $G_{\epsilon'} \geq G_\epsilon$) and hence, taken as a function of ϵ, has a pointwise limit G (that is, $G_\epsilon(T) \underset{\epsilon\to 0}{\longrightarrow} G(T)$ for all $T \geq 0$). Then

$$\overline{V}_x(L,1,\epsilon) \underset{\epsilon\to 0}{\longrightarrow} \int_0^\infty W_T(p(L,x))dG(T)$$

Since $1 \in \alpha(L,0)$, there exists a sequence $\epsilon_n \searrow 0$ such that $\overline{V}_{m(L,0)}(L,1,\epsilon_n) \geq 0$ for all n,[50] whence $\lim_{\epsilon\to 0} \overline{V}_{m(L,0)}(L,1,\epsilon) \geq 0$.

[47] In greater detail, $\int_0^\epsilon W_T(p(L,x))dG(T) \approx \int_0^\epsilon (\alpha_0 T + \alpha_1)dG(T) \leq \int_0^\epsilon (\alpha_0\epsilon + \alpha_1)dG(T) = (\alpha_0\epsilon + \alpha_1)\int_0^\epsilon dG(T) = (\alpha_0\epsilon + \alpha_1)(G(\epsilon) - G(0)) = (\alpha_0\epsilon + \alpha_1)G(\epsilon) = \alpha_0\epsilon G(\epsilon) = G(\epsilon)\mathcal{O}(\epsilon)$ where we have used the fact that we subtracted $\frac{r}{\gamma}$ from every value function to get rid of the constant α_1.

[48] $\overline{V}_{m(L,1)}(L,1,\epsilon) \geq V_{m(L,1)}(L,1)$ is then equivalent to $G(\epsilon)(1-G(\epsilon))\mathcal{O}(\epsilon) \geq -G(\epsilon)V_{m(L,1)}(L,1)$, which is satisfied for $V_{m(L,1)}(L,1) > 0$.

[49] Recall that we have subtracted $\frac{r}{\gamma}$ from every value function.

[50] Suppose for the sake of contradiction that $1 \in \alpha(L,0)$ and such a sequence does not exist. Then for all $\epsilon > 0$ sufficiently small we have $\overline{V}_{m(L,0)}(L,1,\epsilon) < 0$ (note that we have used the fact that we subtract $\frac{r}{\gamma}$ from every value function here). Then $\overline{V}_{m(L,0)}(L,1,\epsilon) < \overline{V}_{m(L,0)}(L,0,\epsilon) = 0$, which contradicts $1 \in \alpha(L,0)$ by Condition (iii).

There are now two cases. First, if $E_G[T] > 0$, we can use the following argument. $\lim_{\epsilon\to 0} \overline{V}_{m(L,0)}(L,1,\epsilon) \geq 0$ implies that $\lim_{\epsilon\to 0} \overline{V}_{m(L,1)}(L,1,\epsilon) > 0$ because $m(L,1) > m(L,0)$ and $x \mapsto V_x(L,1)$ is strictly increasing by Lemma 4 (note that we have used the fact that $E_G[T] > 0$ to apply Lemma 4 here). Because $\epsilon \mapsto \overline{V}_{m(L,1)}(L,1,\epsilon)$ is continuous, it follows that $\overline{V}_{m(L,1)}(L,1,\epsilon) > 0$ for all $\epsilon > 0$ sufficiently small. But then $\alpha(L,1) = 1$ by Condition (iii), a contradiction.

Second, if $E_G[T] = 0$, then we can employ a similar argument using the fact that, by part (ii) of lemma 7, $\left.\frac{\partial W_\epsilon(p(L,x))}{\partial \epsilon}\right|_{\epsilon=0}$ is strictly increasing in x and that, by Lemma 7, we have

$$\lim_{\epsilon\to 0} \frac{\overline{V}_x(L,1,\epsilon)}{E_{G_\epsilon}[T]} = \lim_{\epsilon\to 0} \frac{W_\epsilon(p(L,x))}{\epsilon} = \left.\frac{\partial W_\epsilon(p(L,x))}{\partial \epsilon}\right|_{\epsilon=0}$$

■

Proof of Proposition 1. We first argue that if $\inf_{t\geq 0} V(p_t(m_t)) \geq \frac{r}{\gamma}$, then experimenting forever is an equilibrium if the organization is experimenting at time $t = 0$.

Consider the following strategy profile: $\alpha(L,1) = 1$ for all $L \in \{0\} \cup [1,\infty)$, $\alpha(L,0) = 1$ if $V(p(L,m(L,0))) > \frac{r}{\gamma}$, and $\beta(x,L,\pi)$ is given by (i) in the definition of the equilibrium. The path of play is as follows. If the organization is in state $(L,1)$ at time t_0, then $\pi_t = 1$ for all $t > t_0$. If the organization is in state $(L,0)$ at time t_0, then $t_1 = \inf\left\{t \geq t_0 : V(p(L_t, m(L_t,0))) > \frac{r}{\gamma}\right\}$. It follows that $\pi_t = 1$ for all $t \geq t_1$ and $\pi_t = 0$ for $t < t_1$. We can check that Conditions (a)-(d) and (i)-(iii) hold, so this is an equilibrium.

Next, we argue that if $V(p_t(m_t)) > \frac{r}{\gamma}$ for all $t \geq 0$, then any equilibrium must be of this form. Let σ be an equilibrium. By Lemma 11, the risky policy must always be used after a success. By Lemma 13, after a switch from the risky policy to the safe policy, the safe policy will be used forever.

Now suppose for the sake of contradiction that σ is *not* such that, starting with policy 1, policy 1 is used forever with probability one. In other words, suppose that there is L for which $0 \in \alpha(L,1)$. By Condition (ii) in the definition of the equilibrium, this requires that $\frac{r}{\gamma} \geq V_{m(L,1)}(L,1)$.

Let T be the (possibly random) time until the policy first switches to 0, starting in state $(L,1)$. Then $V_{m(L,1)}(L,1) = E[W_T(p(L,m(L,1)))]$, where $W_T(x)$ is as in Lemma 7 and the expectation is taken over T.

Recall that $V(p(L,m(L,1))) > \frac{r}{\gamma}$ by assumption. By Lemma 7, this implies that $W_T(p(L,m(L,1)))) > \frac{r}{\gamma}$ for all $T > 0$. If $E[T] > 0$, it follows that $E[W_T(p(L,m(L,1)))] > \frac{r}{\gamma}$, implying that $V_{m(L,1)}(L,1) > \frac{r}{\gamma}$. This implies that $\alpha(L,1) = 1$ by Condition (ii), which is a contradiction. If $E[T] = 0$, then $V_{m(L,1)}(L,1) = \frac{r}{\gamma}$ but, by the same argument, $\overline{V}_{m(L,1)}(L,1,\epsilon) > \frac{r}{\gamma} = \overline{V}_{m(L,1)}(L,0,\epsilon)$ for all $\epsilon > 0$ and thus $\alpha(L,1) = 1$ by Condition (iii), which is a contradiction.

Finally, suppose that $\inf_{t\geq 0} V(p_t(m_t)) < \frac{r}{\gamma}$, so that $V(p_{t_0}(m_{t_0})) < \frac{r}{\gamma}$ for some t_0, and suppose that there is an equilibrium in which $\pi_t = 1$ for all t starting in state $(1,1)$. This requires that $1 \in \alpha(L_{t_0},1)$, which implies that $V(p_{t_0}(m_{t_0})) \geq \frac{r}{\gamma}$ by Condition (ii), a contradiction. ■

Proposition 9. *The value function V in Proposition 1 satisfies the following:*

(i) If f is non-decreasing, then

$$\gamma \inf_{t\geq 0} V(p_t(m_t)) = \gamma V\left(\frac{2s}{b+s}\right) = \frac{2bs}{b+s} + \left(\frac{1}{2}\right)^{\frac{\gamma}{b}} \frac{s(b-s)}{b+s}\frac{b}{\gamma+b}$$

(ii) Given $\alpha > 0$, let $f_\alpha(x)$ denote a density with support $[0,1]$ such that $f_\alpha(x) = (\alpha+1)(1-x)^\alpha$ for $x \in [0,1]$. Let f be a density with support $[0,1]$ that MLRP-dominates f_α. Let $\lambda = \frac{1}{2^{\frac{1}{\alpha+1}}}$. Then

$$\gamma \inf_{t\geq 0} V(p_t(m_t)) \geq \gamma V\left(\frac{s}{\lambda b + (1-\lambda)s}\right) = \frac{bs}{\lambda b + (1-\lambda)s} + \lambda^{\frac{\gamma+b}{b}} \frac{s(b-s)}{\lambda b + (1-\lambda)s}\frac{b}{\gamma+b}$$

(iii) Let f be any density with support $[0,1]$. Then

$$\gamma \inf_{t\geq 0} V(p_t(m_t)) \geq \gamma V\left(\frac{s}{b}\right) = s + \frac{s(b-s)}{\gamma+b}$$

Proof of Proposition 9.

We prove each inequality in three steps.

First, we show that the median posterior belief is uniformly bounded below for all t, with different bounds depending on the density. When f is uniform, we have $p_t(m_t) \geq \frac{2s}{b+s}$. For any $\alpha > 0$, when $f = f_\alpha$, we have $p_t(m_t) \geq \frac{s}{\lambda b+(1-\lambda)s}$ for $\lambda = \frac{1}{2^{\frac{1}{\alpha+1}}}$. Finally, if f has full support, we have $p_t(m_t) \geq \frac{s}{b}$. The first two claims follow from Lemma 1, which shows that $p_t(m_t) \searrow \frac{2s}{b+s}$ as $t \to \infty$ when f is uniform, and $p_t(m_t) \searrow \frac{s}{\lambda b+(1-\lambda)s}$ when $f = f_\alpha$. The last claim is implied by the fact that $m_t \geq y_t$ and $p_t(y_t) = \frac{s}{b}$.

Second, we argue that these bounds hold not just for the aforementioned densities but also for any that dominate them in the MLRP sense. This follows from Lemma 8 and the fact that the function $x \mapsto p_t(x)$ is strictly increasing.

Third, we observe that, since $V(x)$ is strictly increasing and continuous in x (by Lemmas 2 and 4), we have $\inf_{t\geq 0} V(p_t(m_t)) = V(\inf_{t\geq 0} p_t(m_t))$. Hence, to arrive at the bounds in the Proposition, it is enough to evaluate V at the appropriate beliefs.

To calculate $V\left(\frac{s}{\lambda b+(1-\lambda)s}\right)$, we use Lemmas 2 and 3. The time it takes for an agent with belief $\frac{s}{\lambda b+(1-\lambda)s}$ to reach posterior $\frac{s}{b}$ is

$$t\left(\frac{s}{\lambda b+(1-\lambda)s}\right) = -\frac{\ln\left(\frac{s}{b-s}\frac{\lambda(b-s)}{s}\right)}{b} = -\frac{\ln \lambda}{b}$$

Thus, taking $x = \frac{s}{\lambda b+(1-\lambda)s}$, we have $e^{-bt(x)} = \lambda$ and $e^{-\gamma t(x)} = \lambda^{\frac{\gamma}{b}}$. Substituting this value of x into the formula for $V(x)$ from Lemma 2, we obtain

$$\gamma V\left(\frac{s}{\lambda b+(1-\lambda)s}\right) = \frac{bs}{\lambda b+(1-\lambda)s} + \frac{(b-s)s}{\lambda b+(1-\lambda)s}\frac{b}{\gamma+b}\lambda^{\frac{\gamma+b}{b}}$$

In particular, for $\alpha = 0$, this becomes

$$\gamma V\left(\frac{2s}{b+s}\right) = \frac{2bs}{b+s} + \left(\frac{1}{2}\right)^{\frac{\gamma}{b}}\frac{(b-s)s}{b+s}\frac{b}{\gamma+b}$$

On the other hand, for $x = \frac{s}{b}$, we have $t(x) = 0$. Substituting this in, we obtain

$$\gamma V\left(\frac{s}{b}\right) = s + \frac{b-s}{b}s - \frac{s}{b}(b-s)\frac{\gamma}{\gamma+b} = s + \frac{(b-s)s}{\gamma+b}$$

An additional argument is required to show that the bound is tight in part (i).

Take f to be any non-decreasing density. Let $\tilde{m}_t$ denote the median at time t under f, and let m_t denote the median at time t under the uniform density. It is sufficient to show that the asymptotic posterior of the median is $\frac{2s}{b+s}$ under f, that is, that $\lim_{t\to\infty} p_t(\tilde{m}_t) = \lim_{t\to\infty} p_t(m_t) = \frac{2s}{b+s}$.

Lemma 9 shows that $\frac{1-\tilde{m}_t}{1-m_t} \to 1$. Note that we have $\tilde{m}_t \geq m_t$ for all t by Lemma 8 and $m_t \to 1$ as $t \to \infty$. Then Lemma 10 applies. Lemma 10 applied to the sequences $\tilde{m}_t$ and m_t guarantees that the ratio $\frac{p_t(\tilde{m}_t)}{p_t(m_t)}$ converges to 1. ■

Proof of Proposition 2. First note that when $f = f_\alpha$, the fact that $\inf_t V(p_t(m_t))$ is increasing in b follows from the formula for $V\left(\frac{s}{\lambda b+(1-\lambda)s}\right)$ given in the proof of Proposition 9. Lemma 8 implies that an MLRP-increase in f increases $\inf_t V(p_t(m_t))$. As for changes in s, note that an increase in s clearly increases $V(x)$ for each x, and also increases y_t, and hence m_t, for each t. Finally, note that a decrease in γ is equivalent to an increase in the learning rate.[51] This leaves y_t and m_t unchanged but increases $V(x)$ for all x, as agents have strictly more information to base their entry and exit decisions on. ■

Lemma 14. *There exist parameters such that* $V\left(\frac{2s}{b+s}\right) \geq \frac{r}{\gamma}$.

Proof of Lemma 14. It is easy to show that there exist parameters such that $\frac{2bs}{b+s} + \left(\frac{1}{2}\right)^{\frac{\gamma}{b}} \frac{s(b-s)}{b+s}\frac{b}{\gamma+b} \geq r$ is satisfied. For example, suppose that $\gamma \approx \infty$ (agents are infinitely impatient). Then we need that $\frac{2bs}{b+s} \geq r$. For this, it is sufficient to have $s > \frac{r}{2}$ and $b \geq \frac{sr}{2s-r}$. ■

Define a *stopping function* $\tau : [0,\infty) \to [0,\infty]$ as follows. For each $t \geq 0$, $\tau(t) \geq t$ is such that m_t is indifferent about switching to the safe policy at time t if she expects a continuation where experimentation will stop at time $\tau(t)$ should she fail to stop at t. If the agent never wants to experiment regardless of the expected continuation, then $\tau(t) = t$, while if she always does, then $\tau(t) = \infty$.

Proposition 10. *Any pure strategy equilibrium σ in which the organization does not*

[51] In other words, it is equivalent to increasing the success rate of the good risky policy to bq, for $q > 1$, and lowering the payoff per success to $\frac{1}{q}$.

experiment forever is given by a sequence of stopping times $t_0(\sigma) \leq t_1(\sigma) \leq t_2(\sigma) \leq \dots$ *such that* $t_n(\sigma) = \tau(t_{n-1}(\sigma))$ *for all* $n > 0$ *and* $t_0(\sigma) \leq \tau(0)$.

There exists $t \in [0, \tau(0)]$ *for which* $(t, \tau(t), \tau(\tau(t)), \dots)$ *constitutes an equilibrium. Moreover, if* τ *is weakly increasing, then* $(t, \tau(t), \tau(\tau(t)), \dots)$ *constitutes an equilibrium for all* $t \in [0, \tau(0)]$.

Proof of Proposition 10. We first argue that the stopping function τ is well-defined.

Let t be the current time and let t^* be the time at which m_t would choose to stop experimenting if she had complete control over the policy. Recall the definition of $W_{T-t}(x)$ from Lemma 7: $W_{T-t}(x)$ is the value function starting at time t of an agent with belief x at time t given a continuation equilibrium path on which the organization experiments until T and then switches to the safe technology. Then, equivalently, $t^* = \operatorname{argmax}_T W_{T-t}(x)$.

There are three cases. If $t^* = t$, then $\tau(t) = t$. If $t^* > t$, that is, if x wants to experiment for a positive amount of time, and $V(p_t(m_t)) < \frac{r}{\gamma}$, then $W_{T-t}(p_t(m_t))$ is strictly increasing in T for $T \in [t, t^*]$ and strictly decreasing in T for $T > t^*$, as shown in Lemma 7, and there is a unique $\tau(t) > t^*$ for which $W_{\tau(t)-t}(p_t(m_t)) = \frac{r}{\gamma}$. Finally, if $t^* > t$ and $V(p_t(m_t)) \geq \frac{r}{\gamma}$, then $\tau(t) = \infty$.

Next, note that τ is continuous. If $\tau(t_0) \in (t_0, \infty)$, then for t in a neighborhood of t_0, $\tau(t)$ is defined by the condition $W_{\tau(t)-t}(p_t(m_t)) = \frac{r}{\gamma}$, where $p_t(m_t)$ is continuous in t, and $W_T(x)$ is differentiable in (T, x) at $(T, x) = (\tau(t), p_t(m_t))$ and strictly decreasing in T,[52] so the continuity of τ follows from the Implicit Function Theorem. The proofs for the cases when $\tau(t_0) = 0$ or $\tau(t_0) = \infty$ are similar.

Consider a pure strategy equilibrium σ in which the organization does not experiment forever on the equilibrium path. Let $t_0(\sigma)$ be the time at which experimentation stops on the equilibrium path. Clearly, we have $t_0(\sigma) \leq \tau(0)$, as otherwise m_0 would switch to the safe policy at time 0. As before, if a success occurs or if the organization switches to the safe policy, everyone joins the organization permanently.

Consider what happens at time $t_0(\sigma)$ if $m_{t_0(\sigma)}$ deviates and continues experi-

[52]The fact that $W_T(x)$ is differentiable in T at $(T, x) = (\tau(t), p_t(m_t))$ and strictly decreasing in T is implied by Lemma 7.

menting. Suppose first that $\tau(t_0(\sigma)) \in (t_0(\sigma), \infty)$. In a pure strategy equilibrium, there must be a time $t_1(\sigma) \geq t_0(\sigma)$ for which experimentation stops in this continuation, and it must satisfy $t_1(\sigma) = \tau(t_0(\sigma))$. To see why, suppose that $t_1(\sigma) > \tau(t_0(\sigma))$. In this case, for $\epsilon > 0$ sufficiently small, $m_{t_0(\sigma)+\epsilon}$ would strictly prefer to stop experimenting, a contradiction. On the other hand, if $t_1(\sigma) < \tau(t_0(\sigma))$, then $m_{\tau_0(\sigma)}$ would strictly prefer to deviate from the equilibrium path and not stop.

Next, suppose that $\tau(t_0(\sigma)) = \infty$, that is, $m_{t_0(\sigma)}$ weakly prefers to continue experimenting regardless of the continuation. Then it must be that $t_1(\sigma) = \infty$ and $V\left(p_{t_0(\sigma)}\left(m_{t_0(\sigma)}\right)\right) = \frac{r}{\gamma}$, and in this case we must still have $t_1(\sigma) = \tau(t_0(\sigma))$.

Now suppose that $\tau(t_0(\sigma)) = t_0(\sigma)$, that is, $m_{t_0(\sigma)}$ weakly prefers to stop regardless of the continuation. In this case, the implied sequence of stopping points is $(t_0(\sigma), t_0(\sigma), \ldots)$. This does not fully describe the equilibrium, as it does not specify what happens conditional on not stopping experimentation by $t_0(\sigma)$, but still provides enough information to characterize the equilibrium path fully, as in any equilibrium experimentation must stop at $t_0(\sigma)$.

Next, we show that if τ is increasing and $t \in [0, \tau(0)]$, then $(t, \tau(t), \tau(\tau(t)), \ldots)$ constitutes an equilibrium. Our construction already shows that $m_{t_n(\sigma)}$ is indifferent between switching to the safe policy at time $t_n(\sigma)$ and continuing to experiment. To finish the proof, we have to show that for t not in the sequence of the stopping times, m_t weakly prefers to continue experimenting. Fix $t \in (t_n(\sigma), t_{n+1}(\sigma))$. Since $t > t_n(\sigma)$ and τ is increasing, we have $\tau(t) \geq \tau(t_n(\sigma)) = t_{n+1}(\sigma)$. Hence the definition of $\tau(t)$ and the fact that $T \mapsto W_T(x)$ is single-peaked by Lemma 7 imply that $W_{t_{n+1}(\sigma)-t}(p_t(m_t)) \geq \frac{r}{\gamma}$, and Conditions (ii) and (iii) imply that m_t weakly prefers to continue experimenting.

Finally, we show that even if τ is not increasing, this construction yields an equilibrium for at least one value of $t \in [0, \tau(0)]$. Note that our construction fails if and only if there is $t \in (t_k(\sigma), t_{k+1}(\sigma))$ for which $\tau(t) < t_{k+1}(\sigma)$. Motivated by this, we say t is *valid* if $\tau(t) = \inf_{t' \geq t} \tau(t')$, and say t is *n-valid* if $t, \tau(t), \ldots, \tau^{(n-1)}(t)$ are all valid. Let $A_0 = [0, \tau(0)]$ and, for $n \geq 1$, let $A_n = \{t \in [0, \tau(0)] : t \text{ is } n\text{-valid}\}$.

Suppose that $\tau(t) > t$ and $\tau(t) < \infty$ for all t. Clearly, $A_n \supseteq A_{n+1}$ for all n, and the continuity of τ implies that A_n is closed for all n. In addition, A_n must be non-empty for all n by the following argument. Take $t_0 = t$ and define a sequence

$\{t_0, t_{-1}, t_{-2}, \ldots, t_{-k}\}$ by $t_{-i} = \max\{\tau^{-1}(t_{-i+1})\}$ for $i \leq -1$, and $t_{-k} \in [0, \tau(0)]$. By construction, $t_{-k} \in A_0$ is k-valid, and, because $\tau(t) < \infty$ for all t, if we choose t large enough, we can make k arbitrarily large.[53] Then $A = \cap_0^\infty A_n \neq \varnothing$ by Cantor's intersection theorem, and any sequence $(t, \tau(t), \ldots)$ with $t \in A$ yields an equilibrium. The same argument goes through if $\tau(t) = \infty$ for some values of t but there are arbitrarily large t for which $\tau(t) < \infty$.

If $\tau(t) = t$ for some t, let $\bar{t} = \min\{t \geq 0 : \tau(t) = t\}$. If there is $\epsilon > 0$ such that $\tau(t) \geq \tau(\bar{t})$ for all $t \in (\bar{t} - \epsilon, \bar{t})$, then we can find a finite equilibrium sequence of stopping times by setting $t_0 = \bar{t}$ and using the construction in the previous paragraph. If there is no such ϵ, then the previous argument works.[54] The only difference is that, to show the non-emptiness of A_n, we take $t \to \bar{t}$ instead of making t arbitrarily large.

If $\tau > t$ for all t and there is $\tilde{t}$ for which $\tau(t) = \infty$ for all $t \geq \tilde{t}$, without loss of generality, take $\tilde{t}$ to be minimal (that is, let $\tilde{t} = \min\{t \geq 0 : \tau(t) = \infty\}$). Then we can find a finite sequence of stopping times compatible with equilibrium by taking $t_0 = \tilde{t}$, assuming that m_{t_0} stops at t_0 and using the above construction. ■

Proof of Proposition 3. Parts 1, 2 and 3 follow from Proposition 10. Part 4 follows from an analogous proof to Proposition 1. ■

A A Model of Bad News

Lemma 15. *In a model of bad news, the value function of an agent with prior x who is in the organization and expects the organization to continue forever unless a failure is observed is*

$$V(x) = (xb + (1-x)r)\frac{1}{\gamma} - (1-x)r\frac{1}{\gamma + b}$$

Proof of lemma 15. Note that an agent receives an expected flow payoff of b only if the technology is good and the organization has not switched to the safe technology upon observing a failure. Because a good technology cannot experience a

[53]Under the assumption that $\tau(t) < \infty$ for all t, since τ is continuous, the image of τ^l restricted to the set $[0, \tau(0)]$ is compact and hence bounded for all l. Thus for any t larger than the supremum of this image, k must be larger than l.

[54]If there is $\epsilon > 0$ with the aforementioned property, then $\tau^{-1}(\bar{t})$ is strictly lower than $\bar{t}$ and reaching $[0, \tau(0)]$ takes finitely many steps. If there is no such ϵ, then $\tau^{-1}(\bar{t}) = \bar{t}$ and there exists a sequence converging to $\bar{t}$.

failure, as long as experimentation continues, an agent with posterior belief x receives an expected flow payoff of b with probability x.

Let $P_t = x + (1-x)e^{-bt}$ denote the probability that an agent with prior belief x assigns to not having a failure by time t. Then

$$\begin{aligned} V(x) &= \int_0^\infty (xb + (1-P_\tau)r)\, e^{-\gamma\tau} d\tau = \int_0^\infty \left(xb + (1-x)(1-e^{-b\tau})r\right) e^{-\gamma\tau} d\tau \\ &= (xb + (1-x)r)\frac{1}{\gamma} - (1-x)r\frac{1}{\gamma+b} \end{aligned}$$

■

Assumption 1. *The parameters b, r, s, γ, f are such that for all $t' > t$, $\frac{\partial}{\partial t} W_{t'-t}(p_t(m_t)) \neq 0$ whenever $W_{t'-t}(p_t(m_t)) = \frac{r}{\gamma}$.*

Assumption 1 guarantees that the agents' value functions are well-behaved: that is, for each t', the function $t \mapsto W_{t'-t}(p_t(m_t))$ crosses the threshold $\frac{r}{\gamma}$ finitely many times, and is never tangent to it. Under this assumption, Proposition 6 characterizes the equilibrium in the bad news model.

Proposition 11. *Under Assumption 1, there is a unique equilibrium. The equilibrium can be described by a finite, possibly empty set of stopping intervals $I_0 = [t_0, t_1]$, $I_1 = [t_2, t_3]$, $\ldots$, I_n such that $t_0 < t_1 < t_2 < \ldots$ as follows: conditional on the risky policy having been used during $[0, t]$ with no failures, the median m_t switches to the safe policy at time t if and only if $t \in I_k$ for some k.*

Proof of Proposition 11. We first argue that there exists T such that for all $t \geq T$, if no failures have been observed during $[0,t]$, then $V(p_t(m_t)) > \frac{r}{\gamma}$ and $p_t(m_t))b > r$. Note that, because in a model of bad news agents do not leave the organization, we have $\liminf_{t\to\infty} m_t > 0$. Moreover, $\lim_{t\to\infty} e^{-bt} = 0$. This implies that $\lim_{t\to\infty} p_t(m_t) = \lim_{t\to\infty} \frac{m_t}{m_t + e^{-bt}(1-m_t)} = 1$, so $\lim_{t\to\infty} p_t(m_t)b = b > r$. Provided that no failures have been observed during $[0,t]$, we have $\lim_{t\to\infty} V(p_t(m_t)) = V(1)$ because V is continuous, and $V(1) = \frac{b}{\gamma} > \frac{r}{\gamma}$. Next, we argue that these agents will always experiment.

Claim 1. If $p_t(m_t)b > r$, then in any equilibrium m_t continues experimenting.

Proof of claim 1. Suppose for the sake of contradiction that this is not the case. Let $t+t_+$ denote the first time after t when the equilibrium prescribes a switch to the safe policy.[55] Because experimentation will stop after a period of length t_+, the payoff to m_t from experimenting is

$$\begin{aligned} W_{t_+}(p_t(m_t)) &= \int_0^{t_+} (p_t(m_t)b + (1-P_\tau)r)e^{-\gamma\tau}d\tau + \int_{t_+}^{\infty} re^{-\gamma\tau}d\tau \\ &\geq \int_0^{t_+} p_t(m_t)be^{-\gamma\tau}d\tau + \int_{t_+}^{\infty} re^{-\gamma\tau}d\tau = p_t(m_t)b\frac{1-e^{-\gamma t_+}}{\gamma} + r\frac{e^{-\gamma t_+}}{\gamma} \end{aligned}$$

The payoff to stopping experimentation is $\frac{r}{\gamma}$. Then, since $p_t(m_t)b > r$ by assumption, we have $\frac{1-e^{-\gamma t_+}}{\gamma}p_t(m_t)b + \frac{e^{-\gamma t_+}}{\gamma}r > \frac{r}{\gamma}$, so m_t strictly prefers to continue experimenting, a contradiction. ■

Our results so far are already enough to deal with one important case. If $V(p_t(m_t)) > \frac{r}{\gamma}$ for all t, then the organization experiments forever. The reason is as follows. For $t \geq T$, all pivotal agents m_t continue experimenting by Claim 1. Let $\mathcal{T} \subseteq [0,T)$ be the set of times for which the pivotal agent at that time stops experimenting in equilibrium, and assume $\mathcal{T}$ is nonempty. Let $t^* = \sup \mathcal{T}$. If $t^* \in \mathcal{T}$, then m_{t^*} stops experimenting even though $V(p_t(m_t)) > \frac{r}{\gamma}$ and m_{t^*} gets perpetual experimentation by continuing, a contradiction. If $t^* \notin \mathcal{T}$, a similar argument can be made leveraging Condition (iii).

Suppose then that there exists $t \leq T$ such that $V(p_t(m_t)) < \frac{r}{\gamma}$.

Claim 2. Suppose that on the equilibrium path, the organization continues experimenting for time t_+ unless a failure occurs and then switches to the safe policy. Then the value function of an agent with prior x in this equilibrium is given by

$$(xb + (1-x)r)\frac{1-e^{-\gamma t_+}}{\gamma} - (1-x)r\frac{1-e^{-(\gamma+b)t_+}}{\gamma+b} + e^{-\gamma t_+}\frac{r}{\gamma}$$

Proof of claim 2. Because experimentation ends after a period of length t_+,

[55] We write the argument assuming that $t_+ > 0$. If $t_+ = 0$, the proof follows a similar argument leveraging Condition (iii).

the value function of agent x is given by

$$\int_0^{t_+} (xb + (1 - P_\tau)r)\, e^{-\gamma\tau} d\tau + \int_{t_+}^{\infty} re^{-\gamma\tau} d\tau$$
$$= (xb + (1 - x)r)\frac{1 - e^{-\gamma t_+}}{\gamma} - (1 - x)r\frac{1 - e^{-(\gamma+b)t_+}}{\gamma + b} + e^{-\gamma t_+}\frac{r}{\gamma}$$

■

Claim 3. Suppose that in some equilibrium m_{t_0} stops experimenting. If for all $t \in [\underline{t}, t_0)$ we have $p_t(m_t)b < r$, then for all $t \in [\underline{t}, t_0)$, m_t stops experimenting.

Proof of claim 3. Suppose for the sake of contradiction that this is not the case. Then there exists a non-empty subset $B \subseteq [\underline{t}, t_0)$ such that for all $t \in B$, m_t continues experimenting.

There are two cases. In the first case, B has a non-empty interior. In this case, for all $\epsilon > 0$ small, there must exist $\tau \in [\underline{t}, t_0)$ such that, starting at time τ, experimentation continues up to time $\tau + \epsilon$ and then stops.[56]

By claim 2, the payoff to m_τ from continuing experimentation is $W_\epsilon(p_\tau(m_\tau)) = (p_\tau(m_\tau)b + (1 - p_\tau(m_\tau))r)\frac{1-e^{-\gamma\epsilon}}{\gamma} - (1 - p_\tau(m_\tau))r\frac{1-e^{-(\gamma+b)\epsilon}}{\gamma+b} + e^{-\gamma\epsilon}\frac{r}{\gamma}$, which is of the form $\frac{r}{\gamma} + (p_\tau(m_\tau))b - r)\epsilon + \mathcal{O}(\epsilon^2)$. The payoff to stopping experimentation is $\frac{r}{\gamma}$. Then, since $p_\tau(m_\tau)b < r$ by assumption, for ϵ small enough m_τ strictly prefers to stop experimenting, a contradiction. In the second case, the interior of B is empty. In this case, the proof follows a similar argument leveraging Condition (iii). ■

Let $t_{2n+1} = \sup\left\{t : V(p_t(m_t)) < \frac{r}{\gamma}\right\}$ denote the largest time for which the median stops experimenting.

Let $T_1 = \{t : p_t(m_t)b \leq r\}$ and $T_2 = \{t : p_t(m_t)b > r\}$. Our genericity assumption (Assumption 1) implies that T_1 and T_2 are finite collections of intervals. Enumerate the intervals such that $T_1 = \cup_{i=0}^{n}[\underline{t}_i, \bar{t}_i]$.

Suppose first that $p_t(m_t)b \leq r$ for all $t < t_{2n+1}$. In this case, by claim 3, for all $t \leq t_{2n+1}$, m_t stops experimentation. Then we set $n = 0$, $t_0 = 0$ and $I_0 = [t_0, t_1]$.

Suppose next that there exists $t < t_{2n+1}$ such that $p_t(m_t)b > r$. Set $t_{2n} =$

[56] To find such τ, let $\tilde{t}$ be in the interior of B, and let $\tilde{\tilde{t}} = \inf\{t \geq \tilde{t} : t \notin B\}$. Then $\tau = \tilde{\tilde{t}} - \epsilon$ works for all $\epsilon > 0$ small enough.

$\sup\{t < t_{2n+1} : p_t(m_t)b > r\}$. Note that, because F admits a continuous density, $t \mapsto p_t(m_t)$ is continuous, which implies that we must have $p_{t_{2n}}(m_{t_{2n}})b - r = 0$. Then claim 3 implies that for all $t \in [t_{2n}, t_{2n+1}]$, m_t stops experimentation.

Let us conjecture a continuation equilibrium path on which, starting at t, the organization experiments until t_{2n}. Recall that $W_{t_{2n}-t}(x)$ denotes the value function of an agent with belief x (at time t) given this continuation equilibrium path. We then let $t_{2n-1} = \sup\left\{t < t_{2n} : W_{t_{2n}-t}(p_t(m_t)) \leq \frac{r}{\gamma}\right\}$.

Note that, because, by construction, for $t \in (t_{2n-1}, t_{2n})$ we have $W_{t_{2n}-t}(p_t(m_t)) > \frac{r}{\gamma}$, the median m_t continues experimentation for all $t \in (t_{2n-1}, t_{2n})$.

Since F admits a continuous density, $t \mapsto W_{t_{2n}-t}(p_t(m_t))$ is continuous, which implies that we must have $t_{2n-1} = \max\left\{t < t_{2n} : W_{t_{2n}-t}(p_t(m_t)) \leq \frac{r}{\gamma}\right\}$. Note that it is then consistent with equilibrium for the median $m_{t_{2n}}$ to stop experimenting.

Now note that if $W_{t_{2n}-t_{2n-1}}(p_{t_{2n-1}}(m_{t_{2n-1}})) = \frac{r}{\gamma}$, then $p_{t_{2n-1}}(m_{t_{2n-1}})b < r$. By continuity, there exists an interval $[\underline{t}_i, \bar{t}_i]$ in T_1 such that $t_{2n-1} \in [\underline{t}_i, \bar{t}_i]$ (and $\underline{t}_i$ satisfies $\underline{t}_i = \min\{t < t_{2n-1} : p_t(m_t)b \leq r\}$).

Set $t_{2n-2} = \underline{t}_i$. Because $p_t(m_t)b \leq r$ for all $t \in [t_{2n-2}, t_{2n-1}]$, claim 1 implies that, for all $t \in [t_{2n-2}, t_{2n-1}]$, m_t stops experimenting.

We then proceed inductively in the above manner, finding the largest t strictly less than t_{2n-2} such that $W_{t_{2n-2}-t}(p_t(m_t)) \leq \frac{r}{\gamma}$. Because T_1 is finite collection of intervals, the induction terminates in a finite number of steps.

The equilibrium is generically unique for the following reason. Under Assumption 1, each t_{2k+1} satisfies not only $W_{t_{2k+2}}(p_{t_{2k+1}}(m_{t_{2k+1}})) = \frac{r}{\gamma}$ but also $\frac{\partial}{\partial t}W_{t_{2k+2}-t}(p_t(m_t))|_{t=t_{2k+1}} > 0$, that is, $W_{t_{2k+2}-t}(p_t(m_t)) < \frac{r}{\gamma}$ for all $t < t_{2k+1}$ close enough to t_{2k+1}. Thus, even if we allow $m_{t_{2k+1}}$ to continue experimenting, all agents in $(t_{2k+1} - \epsilon, t_{2k+1})$ must stop as they strictly prefer to do so. Likewise, each t_{2k} satisfies not only $p_{t_{2k}}(m_{t_{2k}})b - r = 0$ but also $\frac{\partial}{\partial t}p_t(m_t)|_{t=t_{2k}} < 0$, that is, $p_t(m_t)b - r > 0$ for all $t < t_{2k}$ close enough to t_{2k}. Thus, even if we allow $m_{t_{2k}}$ to stop experimenting, all agents in $(t_{2k} - \epsilon, t_{2k})$ must stop as they strictly prefer to do so. ■

Proof of Proposition 6. Follows from Proposition 11. ■

B Imperfectly Informative Experimentation

Lemma 16. *If the organization is experimenting at time t, then an agent with prior belief x is in the organization at time t if and only if $L(k,t) \leq \frac{x(b-s)}{(1-x)(s-c)}$.*

Proof of lemma 16. Because agents make their membership decisions based on expected flow payoffs, agent x is a member at time t iff $p(L,x)b+(1-p(L,x))c \geq s$, that is, if $p(L,x) \geq \frac{s-c}{b-c}$. Since $p(L,x) = \frac{x}{x+(1-x)L(k,t)}$, this is equivalent to $L(k,t) \leq \frac{x(b-s)}{(1-x)(s-c)}$. ∎

Lemma 17. *If the distribution of priors is power law, then $L \mapsto p(L, m(L))$ is decreasing. Moreover, if $L_0 m'(L_0) < m(L_0)(1-m(L_0))$, then $L \mapsto p(L, m(L))$ is strictly decreasing at $L = L_0$, and if $L_0 m'(L_0) > m(L_0)(1-m(L_0))$, then $L \mapsto p(L, m(L))$ is strictly increasing at $L = L_0$.*

Proof of lemma 17. The density of the power law distribution is given by $f(x) = (1-x)^\alpha c$ where c is a constant ensuring that the density integrates to 1. In particular, if the support of the distribution is $[0,1]$, then we have $F(z) = \int_0^z (1-x)^\alpha c dx = \frac{c}{\alpha+1}(1-(1-z)^{\alpha+1})$. Because $F(1) = 1$, we must have $c = \alpha+1$. Then $F(z) = 1-(1-z)^{\alpha+1}$ and the CDF of the distribution with support on $[y,1]$ is given by $\frac{(1-y)^{\alpha+1}-(1-z)^{\alpha+1}}{(1-y)^{\alpha+1}}$.

Recall that $m(L)$ and $y(L)$ denote the median and the marginal members of the organization respectively when the state variable is L. The above argument implies that the median must satisfy $\frac{(1-y(L))^{\alpha+1}-(1-m(L))^{\alpha+1}}{(1-y(L))^{\alpha+1}} = \frac{1}{2}$. Equivalently, we must have $(1-m(L))^{\alpha+1} = \frac{1}{2}(1-y(L))^{\alpha+1}$. Then the median must satisfy $1-m(L) = (1-y(L))2^{-\frac{1}{\alpha+1}}$, or $m(L) = 1-\kappa+\kappa y(L)$ for $\kappa = 2^{-\frac{1}{\alpha+1}}$.

Note that $p(L,m(L)) = \frac{1}{1+\left(\frac{1}{m(L)}-1\right)L}$. Then $\frac{\partial}{\partial L}p(L,m(L)) \propto -\frac{\partial}{\partial L}\left(1+\left(\frac{1}{m(L)}-1\right)L\right)$ and $\frac{\partial}{\partial L}\left(1+\left(\frac{1}{m(L)}-1\right)L\right) = \frac{\partial}{\partial L}\left(\left(\frac{1}{m(L)}-1\right)L\right) = \frac{1}{m(L)} - 1 - \frac{L}{(m(L))^2}m'(L)$.

This implies that if $L_0 m'(L_0) < m(L_0)(1-m(L_0))$, then $L \mapsto p(L,m(L))$ is strictly decreasing at $L = L_0$, and if $L_0 m'(L_0) > m(L_0)(1-m(L_0))$, then $L \mapsto p(L,m(L))$ is strictly increasing at $L = L_0$.

After some algebra, using the fact that $y(L) = \frac{s-c}{s-c+(b-s)\frac{1}{L}}$, we get that if the distribution of priors is power law, then $Lm'(L) < m(L)(1-m(L))$ is equivalent to $0 < (1-\kappa)(1-\zeta)$, where $\zeta = \frac{s-c}{b-c}$. Since κ and ζ are between 0 and 1, this always holds. ■

Lemma 18. *There exist distributions for which there exist states $L_1 < L_2$ such that L_1 is a unique minimizer of $p(L, m(L))$ and $L \mapsto p(L, m(L))$ is strictly increasing on (L_1, L_2).*

Proof of lemma 18. Consider a distribution with a density $f(x) = a_1$ for $x \in [0, b_1]$ and $f(x) = a_2$ for $x \in [b_1, 1]$. Note that we must have $a_1 b_1 + a_2(1 - b_1) = 1$ so that f integrates to 1. Define $\overline{b} = b - c$, $\overline{s} = s - c$, $y = y(L)$, $m = m(L)$, $z = p(L, m(L))$. Let L_1 be such that $m(L_1) = b_1$ and let L_2 be such that $y(L_2) = b_1$. Clearly, $0 < L_1 < L_2$. For $L > L_2$, $m(L)$ and $p(L, m(L))$ are the same as in the uniform case. In particular, $p(L, m(L)) = \frac{2L\overline{s}+\overline{b}-\overline{s}}{L(\overline{s}+\overline{b})+\overline{b}-\overline{s}}$, which is decreasing in L. Moreover, with the notation we have defined, the formula for $y(L)$ can be written as $y = \frac{L\overline{s}}{L\overline{s}+\overline{b}-\overline{s}}$.

For $L \in (L_1, L_2)$, we have $a_1(b_1 - y) + a_2(m - b_1) = a_2(1 - m)$, that is, $m = \frac{1+b_1}{2} - \frac{a_1 b_1}{2a_2} + \frac{a_1}{2a_2} y$. Equivalently, $m = \left(1 - \frac{1}{2a_2}\right) + \frac{a_1}{2a_2} y = \left(1 - \frac{1}{2a_2}\right) + \frac{a_1}{2a_2} \frac{L\overline{s}}{L\overline{s}+\overline{b}-\overline{s}}$. Then

$$\frac{1}{z} - 1 = \frac{L(1-m)}{m} = L \frac{L\frac{1-a_1}{2a_2}\overline{s} + \frac{1}{2a_2}\left(\overline{b}-\overline{s}\right)}{\left(1 - \frac{1}{2a_2} + \frac{a_1}{2a_2}\right)L\overline{s} + \left(1 - \frac{1}{2a_2}\right)\left(\overline{b}-\overline{s}\right)}$$

For $L < L_1$, we have $a_1(m - y) = a_1(b_1 - m) + a_2(1 - b_1)$, that is, $2a_1 m = a_1 b_1 + a_2(1 - b_1) + a_1 y = 1 + a_1 y$, so $m = \frac{1}{2a_1} + \frac{1}{2} y$, and

$$\frac{1}{z} - 1 = \frac{L(1-m)}{m} = L \frac{L\left(\frac{1}{2} - \frac{1}{2a_1}\right)\overline{s} + \left(1 - \frac{1}{2a_1}\right)\left(\overline{b}-\overline{s}\right)}{\left(\frac{1}{2a_1} + \frac{1}{2}\right)L\overline{s} + \frac{1}{2a_1}\left(\overline{b}-\overline{s}\right)}.$$

Now take $a_2 = \frac{1}{2}$ and any $a_1 > 1$ (note that choosing both pins down $b_1 = \frac{1}{2a_1 - 1}$). Then we can verify that $L \mapsto \frac{1}{p(L,m(L))} - 1$ is increasing on $(0, L_1)$ and decreasing on (L_1, L_2). In other words, $L \mapsto p(L, m(L))$ is decreasing on $(0, L_1)$ and (L_2, ∞) but increasing on (L_1, L_2), so L_1 is a local minimizer for $p(L, m(L))$.

Moreover, we can verify that under some extra conditions L_1 is a global minimizer: note that $\lim_{L\to\infty} \frac{1}{p(L,m(L))} - 1 = \frac{\overline{b}-\overline{s}}{2\overline{s}}$, while $\frac{1}{p(L_1,m(L_1))} - 1 = \frac{L_1(1-a_1)\overline{s}+\overline{b}-\overline{s}}{a_1\overline{s}}$.

Since $m(L_1) = b_1$, we have

$$\frac{1}{p(L_1, m(L_1))} - 1 = \frac{L_1}{m(L_1)} - L_1 = \frac{L_1}{b_1} - L_1 = \frac{L_1(1-a_1)\bar{s} + \bar{b} - \bar{s}}{a_1\bar{s}}$$

$$L_1 = \frac{\bar{b} - \bar{s}}{\bar{s}\left(\frac{a_1}{b_1} - 1\right)}$$

$$\begin{aligned}\frac{1}{p(L_1, m(L_1))} - 1 &= \frac{L_1}{b_1} - L_1 = \frac{\bar{b} - \bar{s}}{\bar{s}} \frac{\frac{1}{b_1} - 1}{\frac{a_1}{b_1} - 1} = \frac{\bar{b} - \bar{s}}{\bar{s}} \frac{1 - b_1}{a_1 - b_1} \\ &= \frac{\bar{b} - \bar{s}}{\bar{s}} \frac{2a_1 - 2}{2a_1^2 - a_1 - 1} = \frac{\bar{b} - \bar{s}}{\bar{s}} \frac{1}{a_1 + \frac{1}{2}}\end{aligned}$$

so L_1 is a global minimizer if we take $a_1 \in \left(1, \frac{3}{2}\right)$. ■

Proposition 12. *If $V_{m(L)}(L) > \frac{r}{\gamma}$ for all L, then there is a unique equilibrium. In it, the organization experiments forever. Moreover, if f is non-decreasing, then $V_{m(L)}(L) \geq V\left(\frac{2(s-c)}{(b-c)+(s-c)}\right) \geq \frac{1}{\gamma}\frac{(b-c)s+(s-c)b}{(b-c)+(s-c)}$, so there exist parameter values such that $V_{m(L)}(L) > \frac{r}{\gamma}$ for all L.*

Proof of Proposition 12. The proof is similar to the proof for the baseline model (Propositions 1 and 9). If $V_{m(L)}(L) > \frac{r}{\gamma}$ for all L, perpetual experimentation is clearly an equilibrium, as each pivotal agent $m(L)$ has a choice between $V_{m(L)}(L)$ and $\frac{r}{\gamma}$, and strictly prefers the former. The equilibrium is unique by the following argument. Suppose for the sake of contradiction that there is another equilibrium in which experimentation stops whenever $L \in \mathcal{L} \neq \varnothing$. Let $W_{\mathcal{L}}(x)$ denote the continuation utility of an agent with current belief x when she expects the organization to stop whenever $L \in \mathcal{L}$.

For L close enough to 0, it can be shown that pivotal agents will prefer to experiment no matter what equilibrium continuation they expect. That is, $W_{\mathcal{L}}(x) \geq \frac{r}{\gamma}$ for all $\mathcal{L}$ and x close enough to 1. In other words, there is $L_0 > 0$ such that $\mathcal{L} \subseteq (L_0, +\infty)$.

Let $L_1 = \inf \mathcal{L}$. It can be shown that, because $m(L_1)$ would rather experiment forever than not at all, she would also prefer to experiment until L hits $\mathcal{L}$. That is, if $V_{m(L_1)}(L_1) > \frac{r}{\gamma}$ then $W_{\mathcal{L}}(p(L_1, m(L_1))) > \frac{r}{\gamma}$. To see why, suppose that $W_{\mathcal{L}}(p(L_1, m(L_1))) \leq \frac{r}{\gamma}$. Clearly, this implies $W_{\mathcal{L}}(p(L, m(L_1))) < \frac{r}{\gamma}$ for any $L > L_1$.

Note that changing the set of stopping states from $\mathcal{L}$ to $\varnothing$ changes the payoff $m(L_1)$ gets from some continuations—namely, continuation starting at states $L > L_1$—from $\frac{r}{\gamma}$ to objects of the form $W_{\mathcal{L}}(p(L, m(L_1)))$ for $L > L_1$. Hence we must have $V_{m(L_1)}(L_1) < \frac{r}{\gamma}$, a contradiction.

This proves the first statement. Next, we provide an explicit bound on V when f is non-decreasing.

Claim 1. *If f is uniform,* $\lim_{L\to\infty} p(L, m(L)) = \frac{2(s-c)}{(b-c)+(s-c)}$.

Proof of claim 1. By Lemma 16, the marginal member $y(L)$ satisfies $L = \frac{y(L)(b-s)}{(1-y(L))(s-c)}$, whence $y(L) = \frac{s-c}{s-c+(b-s)\frac{1}{L}}$. If f is uniform, we have $m(L) = \frac{1+y(L)}{2}$, so $m(L) = \frac{1}{2}\frac{2L(s-c)+b-s}{L(s-c)+b-s}$. Recall that $p(L, m(L)) = \frac{1}{1+\left(\frac{1}{m(L)}-1\right)L}$. Then $p(L, m(L)) = \frac{2L(s-c)+b-s}{L(2(s-c)+b-s)+b-s}$, so $\lim_{L\to\infty} p(L, m(L)) = \frac{2(s-c)}{(b-c)+(s-c)}$. ■

Assume that f is uniform. By Lemma 4, $x \mapsto V(x)$ is strictly increasing and, by Lemma 17, $L \mapsto p(L, m(L))$ is decreasing, so $L \mapsto V(p(L, m(L)))$ is decreasing. Thus, for all L, $V(p(L, m(L))) \geq \lim_{L'\to\infty} V(p(L', m(L'))) \geq \frac{r}{\gamma}$. By Claim 1, $\lim_{L'\to\infty} V(p(L', m(L'))) \geq V\left(\frac{2(s-c)}{(b-c)+(s-c)}\right)$. By Lemma 8, this result extends to all non-decreasing densities f.

Next, we show that $V\left(\frac{2(s-c)}{(b-c)+(s-c)}\right) \geq \frac{1}{\gamma}\frac{(b-c)s+(s-c)b}{(b-c)+(s-c)}$. Note that, in an equilibrium in which the organization experiments forever, the payoff of an agent x is bounded below by her payoff from staying in the organization forever. This is given by $\frac{b}{\gamma}$ if the risky policy is good and $\frac{c}{\gamma}$ if not. Then $V(x) \geq x\frac{b}{\gamma} + (1-x)\frac{c}{\gamma}$, which implies that $V\left(\frac{2(s-c)}{(b-c)+(s-c)}\right) \geq \frac{1}{\gamma}\frac{(b-c)s+(s-c)b}{(b-c)+(s-c)}$.

Finally, note that there exist parameter values such that $\frac{1}{\gamma}\frac{(b-c)s+(s-c)b}{(b-c)+(s-c)} \geq \frac{r}{\gamma}$ is satisfied. In general, for any values of b, s and c satisfying $b > s > c > 0$, there is $r^*(b,s,c)$ such that the condition holds if $r \leq r^*(b,s,c)$, and, moreover, $r^*(b,s,c) \in (s,b)$. ■

Proof of Proposition 4. Follows from Proposition 12. ■

Proposition 13. *There exist b, r, s, c and f, $\epsilon \in (0,1]$ and $L^* > 0$ such that an equilibrium of the following form exists: whenever $L = L^*$, the organization stops experimenting with probability ϵ, and whenever $L \neq L^*$, the organization continues* *experimenting with probability one.*

Proof of Proposition 13. For convenience, we multiply all the value functions in this proof by γ. Let $V_x^\epsilon(L)$ denote the value function of agent x given that the state is $L(k,t) = L$ and the behavior on the equilibrium path is as described in the Proposition. Note that $V_x^0(L)$ is the value function of agent x when the state is L and there is perpetual experimentation.

We claim that we can choose the density f such that there is a unique global minimum of $L \mapsto V^0_{m(L)}(L)$, which we will call L^*, and in addition so that $V^0_{m(L)}(L)$ has a kink at L^*. Because, by Corollary 1, $V^0_{m(L)}(L) = V^0_{p(L,m(L))}(1)$ and $x \mapsto V_x(1)$ is smootly increasing, this is equivalent to $p(L, m(L))$ being uniquely minimized at L^* with a kink at L^*. This claim follows from Lemma 18.[57]

Note that $V^0_{m(L)}(L)$ does not depend on r and the density f constructed in the proof of Lemma 18 does not depend on r, which implies that L^* does not depend on r. Then we can choose r such that

$$V^0_{m(L^*)}(L^*) = r \tag{5}$$

Then, because L^* is the unique minimizer of $L \mapsto V^0_{m(L)}(L)$, we have $V^0_{m(L)}(L) > r$ for all $L \neq L^*$.

We aim to show that if we change the equilibrium to require that experimentation stops at $L = L^*$ with an appropriately chosen probability $\epsilon > 0$, the constraints $V^\epsilon_{m(L^*)}(L^*) = r$ and $V^\epsilon_{m(L)}(L) \geq r$ for all $L \neq L^*$ still hold.

It is useful to note at this point that the value function can be written recursively. Towards this end, we introduce the following notation. For any strategy profile and

[57] Technically, we also need the condition that $V^0_{m(L^*)}(L^*) < \lim_{L\to\infty} V^0_{m(L)}(L)$, but this is also satisfied by the construction in Lemma 18.

any L, $L' \in \mathbb{R}$, define

$$T_{x,L'}(L) = \int_0^\infty \gamma e^{-\gamma t} Pr[\exists s \in [0,t] : L_s = L' | L_0 = L] dt$$
$$\hat{V}_{x,L'}(L) = \int_0^\infty \gamma e^{-\gamma t} E[u_x(h^t) \mathbb{1}_{\nexists s \in [0,t]: L_s = L'} | L_0 = L] dt$$
$$\tilde{V}_{x,L'}(L) = \frac{\hat{V}_{x,L'}(L)}{1 - T_{x,L'}(L)}$$

where $u_x(h^t)$ is agent x's flow payoff at time t and history h^t and the expectation is taken with respect to the stochastic process induced by the equilibrium strategy and the stochastic process $\left(\tilde{L}_\tau\right)_\tau$.

Intuitively, $T_{x,L'}(L)$ is the weighted discounted probability that the stochastic process $(L_s)_s$ hits the value L' at least once, $\hat{V}_{x,L'}(L)$ is the expected utility of agent x starting with $L_0 = L$ but setting the continuation value to zero when $(L_s)_s$ hits L', and $\tilde{V}_{x,L'}(L)$ is a normalization. Then the value function can be written recursively as

$$V_x(L) = (1 - T_{x,L'}(L))\tilde{V}_{x,L'}(L) + T_{x,L'}(L)V_x(L')$$

Taking $L' = L^*$, this implies that for any $\epsilon \in [0,1]$, we have

$$V_x^\epsilon(L) = (1 - T_{x,L^*}(L))\tilde{V}_{x,L^*}(L) + T_{x,L^*}(L)V_x^\epsilon(L^*) \tag{6}$$

where $T_{x,L^*}(L)$ is independent of ϵ, since changing ϵ has no impact on the policy path except when $L = L^*$. Let

$$T_{x,L'}(L^+) = \int_0^\infty \gamma e^{-\gamma t} Pr[\exists s \in (0,t] : L_s = L' | L_0 = L] dt$$

$$\hat{V}_{x,L'}(L^+) = \int_0^\infty \gamma e^{-\gamma t} E[u_x(h^t) \mathbb{1}_{\nexists s \in (0,t]: L_s = L'} | L_0 = l] dt$$

and

$$\tilde{V}_{x,L'}(L^+) = \frac{\hat{V}_{x,L'}(L^+)}{1 - T_{x,L'}(L^+)}$$

Observe that $\tilde{V}^\epsilon_{x,L^*}(L^{*+}) = \lim_{L \searrow L^*} \tilde{V}^\epsilon_{x,L^*}(L)$ and $T_{x,L^*}(L^{*+}) = \lim_{L \searrow L^*} T_{x,L^*}(L)$. Let $\tilde{W}^\epsilon_x = \tilde{V}^\epsilon_{x,L^*}(L^{*+})$ and $W^\epsilon_x = \lim_{L \searrow L^*} V^\epsilon_x(L)$. W^ϵ_x is the expected continuation value of agent x when $L = L^*$ and the median member, $m(L^*)$, has just decided *not* to stop experimenting. This is closely related to $V_x^\epsilon(L^*)$, which is the expected continuation value where the expectation is taken before $m(L^*)$ has decided whether to stop experimenting or not. Specifically, we have

$$V_x^\epsilon(L^*) = \epsilon r + (1-\epsilon) W_x^\epsilon = \epsilon r + (1-\epsilon)\left((1 - T_{x,L^*}\left(L^{*+}\right))\tilde{W}_x^\epsilon + T_{x,L^*}\left(L^{*+}\right) V_x^\epsilon(L^*)\right)$$

Solving this for $V_x^\epsilon(L^*)$, we obtain

$$\begin{aligned} V_x^\epsilon(L^*) &= \frac{\epsilon r + (1-\epsilon)(1 - T_{x,L^*}(L^{*+}))\tilde{W}_x^\epsilon}{1 - (1-\epsilon) T_{x,L^*}(L^{*+})} \\ &= V_x^0(L^*) + \epsilon \frac{r - V_x^0(L^*)}{1 - (1-\epsilon) T_{x,L^*}(L^{*+})} \end{aligned} \tag{7}$$

where the second equality follows from the fact that $\tilde{W}_x^\epsilon = V_x^0(L^*)$ because $\tilde{W}_x^\epsilon$ is the continuation value of the agent conditional on the event that $(L_s)_s$ never hits L^* again, which means that in this case experimentation continues forever. Hence, substituting (7) into (6), we obtain

$$\begin{aligned} V_x^\epsilon(L) &= (1 - T_{x,L^*}(L))\tilde{V}_{x,L^*}(L) + T_{x,L^*}(L)\left(V_x^0(L^*) + \epsilon \frac{r - V_x^0(L^*)}{1 - (1-\epsilon) T_{x,L^*}(L^{*+})}\right) \\ &= T_{x,L^*}(L) \epsilon \frac{r - V_x^0(L^*)}{1 - (1-\epsilon) T_{x,L^*}(L^{*+})} + V_x^0(L) \end{aligned} \tag{8}$$

At the same time, because we have assumed that $V_{m(L)}(L)$ is minimized at L^* with a kink at L^*, there exist $\delta > 0$ and $K > 0$ such that for all $L \in (L^* - \delta, L^* + \delta)$

$$V_{m(L)}^0(L) = V_{p(L,m(L))}^0(1) \quad \geq V_{p(L^*,m(L^*))}^0(1) + K|L - L^*| = r + K|L - L^*| \tag{9}$$

where the first equality follows from Corollary 1, and the last equality follows from (5). On the other hand, for $L \notin (L^* - \delta, L^* + \delta)$ there exists $K' > 0$ such that

$$V_{m(L)}^0(L) = V_{p(L,m(L))}^0(0) \geq V_{p(L^*,m(L^*))}^0(0) + K' = r + K' \tag{10}$$

where the first equality follows from Corollary (1), the inequality follows from the fact that $p(L, m(L)) - p(L^*, m(L^*))$ is bounded away from zero in this case, and the second equality follows from (5).

By (8), $V^{\epsilon}_{m(L)}(L) \geq r$ is equivalent to $V^0_{m(L)}(L) \geq r - T_{m(L),L^*}(L)\epsilon\frac{r-V^0_{m(L)}(L^*)}{1-(1-\epsilon)T_{m(L),L^*}(L^{*+})}$. If $V^0_{m(L)}(L^*) - r \leq 0$, then we are done, so suppose that $V^0_{m(L)}(L^*) - r > 0$.

Suppose that $L \in (L^* - \delta, L^* + \delta)$. Then, by (9), it is sufficient that

$$
\begin{aligned}
& r + K|L - L^*| \geq r - T_{m(L),L^*}(L)\epsilon\frac{r - V^0_{m(L)}(L^*)}{1 - (1-\epsilon)T_{m(L),L^*}(L^{*+})} \\
\iff & K|L - L^*| \geq T_{m(L),L^*}(L)\epsilon\frac{V^0_{m(L)}(L^*) - r}{1 - (1-\epsilon)T_{m(L),L^*}(L^{*+})} \\
\Longleftarrow & K|L - L^*| \geq \epsilon\frac{V^0_{m(L)}(L^*) - r}{1 - T_{m(L),L^*}(L^{*+})} \iff \epsilon \leq \frac{K|L - L^*|(1 - T_{m(L),L^*}(L^{*+}))}{V^0_{m(L)}(L^*) - V^0_{m(L^*)}(L^*)}
\end{aligned}
$$

where we have used that $T_{m(L),L^*}(L) \in (0,1)$.

Suppose next that $L \notin (L^* - \delta, L^* + \delta)$. Then, because $V^0_{m(L)}(L) \geq r + K'$ by (10), it is sufficient that

$$
\begin{aligned}
& r + K' \geq r - T_{m(L),L^*}(L)\epsilon\frac{r - V^0_{m(L)}(L^*)}{1 - (1-\epsilon)T_{m(L),L^*}(L^{*+})} \\
\iff & K' \geq T_{m(L),L^*}(L)\epsilon\frac{V^0_{m(L)}(L^*) - r}{1 - (1-\epsilon)T_{m(L),L^*}(L^{*+})} \\
\Longleftarrow & K' \geq \epsilon\frac{b - r}{1 - T_{m(L),L^*}(L^{*+})} \iff \epsilon \leq K'(1 - T_{m(L),L^*}(L^{*+}))\frac{1}{b - r}
\end{aligned}
$$

where we have used that $T_{m(L),L^*}(L) \in (0,1)$ and $V^0_{m(L)}(L^*) \leq b$.

Let $\epsilon_1 = \inf_{L\in(L^*-\delta,L^*+\delta)} K\left(1 - T_{m(L),L^*}(L^{*+})\right)\frac{|L-L^*|}{V^0_{m(L)}(L^*)-V^0_{m(L^*)}(L^*)}$ and $\epsilon_2 = \inf_{L\notin(L^*-\delta,L^*+\delta)} K'\left(1 - T_{m(L),L^*}(L^{*+})\right)\frac{1}{b-r}$. To have $\min\{\epsilon_1, \epsilon_2\} > 0$ we need to verify that $\sup_x T_{x,L^*}(L^{*+}) < 1$ and that $\sup_{L\in(L^*-\delta,L^*+\delta)}\left|\frac{V^0_{m(L)}(L^*)-V^0_{m(L^*)}(L^*)}{L-L^*}\right|$ is finite. $\sup_x T_{x,L^*}(L^{*+}) < 1$ is immediate. The fact that $\sup_{L\in(L^*-\delta,L^*+\delta)}\left|\frac{V^0_{m(L)}(L^*)-V^0_{m(L^*)}(L^*)}{L-L^*}\right|$ is finite follows from the fact that $\frac{\partial}{\partial x}V^0_x(L^*)$ and $m'(L)$ are bounded.

Then choosing $\epsilon \in (0, \min\{\epsilon_1, \epsilon_2\})$ delivers the result. ∎

Proof of Proposition 5. Take the example constructed in Proposition 13, and assume that $L_0 > L^*$.[58]

Let $P_\theta(L_0)$ be the probability that, conditional on starting at L_0 and the state being $\theta \in \{G, B\}$, the organization stops experimenting at any finite time $t < \infty$. We will show that $P_G(L_0) > P_B(L_0)$ for L_0 large enough. In fact, we will prove a stronger result: we will show that there is $C > 0$ such that $P_G(L_0) \geq C > 0$ for all $L_0 > L^*$, but $\lim_{L_0 \to \infty} P_B(L_0) = 0$.

Let $Q_\theta(L_0, L^*)$ denote the probability that there exists $t < \infty$ such that $L_t \in \left(\left(\frac{c}{b}\right) L^*, L^*\right]$ when the state is $\theta \in \{G, B\}$. $Q_\theta(L_0, L^*)$ is the probability that L_t ever crosses over to the left of L^*.

We claim that $Q_G(L_0, L^*) = 1$ for all $L_0 > L^*$ but $\lim_{L_0 \to \infty} Q_B(L_0, L^*) = 0$.

Let $l(k,t) = \ln L(k,t)$, and note that $l(k,t) = \ln\left(\left(\frac{c}{b}\right)^k e^{(b-c)t}\right) = k \ln\left(\frac{c}{b}\right) + \ln\left(e^{(b-c)t}\right) = k(\ln(c) - \ln(b)) + (b-c)t$. Let $l_0 = \ln(L_0)$.

When $\theta = G$, we then have $(l_t)_t = l_0 + (b-c)t - [\ln(b) - \ln(c)]\, N(t)$, where $(N(t))_t$ is a Poisson process with rate b, that is, $N(t) \sim P(bt)$. This can be written as a random walk: for integer values of t, $l_t - l_0 = \sum_{i=0}^t S_i$, where $S_i = b - c - [\ln(b) - \ln(c)]\, N_i$, and $N_i \sim P(b)$ are iid. Note that $E[S_i] = b - c - b\,(\ln(b) - \ln(c)) < 0$.[59] Then, by the strong law of large numbers, we have $\frac{l_t}{t} \xrightarrow[t\to\infty]{} E[S_i] < 0$ a.s., whence $l_t \xrightarrow[t\to\infty]{} -\infty$ a.s., implying the first claim.

On the other hand, when $\theta = B$, we have $(l_t)_t = l_0 + (b-c)t - \ln(b-c)N(t)$, where $(N(t))_t$ is a Poisson process with rate c. This can be written as a random walk with positive drift: $l_t - l_0 = \sum_{i=0}^t S_i$, where $S_i = b - c - [\ln(b) - \ln(c)]\, N_i$, $N_i \sim P(c)$, and $E[S_i] = b - c - c\,(\ln(b) - \ln(c)) > 0$. As above, by the strong law of large numbers, we have $l_t \xrightarrow[t\to\infty]{} \infty$ a.s.

Note that $Q_B\left(L, \frac{c}{b}L\right) = q$ is independent of L because $(l_t)_t$ follows a random

[58]Technically, our definition of L_t requires that $L_0 = 1$, but we can relax this assumption by considering a continuation of the game starting at some $t_0 > 0$, where, by assumption, the number of successes at time t_0 is such that the state variable at t_0 is L_0. This example can be fit into our original framework by redefining the density of prior beliefs $\tilde{f}$ to be the density of the posteriors held by agents when $L = L_0$ and f is as in Proposition 13. With this relabeling, L_0 would equal 1 and L^* would shift to some value less than 1. We find it is easier to think in terms of shifting L_0 and leaving f unchanged.

[59]Let $\frac{b}{c} = 1 + x$. Then $b - c - b\,(\ln(b) - \ln(c)) = c(x - (1+x)\ln(1+x))$, where $x - (1+x)\ln(1+x)$ is negative for all $x > 0$. Similarly, $b - c - c\,(\ln(b) - \ln(c)) = c(x - \ln(1+x))$, where $x - \ln(1+x)$ is positive for all $x > 0$.

walk. Now suppose for the sake of contradiction that $\limsup_{L\to\infty} Q_B(L, L^*) > 0$. We claim that this implies $q = 1$. Suppose towards a contradiction that $q < 1$. Fix $J \in \mathbb{N}$. Then, for L_0 large enough that $\left(\frac{c}{b}\right)^{2J+1} L_0 > L^*$,

$$Q_B(L_0, L^*) \leq \prod_{j=0}^{J} Q_B\left(\left(\frac{c}{b}\right)^{2j} L_0, \left(\frac{c}{b}\right)^{2j+1} L_0\right) = q^{J+1}$$

This implies that, whenever $\limsup_{L\to\infty} Q_B(L, L^*) > 0$, we have $q = 1$, as the above equation must hold for arbitrarily large J. Hence $(l_t)_t$ is recurrent, that is, it visits the neighborhood of every $l \in \mathbb{R}$ infinitely often (Durrett 2010: pp. 190–201). However, this contradicts the fact that $\lim_{t\to\infty} l_t = \infty$ a.s. Therefore, $\limsup_{L\to\infty} Q_B(L, L^*) = 0$.

This implies that $P_B(L_0) \leq Q_B(L_0, L^*) \to 0$ as $L_0 \to \infty$. On the other hand, $P_G(L_0) \geq Q_G(L_0, L^*) \inf_{L \in \left(\left(\frac{c}{b}\right)L^*, L^*\right]} P_G(L) > 0$. The first inequality holds for the following reason. With probability 1, if $L_t = L^*$ for some t, there must be $t' < t$ such that $L_{t'} \in \left(\frac{c}{b}L^*, L^*\right)$, which happens with probability $Q_G(L_0, L^*)$. Conditional on this event, the probability of hitting state L^* in the continuation is $P_G(L_{t'})$. Note that $\inf_{L \in \left(\left(\frac{c}{b}\right)L^*, L^*\right]} P_G(L) > 0$ because it is equal to $P_G\left(\left(\frac{c}{b}\right) L^*\right)$. ■

Proof of Proposition 7. Fix an equilibrium σ in which organization experiments forever and let μ_t be the size of the organization at time t on the equilibrium path. Let $g_t = g(\mu_t)$. The *first success* that happens at time t yields the per-capita payoff of g_t, and all further successes pay 1 (because all agents enter the organization after the first success).

Let $P_t = 1 - e^{-bt}$ denote the probability that there is a success by time t given that the risky technology is good. In the above problem, an agent with belief x who expects experimentation to continue forever has utility

$$\begin{aligned} V_{(g_t)_t}(x) =& x \int_0^{t^*} e^{-\gamma t} \left(P_t b + (1 - P_t) g_t b\right) dt + \\ & x \int_{t^*}^{\infty} e^{-\gamma t} \left(P_t b + (1 - P_t) s\right) dt + (1 - x) \int_{t^*}^{\infty} e^{-\gamma t} s dt \end{aligned}$$

where t^* is the time at which the agent leaves, that is, when her posterior reaches $\frac{s}{g_t b}$.

Now consider the case in which $g_t = g$ for all t. Then the above expression equals

$$V_g(x) = x\left(\frac{b}{\gamma} - \frac{b}{\gamma+b} + \frac{gb\left(1-e^{-(\gamma+b)t^*}\right)}{\gamma+b} + \frac{e^{-(\gamma+b)t^*}s}{\gamma+b} - \frac{e^{-\gamma t^*}s}{\gamma}\right) + \frac{e^{-\gamma t^*}s}{\gamma}$$

Suppose that $f = f_\alpha$, as in Proposition 9. By the same arguments as in that Proposition, if y_t satisfies $p_t(y_t) = \frac{s}{gb}$ for all t, then $p_t(m_t) \searrow \frac{s}{\lambda(gb-s)+s}$ as $t \to \infty$, and, by Lemma 3, we have $t^* = -\frac{\ln(\lambda)}{b}$.[60] Then

$$V_g\left(\frac{s}{\lambda(gb-s)+s}\right) = \frac{s}{\lambda gb + (1-\lambda)s}\left(\frac{b}{\gamma} - \frac{b}{\gamma+b} + \frac{gb\left(1-\lambda^{\frac{\gamma+b}{b}}\right)}{\gamma+b} + \frac{\lambda^{\frac{\gamma+b}{b}}s}{\gamma+b} - \frac{\lambda^{\frac{\gamma}{b}}s}{\gamma}\right) + \frac{\lambda^{\frac{\gamma}{b}}s}{\gamma}$$

Since this is a hyperbola in g, it is either increasing in g for all $g > 0$ or decreasing in g for all $g > 0$. In particular, when the congestion effect is maximal, that is, when $g \to \infty$, we have

$$\lim_{g\to\infty} \gamma V_g\left(\frac{s}{\lambda(gb-s)+s}\right) = \frac{s}{\lambda}\left(1-\lambda^{\frac{\gamma+b}{b}}\right)\frac{\gamma}{\gamma+b} + \lambda^{\frac{\gamma}{b}}s = \frac{s}{\lambda}\frac{\gamma}{\gamma+b} + \lambda^{\frac{\gamma}{b}}s\frac{b}{\gamma+b}$$

On the other hand, when the economies of scale are maximal, that is, as $g \to \frac{s}{b}$,[61]

$$\lim_{g\to\frac{s}{b}} \gamma V_g\left(\frac{s}{\lambda(gb-s)+s}\right) = \gamma\left(\frac{b}{\gamma} - \frac{b}{\gamma+b} + \frac{s\left(1-\lambda^{\frac{\gamma+b}{b}}\right)}{\gamma+b} + \frac{\lambda^{\frac{\gamma+b}{b}}s}{\gamma+b}\right) = b\frac{b}{\gamma+b} + s\frac{\gamma}{\gamma+b}$$

Thus $\frac{s}{\lambda}\frac{\gamma}{\gamma+b} + \lambda^{\frac{\gamma}{b}}s\frac{b}{\gamma+b} > b\frac{b}{\gamma+b} + s\frac{\gamma}{\gamma+b}$ is equivalent to $\lim_{g\to\infty} V_g\left(\frac{s}{\lambda(gb-s)+s}\right) > \lim_{g\to\frac{s}{b}} V_g\left(\frac{s}{\lambda(gb-s)+s}\right)$. Because $V_g\left(\frac{s}{\lambda(gb-s)+s}\right)$ is either increasing in g for all $g > 0$ or decreasing in g for all $g > 0$, this condition implies that $V_g\left(\frac{s}{\lambda(gb-s)+s}\right)$ is increasing in g for all $g > 0$. The argument in the case when the inequality is reversed is similar.

In addition, note that if $V_g\left(\frac{s}{\lambda(gb-s)+s}\right)$ is increasing in g, then we can guarantee

[60] Note that this is the same t^* as in the baseline model.
[61] If $g < \frac{s}{b}$, we enter a degenerate case in which the organization becomes empty immediately.

that, with a congestion effect,

$$V_{(g_\tau)_{\tau\geq t}}\left(p_t(m_t)\right) > V_{g_t}\left(p_t(m_t)\right) > V_{g_t}\left(\frac{s}{\lambda(g_t b - s) + s}\right) > V\left(\frac{s}{\lambda(b-s)+s}\right)$$

Here the first inequality follows because $g_\tau \mapsto V_{g_t,\ldots,g_\tau,\ldots}(x)$ is increasing, under congestion effect $\mu \mapsto g(\mu)$ is decreasing and under perpetual experimentation $t \mapsto \mu_t$ is decreasing, so $t \mapsto g_t = g(\mu_t)$ is increasing. The second inequality follows because $x \mapsto V_{g_t}(x)$ is strictly increasing and $p_t(m_t) \searrow \frac{s}{\lambda(gb-s)+s}$ as $t \to \infty$. The last inequality follows because $g \mapsto V_g\left(\frac{s}{\lambda(gb-s)+s}\right)$ is increasing, under congestion effect $\mu \mapsto g(\mu)$ is decreasing and in the baseline model we have $g(\mu) = g(1) = 1$ for all μ.

Thus the condition to obtain experimentation forever is slacker with a congestion effect than in the baseline model at every t, not just in the limit. By the same argument, the condition for experimentation forever is tighter for all t under economies of scale.[62] ■

Proof of Proposition 8.

We have

$$\begin{aligned}\tilde{V}_t(x) =& x\int_t^\infty e^{-\gamma(\tau-t)}\left[P_\tau b + (1-P_\tau)(sF(y_\tau) + b(1-F(y_\tau)))\right]d\tau \\ &+ (1-x)\int_t^\infty e^{-\gamma(\tau-t)} sF(y_\tau) d\tau,\end{aligned}$$

where $P_\tau = 1 - e^{-b(\tau-t)}$ is the probability that there has been a success by time τ, conditional on the state being good and there being no success up to time t, and $F(y_\tau)$ is the fraction of the population that are outsiders at time τ, conditional on no successes. We can rewrite this equation as

$$\begin{aligned}\tilde{V}_t(x) &= x\int_t^\infty e^{-\gamma(\tau-t)}\left[b - (1-P_\tau)(b-s)F(y_\tau)\right]d\tau + (1-x)\int_t^\infty e^{-\gamma(\tau-t)} sF(y_\tau) d\tau \\ &= x\frac{b}{\gamma} - x\int_t^\infty e^{-(\gamma+b)(\tau-t)}(b-s)F(y_\tau)d\tau + (1-x)\int_t^\infty e^{-\gamma(\tau-t)} sF(y_\tau) d\tau \quad (11)\end{aligned}$$

[62] If $g \mapsto V_g\left(\frac{s}{\lambda(gb-s)+s}\right)$ is decreasing, it is more difficult to make general statements about what happens away from the limit because in this case the effect of moving away from the limit goes against the result: for instance, under congestion effect the condition becomes tighter in the limit but increasing g_t slackens the condition.

Let $F_\tau = F(y_\tau)$ and note that $\tau \mapsto F_\tau$ is weakly increasing.

The upper bound for $\tilde{V}_t(x)$ is now obtained as follows. Note that, given $\tau \geq t$, the derivative of (11) with respect to F_τ is proportional to $-xe^{-b(\tau-t)}(b-s)+(1-x)s$. It follows that if the agent could choose F_τ everywhere at will to maximize her payoff, she would choose $F_\tau = 1$ for $\tau \geq t(x)$ and $F_\tau = 0$ for $\tau < t(x)$, where $t(x)$ is defined by the condition $xe^{-b(t(x)-t)}(b-s) = (1-x)s$ (obtained by setting the derivative equal to 0). The result of this choice is $V(x)$, her utility in the private values case, in which she only cares about her own entry and exit decisions and gets to choose them optimally. Because in the common values case the entry and exit decisions of other agents are not optimal from x's point of view, $\tilde{V}_t(x)$ must be weakly lower than $V(x)$.

As for the lower bound, assume for the sake of argument that F_τ is constant for all y_τ and equal to $\overline{F} \in [0,1]$. Then the expression in (11) is

$$x\frac{b}{\gamma} - x\frac{(b-s)\overline{F}}{\gamma+b} + (1-x)\frac{s\overline{F}}{\gamma}$$

which is linear in $\overline{F}$ and is minimized either when $\overline{F} = 0$ or when $\overline{F} = 1$. In the first case, the expression equals $x\frac{b}{\gamma}$. In the second case, it equals $x\frac{b}{\gamma} - x\frac{(b-s)}{\gamma+b} + (1-x)\frac{s}{\gamma}$.

To finish the proof, we argue that whenever F_τ is weakly increasing in τ, the expression in 11 is higher than the expression that is obtained when F_τ is replaced by a suitably chosen constant $\overline{F}$. Hence the lower bound obtained for constant F_τ applies in all cases.

The argument is as follows. Take $\overline{F} = F_{t(x)}$. Then for $\tau > t(x)$, F_τ is weakly greater than $\overline{F}$ and $\tilde{V}_t(x)$ is increasing in the value of F at τ. Conversely, for $\tau < t(x)$, F_τ is weakly lower than $\overline{F}$ and $\tilde{V}_t(x)$ is decreasing in the value of F at τ. Hence the agent's utility is weakly higher under F_τ than under a constant $F_{t(x)}$. ■