EXPERIMENTATION IN ENDOGENOUS ORGANIZATIONS[*]

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

March 2023

Abstract

We study policy experimentation in organizations with endogenous membership. An organization decides when to stop a policy experiment based on its results. As information arrives, agents update their beliefs, and enter or leave the organization based on their expected flow payoffs. Unsuccessful experiments make all agents more pessimistic, but also drive out conservative members. We identify sufficient conditions under which the latter effect dominates, leading to excessive experimentation. In fact, the organization may experiment forever in the face of mounting negative evidence. Ex post heterogeneous payoffs exacerbate the problem, as optimists can join forces with guaranteed winners. Control by shareholders who own all future payoffs, however, can have a corrective effect. Our results contrast with models of collective experimentation with fixed membership, in which under-experimentation is the typical outcome.

JEL codes: D71, D83
Keywords: experimentation, median voter, exit, endogenous population

1 Introduction

Organizations frequently face opportunities to experiment with promising but untested policies. According to conventional wisdom, experimentation should respond to information: agents should become more pessimistic after an adverse outcome, and they should abandon

[*]We would like to thank Daron Acemoglu, Avidit Acharya, Sandeep Baliga, Dana Foarta, Robert Gibbons, Faruk Gul, Matías Iaryczower, Navin Kartik, Elliot Lipnowski, Wolfgang Pesendorfer, Nicola Persico, Andrea Prat, Caroline Thomas, Juuso Toikka, Rui Wang, the audiences at the 2018 APSA Annual Meeting (Boston), the 2018 Columbia Political Economy Conference, the 2019 Canadian Economic Theory Conference, the 2020 Utah Winter Organizational and Political Economics Conference and the SITE Political Economy Workshop 2020, as well as seminar participants at MIT, Princeton, Indiana, Oxford, Northwestern, Berkeley, Calgary and UCEMA for helpful comments. All remaining errors are our own.

[†]Department of Politics, Princeton University. Address: 311 Fisher Hall, Princeton NJ 08540. Phone number: 6172330642. E-mail: germang@princeton.edu.

[‡]Department of Economics, University of Pittsburgh. Address: Wesley W. Posvar Hall, Pittsburgh PA 15213. Phone number: 9179756491. Email: svk14@pitt.edu.

an experiment if enough negative information accumulates. In addition, when experimentation is collective, the temptation to free-ride and fears that information will be misused by other agents lower incentives to experiment (Keller, Rady and Cripps 2005; Strulovici 2010). Thus, if anything, organizations should experiment too little.

Yet history is littered with examples of organizations that have stubbornly persisted with unsuccessful policies to the bitter end. The Communist experiments of the 20th century are a dramatic example: many Communist societies maintained rigid command economies in the face of prolonged economic decline, in some cases all the way up until their governments were violently overthrown. Meanwhile, some like-minded European parties—notably the French Communist Party—held fast in their support for the Soviet Union even as it collapsed, and they themselves faded into irrelevance. Of course, these collective projects had detractors. But rather than fight to change the course, many of them left.

The same sort of collective rigidity is displayed by firms that relentlessly pursue a revolutionary vision or new technology all the way to either ultimate success or bankruptcy. This phenomenon is common at Silicon Valley companies and other start-ups, such as Theranos and Moderna, many of whose employees are almost religiously devoted to the company's mission (Chen 2022). These "true believers" become especially overrepresented during hard times, because they are the least likely to quit.

Motivated by these and similar examples, we propose an explanation for obstinate behavior by organizations. In our baseline model, an organization chooses in each period between a safe policy, which yields a known flow payoff, and a risky policy of uncertain quality, which yields lump sums arriving at random times if it is good, and nothing if it is bad. There is a continuum of agents. They hold heterogeneous prior beliefs about the type of the risky policy, but are otherwise identical. In every period, each agent decides whether to participate in the organization, and obtain the flow payoff generated by its policy, or receive a known outside option. All agents who participate in the organization today are granted voting rights over tomorrow's policy. More precisely, we assume that the median voter—the member with the median prior belief—chooses tomorrow's policy. Whenever the risky policy is used, the results are publicly observed.

Our assumptions reflect three premises of our theory: agents can influence an organization's policy if they are members; they can enter and leave in response to information; and some are more optimistic than others. The key observation is that, under these conditions, new information affects both the beliefs of all agents and the set of agents who desire membership. These effects offset each other: for instance, bad news make all agents pessimistic, but also disproportionately induce those with low priors to exit—and stop expressing dissent. As a result, the distribution of beliefs in an organization can display a dampened or even contrary response to information. Our paper thus formalizes Hirschman (1970)'s argument that members of a declining organization may react by leaving ("exit") or pushing for policy changes ("voice"), and that these two forces can substitute for one another.

Our first main result provides conditions under which this logic leads to excessive experimentation from the point of view of all agents. We show that over-experimentation can take a particularly stark form, in which the organization *never* stops experimenting in the face of failure. Perpetual experimentation is more likely when agents are patient, the distribution of priors contains enough optimists, and the outside option is attractive, so that exit is tempting. In fact, perpetual experimentation always obtains when the outside option offers a close enough payoff to that of the organization's safe policy.

Relative to a benchmark with a fixed decision-maker, two forces affect the pivotal agent's decision to experiment. On the one hand, the identity of the pivotal agent gradually shifts to an ex ante more optimistic member as bad news accumulate. On the other hand, the current pivotal agent is reluctant to continue experimenting precisely because she has limited control over future policy choices. The first force pushes the organization to over-experiment, while the second makes each agent more cautious. Excessive experimentation results when the first force dominates. When perpetual experimentation does not obtain, this interplay of forces can lead to too much or too little experimentation from the point of view of the initial pivotal agent.

We also show that the emergence of perpetual experimentation is robust to several variations in assumptions, including general voting rules, size-dependent flow payoffs or learning rates, barriers to reentry, and different learning processes, such as bad news or imperfectly informative good news. Moreover, when news are imperfectly informative, it is possible for an organization to abandon the risky policy *only after a successful streak*. Paradoxically, the organization may thus experiment more precisely when the risky policy is bad; failure may lead to radicalization, while success may render the organization more conservative and prone to abandoning the very engine of its success.

Our main results also extend to an alternative model in which the risky policy is good for some agents and bad for others ex post, and winners and losers are revealed through experimentation (as in Strulovici 2010). In fact, the problem of over-experimentation becomes more severe in this case, as ex ante optimists can make common cause with revealed winners. Finally, perpetual experimentation is also possible if the intensity of membership is adjustable, and the agents are risk-averse. In that case, there is additional selection at the intensive margin: optimists are all in, and gain outsize influence even relative to other members. However, we show that perpetual experimentation is impossible if the organization is a publicly traded firm, controlled by (risk-averse) investors whose stakes represent ownership of the firm's present *and future* payoffs. Although there is again selection at the intensive margin, in the long run, a shrinking population of optimists will struggle to hold all of the company's volatile stock, leading to falling share prices and an eventual takeover by pessimists. Capital markets can hence have a corrective effect on the tendency of organizations towards obstinate behavior.

The rest of the paper proceeds as follows. The rest of this section reviews the related literature and several applications of the model. Section 2 introduces the baseline model, and Section 3 analyzes its equilibria. Section 4 presents two extensions: one allowing for ex post winners and losers, and another that models a publicly traded firm. Section 5 concludes the paper. All proofs are in Appendix A. Additional extensions are presented in Appendix B.

1.1 Related Literature

This paper is related to the literature on strategic experimentation with multiple agents (Keller et al. 2005, Keller and Rady 2010, 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 2021).

In Keller, Rady and Cripps (2005) and Keller and Rady (2010), multiple agents with common priors control two-armed bandits of the same type which may have breakthroughs at different times. In this setting, there is under-experimentation due to free-riding, but encouragement effects can also arise. This is especially true if the agents have asymmetric information (Dong 2021). These effects are not present in our model, as we assume a single collective decision in each period about whether to experiment, and there is no asymmetric information.[1]

In Strulovici (2010), a group of agents decides by voting whether to collectively experiment with a risky technology. Agents have common priors, but experimentation gradually reveals each to be a *winner* or *loser* from the risky technology. In equilibrium, there is too little experimentation because agents fear being trapped into using the new technology as losers if there are enough winners in favor, and vice versa.

There is a similar motive to under-experiment in our model: because pessimists exit after bad news, a pivotal agent may halt experimentation early to avoid a takeover by over-experimenting optimists. However, when each pivotal agent is optimistic enough to take that risk, the selection effect dominates, and the same exit option instead causes over-

[1]While there is free-riding insofar as outsiders benefit from the option value of experimentation, it is not socially costly because the learning rate is independent of the organization's size. However, perpetual experimentation can result even when the learning rate is endogenous, as shown in Appendix B.

experimentation. The two models also differ in that, in Strulovici's model, learning exacerbates the conflict between agents, while in our model learning helps agents converge to a common belief. However, our main results survive in a "heterogeneous outcomes" version of our model that is directly comparable to Strulovici (2010) (see Section 4.1).

The literature on decision-making in clubs studies dynamic policy choices that determine current flow payoffs as well as control over future decisions. Most papers on this topic assume discrete policy spaces (Acemoglu et al. 2008, 2012, 2015, Roberts 2015), as we do. In contrast, Bai and Lagunoff (2011) and Gieczewski (2021) study the case of a continuous policy space, which yields very different results—namely, the policy converges along a smooth transition path to a myopically stable state. This literature has focused on models with fixed, known environments,[2] with tensions arising due to conflicting preferences. In contrast, our agents differ only in their beliefs. And, more importantly, they live in an uncertain environment that they can learn about depending on their choices. In particular, our result that the long-run equilibrium policy may be desired by almost nobody—as in the case of perpetual experimentation—is driven by learning and is novel to the literature. Finally, our paper shares with Gieczewski (2021) an interest in organizations that allow agents to join or leave. This is mainly a superficial connection, as the model in Gieczewski (2021) can be relabeled to fit the more standard case of policy choices that directly shape the set of decision-makers (e.g., immigration policy). Our paper is also the first in this literature to consider membership of variable intensity.

1.2 Applications

In this Section, we discuss how our assumptions map to different applications such as political parties, political reforms, firms and cooperatives, and give examples of each.

Political Parties

Our model captures the internal dynamics of political parties choosing between a "safe" mainstream platform—for example, social democracy—and a more extreme alternative—for example, a communist platform preaching the imminent collapse of capitalism. The selection of extremists into extremist parties, which intensifies when such parties are unsuccessful, explains their rigidity in the face of setbacks.

The decline of the French Communist Party (FCP) fits this pattern. In the 1980s, many high-profile FCP members became disillusioned with the party's platform as they absorbed

[2] An exception is Acemoglu et al. (2015), which only proves some general results in a framework with *exogenous* shocks that does not nest our model.

a stream of negative signals—namely, the unraveling of the Communist experiment in the Eastern Bloc. Ross (1992: 54), for example, writes of the dissenters in the party that "by autumn 1989, in the face of eastern European disasters, rebel ranks grew larger and larger".

Yet many more detractors left the party, as the FCP was "no longer capable of appealing to the broader community of French intellectuals" (Hazareesingh 1991: 3). For example, Pierre Juquin, a prominent member who became a leader of the moderate faction, was "ousted from the Politbureau Central Committee in 1987, and expelled from the party after declaring his independent presidential candidacy" (Bell and Criddle 1989: 524).

As a result, the party failed to adapt and remained loyal to the Soviet Union, to the point that it came "to be equated with the televised image of bureau politique member Pierre Blotin enthusiastically attending the Romanian Communist Party congress days before the deservedly ignominious end of the Ceaucescus" (Ross 1992: 54). The FCP's electoral support thus declined from a base of roughly 20% in the postwar period to less than 3% in the late 2010s (Bell 2003, Damiani and De Luca 2016), with a precipitous drop in the 1980s. Indeed, "by 1990 what little attention was paid to it portrayed it as a crank, marginalized organization" (Ross 1992: 44). Even decades later, it retained the main tenets of its platform, such as the claim that capitalism is on the verge of collapse.[3]

Reforms

Our model also speaks to the dynamics of political reforms. In this application, agents are residents of a country or city that is trying a reform with uncertain results. The residents can stay and try to influence policy, or they can leave. Our baseline model is appropriate if the reform is equally good or bad for all. The "heterogeneous outcomes" variant of our model in Section 4.1 covers reforms that create unexpected winners and losers (c.f. Strulovici (2010), who suggests trade liberalization or a switch to a centralized economy as examples of ex post-unequal reforms).

The Communist experiments of the 20th century illustrate the relationship between emigration and political pressure. Some Communist countries—most notably, the Soviet Union and East Germany—imposed very strict barriers to emigration, while others, such as China and Cuba, had milder restrictions (Dowty 1988). In accord with our model, the Communist regimes of the Soviet Union and East Germany collapsed, but not those of China or Cuba.[4] Even the regimes that failed, however, took a long time to do so. One possible reason is that, as we show in Section 4.1, support for experimentation is even more robust when outcomes

[3] See, for example, the FCP's 2013 manifesto: `http://congres.pcf.fr/35745`.

[4] Notably, many Cuban emigrants were dissidents, as reflected in the high numbers of Republican-leaning Cuban Americans (Bishin and Klofstad 2012).

are heterogeneous, as ex ante optimists can join forces with revealed winners.

In a similar vein, Sellars (2019) argues that emigration served to preserve the political status quo in Mexico and Japan in the 1920s, as many detractors (e.g., supporters of agrarian reform in Mexico) were young men in search of economic opportunity that they could also find abroad. Finally, examples abound of the "Curley effect" (Glaeser and Shleifer 2005), whereby politicians shape their electorate to maintain power. For instance, the eponymous mayor Curley of Boston induced the rich to emigrate with redistributive policies favoring his base of poor Irish constituents; mayor Coleman Young of Detroit drove white residents and businesses out of the city; and Robert Mugabe of Zimbabwe harassed white farmers and seized their property, precipitating their emigration (Meredith 2002).

Firms

Finally, our model can explain the behavior of rigidly ambitious firms. An extreme example is 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, a vision as revolutionary as it was difficult to realize. Over the course of ten years, the firm spent over a hundred million dollars in pursuit of this vision, while doing little to develop incremental innovations as a fall-back plan. It eventually launched in 2013 with inaccurate and fraudulent tests, and never recovered from the ensuing scandal.

Over the ten years leading up to Theranos's turn to 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.[5] As a result, many left the company,[6] even as those who saw Holmes as a visionary remained. Though the board came close to removing Holmes as CEO early on (Carreyrou 2018: 50), she retained control for many years after, because too many who had lost faith in her leadership had quit before they could form a majority.

The selection of "true believers" into the company was thus exacerbated by its lack of progress with its technology. In a similar fashion, Moderna, the biotech company later famous for developing novel mRNA vaccines for COVID-19, was characterized in 2016 as having run into roadblocks in its ambitious projects, lost top talent, and simultaneously retained employees that "live the mission" and "speak with respect bordering on awe about Moderna's promise" (Garde 2016).

[5]For instance, Theranos's lead scientist, Ian Gibbons, told his wife that "nothing at Theranos was working," years after joining the company (Carreyrou 2018: 146).

[6]For example, while deciding whether to buy more shares of the company at a low price, board member Avie Tevanian was asked by a friend: "'Given everything you now know about this company, do you really want to own more of it?' When Avie thought about it, the answer was no" (Carreyrou 2018: 40).

The mapping of our model to firms depends on where the locus of decision-making lies in the firm. In a start-up, the relevant decision-makers may be all employees above a certain level, with comparable influence over decisions. In this case, our baseline model is appropriate. For a larger firm controlled by investors free to trade shares on the secondary market, a better fit is the model we develop in Section 4.2.

Finally, cooperatives are a related mode of organization that closely fit the assumptions of our baseline model. Here agents are individual producers who own factors of production. In a dairy cooperative, for example, each farmer owns a cow. The farmer can manufacture and sell his own dairy products, 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 follow a safe strategy, such as selling fresh milk and yogurt, or pursue a risky one—for example, developing premium cheeses that may or may not become profitable. Should the latter strategy be used, only farmers optimistic enough about its viability will join or remain in the cooperative. Moreover, cooperatives typically allow their members to elect directors.

2 The Baseline 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* ($\vartheta = G$) or *bad* ($\vartheta = B$) and its type, ϑ, is persistent.

There is a continuum of agents, distributed according to a continuous density f over $[0, 1]$. An agent's position 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. Agents who choose not to be members work independently and obtain a guaranteed *autarkic* flow payoff a. The flow payoffs of members depend on the organization's policy.

While the organization uses the safe policy ($\pi_t = 0$), all members receive a guaranteed flow payoff s. When the risky policy is used ($\pi_t = 1$), their payoffs depend on its type. If the risky policy is good, it produces successes which arrive at the jump times of a Poisson process with rate λ. If it is bad, it never succeeds. Each time the risky policy succeeds, all members receive a lump-sum payoff of size h. At other times, they receive zero. We denote by $g = \lambda h$ the expected flow payoff of the good risky policy.

We assume that $0 < a < s < g$: the good risky policy dominates all other policies, the bad risky policy is the worst option, and the organization's safe policy is preferable to the outside option.[7]

When the risky policy is used, its successes are observed by everyone. By Bayes' rule, an agent with prior x who has seen the organization experiment unsuccessfully for a length of time t believes that the risky policy is good with probability

$$p_t(x) = \frac{xe^{-\lambda t}}{xe^{-\lambda t} + (1-x)}. \tag{1}$$

Of course, all posteriors jump to 1 after a success.

The structure of the game is as follows. At each instant $t > 0$, *policy* and *membership* decisions are made. That is, first the organization's median member chooses the policy to be used in the immediate future.[8] After this, all agents are allowed to enter or leave the organization.

To simplify the presentation, we make two assumptions. First, we assume that the risky policy is being used at the start of the game, that is, $\pi_0 = 1$.[9] Second, we assume that a switch to the safe policy is irreversible.[10] We focus on Markov Perfect Equilibria, that is, equilibria in which strategies condition only on the information about the risky policy revealed so far and on the incumbent policy.

Since optimal membership decisions are quite simple, it is convenient to embed them directly into the definition of equilibrium, as follows. Note that optimal membership decisions must condition only on flow payoffs, even though the agents are forward-looking: x wants to be a member at time t if and only if $s + \pi_t(p_t(x)g - s) \geq a$. This is because there is free entry and exit, so there is no need to remain a member during lean times to retain access to future payoffs, or vice versa; and because there is a continuum of agents, so an agent derives no value from her ability to vote. In particular, if the risky policy is being used at time t and no successes have occurred, x will be a member if and only if $p_t(x)g \geq a$. Clearly, $p_t(x)$ is increasing in x: ex ante optimists remain more optimistic after observing

[7] Our model features a single organization with access to a risky technology. We can, however, allow for the existence of other organizations that only have access to safe technologies. a can be interpreted as their (maximal) productivity. The assumption $a < s$ means that the organization with access to the risky technology also enjoys a competitive edge in the realm of safe technologies. Our main results go through if $a \geq s$, but become less interesting as there is no opportunity cost to having the organization experiment.

[8] The set of members will in fact always be an interval, hence Lebesgue measurable, so the median is well defined. Equivalent results are obtained if we instead assume majority voting, as the median will be decisive.

[9] Starting with the safe policy at $t = 0$ is equivalent to starting with the risky policy, unless the population median finds the continuation in the latter scenario inferior to the payoff from never experimenting, in which case experimentation never begins.

[10] We show in the Appendix that this assumption is without loss of generality: in a more general model with unlimited policy changes, switches to the safe policy are permanent in every equilibrium. The reason is that switching to the safe policy brings in more pessimistic members, and hence yields control to a median even more pessimistic than the one who chose to stop experimenting.

information. Hence, if the organization has experimented unsuccessfully until t, the set of members at t will be an interval of the form $[y_t, 1]$, where y_t is defined by the condition $p_t(y_t) = \frac{a}{g}$. Equation 1 implies that $y_t = \frac{a}{a+(g-a)e^{-\lambda t}}$. This, in turn, pins down the identity of m_t, the median member at time t under experimentation, as the median of F restricted to $[y_t, 1]$. On the other hand, if the safe policy is being used, or the risky policy is being used *after* a success, then all agents will choose to be members, as $s, g > a$. And the organization should of course always use the risky policy after a success.

In this "reduced-form" model, the only strategic decision left is the policy choice made by the pivotal agent at each time t to continue experimenting or not, assuming there have been no successes. We say t is a stopping time if m_t would choose to stop experimentation at time t. We can then define an equilibrium as follows.[11]

Definition 1. An equilibrium is given by a set of stopping times $\mathcal{T} \subseteq [0, \infty)$ such that:

(i) If the organization has experimented unsuccessfully until time t, it continues to experiment ($t \notin \mathcal{T}$) if and only if m_t's payoff from the equilibrium continuation is greater than the payoff from switching to the safe policy, $\frac{s}{\gamma}$.

(ii) If m_t is indifferent because experimentation will stop immediately regardless of her action, but she strictly prefers experimentation (not) to continue for any length of time $\epsilon > 0$ small enough rather than stop, then she chooses (not) to continue experimenting.

To state Conditions (i) and (ii) more formally, it is useful to define the following value functions. Let $V_T(y)$ be the continuation utility of an agent with *current* belief y who expects experimentation to continue for a length of time T, counting from the present. Let $V(y)$ be the same agent's continuation utility if she expects experimentation to continue forever, i.e., $V(y) = \lim_{T\to\infty} V_T(y)$. Note that $V_T(y)$ and $V(y)$ are exogenous functions of the primitives, *not* equilibrium objects. (Explicit formulas are given in Lemma 2 in the Appendix.)

Then, at time t, m_t expects experimentation to continue until time $t' = \inf\{t'' > t : t'' \in \mathcal{T}\}$ if she does not stop. Condition (i) then requires that $t \in \mathcal{T}$ if $V_{t'-t}(p_t(m_t)) < \frac{s}{\gamma}$ and $t \notin \mathcal{T}$ if $V_{t'-t}(p_t(m_t)) > \frac{s}{\gamma}$. Condition (ii) requires that, if $t' = t$, then $t \in \mathcal{T}$ if $V_\epsilon(p_t(m_t)) < \frac{s}{\gamma}$ for all $\epsilon > 0$ small enough, and $t \notin \mathcal{T}$ if $V_\epsilon(p_t(m_t)) > \frac{s}{\gamma}$ for all $\epsilon > 0$ small enough.

Though part (i) of the definition is straightforward, it embeds an important assumption about the timing of policy and membership decisions: taking m_t to be pivotal at time t presumes that, for the organization to stop experimenting, a majority of those who chose to be members *under experimentation* must be in favor of stopping. We are thus implicitly

[11] In Appendix B, we provide a formal definition of equilibrium that includes full membership and policy strategies as primitives.

ruling out the possibility of a large number of detractors of the current policy coordinating to join the organization and immediately changing its policy.

One way to microfound this restriction would be to assume that agents only gain voting power with a time lag $\nu > 0$, so that it is in fact $m_{t-\nu}$ who chooses the policy at time t. In such a model, if the organization switched to the safe policy at time t_0 due to an "invasion" by pessimists, agents with no faith in the risky policy would strictly prefer to delay joining until t_0, and thus would not vote until $t_0 + \nu$, so the invasion would not actually materialize. As this argument applies for all $\nu > 0$, we require our equilibria to obey this property, even if we are in fact taking $\nu = 0$ for simplicity.

Condition (ii) imposes an additional tie-breaking rule in order to eliminate undesirable equilibria of the following variety. $\mathcal{T} = [0, \infty)$, for instance, satisfies Condition (i) vacuously even if experimentation is desired by all agents, because any agent who deviates and chooses experimentation would see her decision immediately overturned. To rule out such equilibria, we require optimal behavior even when the agent's policy choice only affects the path of play for an infinitesimal amount of time.[12]

3 Analysis

In this Section we characterize the set of equilibria of the baseline model. We first provide conditions under which perpetual experimentation is the unique equilibrium outcome, and then show the range of possible outcomes when these conditions are not met. Finally, we discuss the welfare properties of the model, and a simple extension with noisy news.

3.1 Perpetual Experimentation

It is useful to first note a few properties of our reduced-form model. First, the equilibrium stopping time, which we will denote by t_0, is the smallest (or, more generally, the infimal) element of $\mathcal{T}$. (If the risky policy is used forever, that is, $\mathcal{T} = \emptyset$, we write $t_0 = \infty$.) Other elements of $\mathcal{T}$ only serve to inform the agents' expectations about what will happen if they deviate. Second, the population dynamics implied by the optimal membership decisions are as follows. As long as no successes are observed, all agents become more pessimistic and the organization contracts. That is, $p_t(x)$ decreases in t for all x, and y_t increases towards 1. Of course, after a success or a switch to the safe policy, all agents join and remain members forever, and there is no further learning.

[12]Condition (ii) is in the spirit of weak dominance: an agent who prefers experimentation should experiment if she expects her successors to tremble and continue experimenting with some positive probability.

We can now state Proposition 1, which provides necessary and sufficient conditions for perpetual experimentation to arise in equilibrium.

Proposition 1. *Perpetual experimentation ($\mathcal{T} = \emptyset$) is an equilibrium of the game if and only if $V(p_t(m_t)) \geq \frac{s}{\gamma}$ for all t. It is the unique equilibrium if and only if the inequality is strict for all t.*

The first part of Proposition 1 is straightforward. Recall that m_t is the pivotal agent at time t under unsuccessful experimentation, and $p_t(m_t)$ is her posterior belief *when she is pivotal*. $V(p_t(m_t))$ is thus her expected continuation value when pivotal, if she chooses to continue experimenting and expects that no future pivotal agent will stop. $\frac{s}{\gamma}$, on the other hand, is her payoff if she stops. It follows that, if $V(p_t(m_t)) < \frac{s}{\gamma}$ for any t, then perpetual experimentation is not possible: if no one will stop experimenting after t, then m_t would herself make the choice to stop. On the other hand, if $V(p_t(m_t)) \geq \frac{s}{\gamma}$ for all t, then perpetual experimentation is an equilibrium by the same logic: if all pivotal agents expect experimentation to never end, they are reduced to making a binary choice between their respective $V(p_t(m_t))$ and $\frac{s}{\gamma}$, of which they weakly prefer the former.

What is less immediate is why, when perpetual experimentation is an equilibrium, it is the only one.[13] The key here is that if an agent prefers to experiment forever rather than not at all, she also prefers to experiment for any finite amount of time T rather than not at all. Thus, any pivotal agent m_t for whom $V(p_t(m_t)) > \frac{s}{\gamma}$ will *never* choose to halt experimentation in equilibrium, no matter what she conjectures that her successors will do.

The technical reason for this result is that $V_T(y)$ is (strictly) single-peaked in T. That is, letting $T^* = \operatorname{argmax}_T V_T(y)$ be the (finite) stopping time an agent would choose if she could control the policy at all times, her payoff decreases as T deviates from T^* in either direction. Since $V_0(p_t(m_t)) = \frac{s}{\gamma}$ and $\lim_{T\to\infty} V_T(p_t(m_t)) = V(p_t(m_t))$, it follows that, if $V(p_t(m_t)) > \frac{s}{\gamma}$, then $V_T(p_t(m_t)) > \frac{s}{\gamma}$ for any $T > 0$.

We prove the single-peakedness by calculating $V_T(y)$ explicitly (Lemmas 2 and 3 in the Appendix). But it is an intuitive result: $V_T(y)$ is what the agent would get from staying in the organization until her posterior reaches $\frac{a}{g}$ (assuming unsuccessful experimentation), and after that, staying out until there is a success or the safe policy is adopted (at time T). The higher T is, the more pessimistic the agent will be at time T, and the less she would want to prolong experimentation.

Figure 1 illustrates the equilibrium dynamics under perpetual experimentation, for the case $a = 1$, $s = 1.725$, $h = 1$, $\lambda = \gamma = 6$, and f uniform over $[0, 1]$. As the organization experiments unsuccessfully, all agents become more pessimistic. Denoting by x_t the agent

[13] There are multiple equilibria due to indifference if $\min_t V(p_t(m_t)) = \frac{s}{\gamma}$, but this is a knife-edge case.

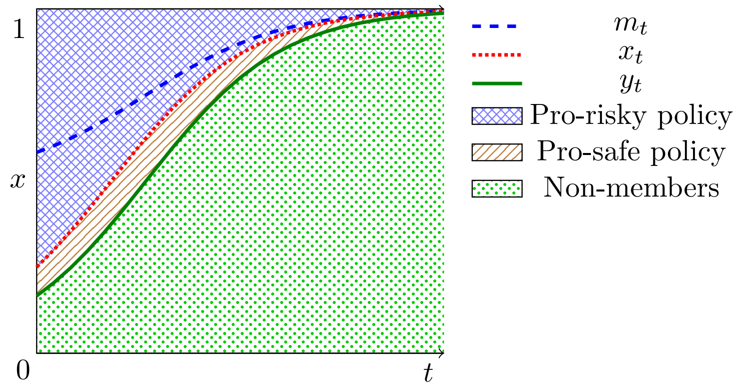

Figure 1: Pivotal member, indifferent agent, and marginal member on the equilibrium path

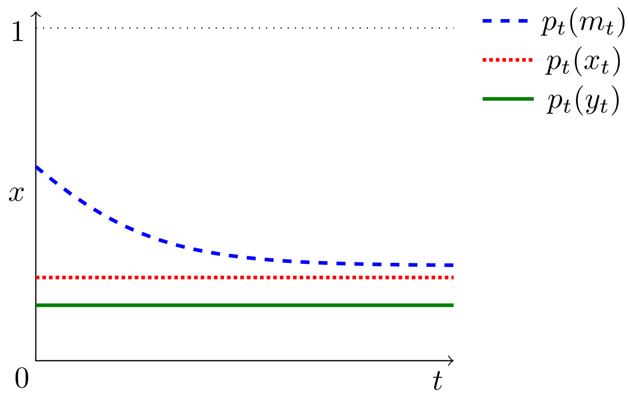

Figure 2: Posterior beliefs on the equilibrium path

indifferent about stopping experimentation at time t (defined by $V(p_t(x_t)) = \frac{s}{\gamma}$), this implies that x_t is increasing in t. Thus there is a shrinking mass of agents in favor of the risky policy (the agents shaded in crossed lines in Figure 1) and a growing mass against it (shaded in lines and dots). For high t, most agents want experimentation to stop.

Growing pessimism, however, induces members to leave. Hence the marginal member becomes more extreme, and so does the median member: as y_t increases, so does m_t. 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 risky policy always retains majority support in the organization due to most of the 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 the selection effect prevents $p_t(m_t)$ 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 pivotal member always remains optimistic enough to continue experimenting.

The result, perpetual experimentation, is clearly excessive: though in a world of heterogeneous priors agents disagree about the optimal length of experimentation, perpetual experimentation is excessive from the point of view of *all* agents except those with prior belief exactly equal to 1.

For what parameter values will $V(p_t(m_t))$ be greater than $\frac{s}{\gamma}$ for all t, leading to perpetual experimentation? Our next set of results aims to answer this question. Firstly, Proposition 2 provides either bounds or explicit closed-form expressions for $\inf_t V(p_t(m_t))$ for several families of prior belief distributions, allowing us to easily check the conditions of Proposition 1 in these cases. Secondly, the comparative statics established in Proposition 3 allow us to use the results Proposition 2 as bounds for *all* distributions.

Proposition 2. *The value function V in Proposition 1 satisfies the following:*

(i) If f is non-decreasing, then

$$\gamma \inf_{t\geq 0} V(p_t(m_t)) = \gamma V\left(\frac{2a}{g+a}\right) = \frac{2ga}{g+a} + \left(\frac{1}{2}\right)^{\frac{\gamma}{\lambda}} \frac{a(g-a)}{g+a}\frac{\lambda}{\gamma+\lambda}.$$

(ii) Suppose $f(x) = f_\omega(x) := (\omega+1)(1-x)^\omega$ for $x \in [0,1]$ and $f(x) = 0$ elsewhere, for any $\omega > 0$. Then, denoting $\eta = 2^{-\frac{1}{\omega+1}}$,

$$\gamma \inf_{t\geq 0} V(p_t(m_t)) = \gamma V\left(\frac{a}{\eta g + (1-\eta)a}\right) = \frac{ga}{\eta g + (1-\eta)a} + \eta^{\frac{\gamma+\lambda}{\lambda}} \frac{a(g-a)}{\eta g + (1-\eta)a}\frac{\lambda}{\gamma+\lambda}.$$

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

$$\gamma \inf_{t\geq 0} V(p_t(m_t)) \geq \gamma V\left(\frac{a}{g}\right) = a + \frac{a(g-a)}{g}\frac{\lambda}{\gamma+\lambda}.$$

In calculating the value of $\inf_t V(p_t(m_t))$, a key step is to find $\inf_t p_t(m_t)$, the infimal posterior belief of a pivotal agent on the equilibrium path. It is shown in part (i) that, if f is non-decreasing, then $\inf_t p_t(m_t) = \frac{2a}{g+a}$. To illustrate the derivation, suppose that f is uniform. Then $m_t \equiv \frac{1+y_t}{2}$, where $y_t = \frac{a}{a+(g-a)e^{-\lambda t}}$ as shown in Section 2. Thus $m_t = \frac{2a+(g-a)e^{-\lambda t}}{2a+2(g-a)e^{-\lambda t}}$ which, by way of Equation 1, implies that $p_t(m_t) = \frac{2a+(g-a)e^{-\lambda t}}{2a+(g-a)(1+e^{-\lambda t})}$, which converges to $\frac{2a}{g+a}$ from above as $t \to \infty$.

The case of a non-increasing density presumes that there are enough optimists in the population. Part (ii) shows that the faster $f(x)$ approaches 0 as $x \to 1$, the lower the value of $\inf_t p_t(m_t)$, as the median m_t is closer to the bottom of the interval $[y_t, 1]$. In particular, if $f(x) \sim (1-x)^\omega$, then $\inf_t p_t(m_t) = \frac{a}{\eta g+(1-\eta)a}$. At the other extreme, part (iii) gives a lower bound based on the principle that $p_t(m_t) \geq p_t(y_t) \equiv \frac{a}{g}$, no matter the shape of f.

The other step in the proof of Proposition 2 is to calculate $V(y)$ for a generic belief y. (A general formula is given in the Appendix.) The resultant expressions—for instance, the expression for $V\left(\frac{2a}{g+a}\right)$—have a natural interpretation. The first term, $\frac{2ga}{g+a}$, is the payoff the agent would get if she was locked into the organization forever: her posterior belief, $\frac{2a}{g+a}$, times the expected flow payoff g of the good risky policy. The second term is the option value of the agent's exit and reentry options.

The following corollary leverages Proposition 2.(iii) to highlight the importance of the gap between s and a:

Corollary 1. *If $a \in \left(\frac{s}{1+\frac{g-s}{g}\frac{\lambda}{\gamma+\lambda}}, s\right]$, there is perpetual experimentation.*

In other words, for *any* values of the other parameters, including the distribution of priors, if a is close enough to s—that is, if the organization's safe policy is not much better than the outside option—then the organization never stops experimenting. The reason is that the selection effect is at its strongest in this case, as most supporters of the safe policy are tempted to exit before their voices can make a difference.

Our next result concerns the comparative statics of our model.

Proposition 3. *If there is an equilibrium with perpetual experimentation under parameters $(\lambda, h, s, a, \gamma, f)$, then the same holds for any set of parameters $(\tilde{\lambda}, \tilde{h}, \tilde{s}, \tilde{a}, \tilde{\gamma}, \tilde{f})$ such that $\tilde{\lambda} \geq \lambda$, $\tilde{\lambda}\tilde{h} = \lambda h$, $\tilde{s} \leq s$, $\tilde{a} \geq a$, $\tilde{\gamma} \leq \gamma$ and $\tilde{f}$ MLRP-dominates f, i.e., $\frac{\tilde{f}(x)}{f(x)}$ is non-decreasing in x.*

The intuition is as follows. A decrease in s makes the safe policy less attractive and has no effect on the payoff from perpetual experimentation. A decrease in γ makes the agents more patient, which increases the option value of experimentation. An increase in λ while holding g fixed increases the learning rate, with similar consequences.[14] An increase in a has two effects that favor experimentation: it increases the expected payoff of experimentation (which entails collecting the outside option with some probability), and it induces agents to quit, leaving the organization with a more radical median member.

Finally, an increase in the number of optimists leaves the value function V and the marginal member y_t unchanged, but results in a more optimistic median—an m_t higher up within the interval $[y_t, 1]$—who is more likely to support experimentation. In particular, then, for any f that MLRP-dominates f_ω as defined in Proposition 2.(ii), $\inf V(p_t(m_t))$ is at least as high as the expression given in Proposition 2.(ii). We can thus give tighter bounds than the general bound in Proposition 2.(iii) whenever f decreases at a rate bounded by a power law.

[14] In contrast, the effect of an increase in h (holding λ constant) is ambiguous: while a higher payoff from the good risky policy encourages experimentation, it also discourages exit, weakening the selection effect. However, the first effect dominates for all the prior distributions covered in Proposition 2.(i)-(ii).

3.2 Finite Experimentation

If perpetual experimentation is not an equilibrium, there may be multiple equilibria featuring different levels of experimentation, supported by different off-path behavior.

To characterize them, it is useful to define a *stopping function* $\tau : [0, \infty) \to [0, \infty]$ as follows: for each $t \geq 0$, $\tau(t)$ is the highest $\tilde{t} \geq t$ such that m_t weakly prefers experimentation to continue until $\tilde{t}$, relative to stopping right away. In particular, $V_{\tau(t)-t}(p_t(m_t)) = \frac{s}{\gamma}$. (If the agent does not want to experiment at all, then $\tau(t) = t$, while if she would accept perpetual experimentation, then $\tau(t) = \infty$.) Proposition 4 characterizes the set of pure strategy equilibria in this setting.

Proposition 4.

(i) Any pure strategy equilibrium with finite experimentation, $\mathcal{T} \neq \emptyset$, must be a sequence of the form $(t_0, \tau(t_0), \tau(\tau((t_0)), \dots)$ for some $t_0 \leq \tau(0)$. The sequence may be finite, ending at a fixed point of τ, or infinite.

(ii) There exists $t_0 \in [0, \tau(0)]$ for which $(t_0, \tau(t_0), \dots)$ is an equilibrium.

(iii) If τ is non-decreasing and $\tau(0)$ is finite, then $(t_0, \tau(t_0), \dots)$ is an equilibrium for all $t_0 \in [0, \tau(0)]$.

(iv) If $\tau(t) = \infty$ for all $t \in [0, T]$, then $t_0 > T$ for any equilibrium stopping time t_0.

Part (i) describes the general structure of a non-empty set of equilibrium stopping times: each element t_n of the sequence must be chosen to leave the *previous* pivotal agent who stops in equilibrium, $m_{t_{n-1}}$, indifferent. The logic is that, if stopping times were any further apart (so that $m_{t_{n-1}}$ strictly wanted to stop, given the next expected stopping time), some later pivotal agent $m_{t_{n-1}+\epsilon}$ would also want to stop, by a continuity argument. Conversely, if they were any closer, $m_{t_{n-1}}$ would not stop at all, by the single-peakedness of V_T. Moreover, the initial (on-path) stopping time t_0 must be weakly before $\tau(0)$, as otherwise m_0 would deviate and stop right away.

Setting $t_0 \in [0, \tau(0)]$ and $t_n \equiv \tau^n(t_0)$ indeed guarantees that m_0 will not stop and that m_{t_n} is indifferent for all n. Part (iii) establishes that, under a regularity condition—if τ is increasing[15]—this is all we need for $\mathcal{T}$ to be an equilibrium, so every t_0 between 0 and $\tau(0)$ can be supported as a stopping time by a (unique) set of conjectures about off-path behavior. If τ is nonmonotonic, not all sequences of the form $(t, \tau(t), \dots)$ will be equilibria, because

[15] τ is guaranteed to be increasing if $p_t(m_t)$ does not decrease too steeply. For example, if $f(x) \propto \frac{1}{x^2}$ for all $x \geq \frac{a}{g}$, then $p_t(m_t) \equiv \frac{2a}{g+a}$ is constant, so $\tau(t) - t$ is constant, and τ is obviously increasing.

some pivotal agents between m_{t_n} and $m_{t_{n+1}}$ are more eager to stop than m_{t_n} is. But, by part (ii), there is always some t_0 for which this construction does yield an equilibrium.

It is easy to see that m_0's optimal stopping time lies between 0 and $\tau(0)$. Thus, from her point of view, both over and under-experimentation are possible depending on which equilibrium is played. Under-experimentation obtains if an early pivotal agent expects that, should she continue experimenting, the next stopping time will be too far in the future—that the organization will go down a "slippery slope" of excessive experimentation. In this scenario, the agent is compelled to stop experimentation while the decision is still in her hands, even at a time too early for her liking.[16]

Finally, part (iv) shows that perpetual experimentation is, in a sense, robust: if the condition $V(p_t(m_t)) > \frac{s}{\gamma}$ holds for all t up to some T, then experimentation must continue until at least T. (As noted previously, agents willing to experiment forever will *never* stop experimentation.) This implies that, if there is perpetual experimentation under a density $f(x)$, then there is *almost* perpetual experimentation (until an arbitrarily late T) under a truncated density of the form $f(x)\mathbb{1}_{x\leq 1-\epsilon}$ for $\epsilon > 0$ small enough. Selection forces can thus have powerful consequences even if the distribution of priors is bounded away from 1.

There are another two ways in which our results have bite even if the baseline model, as presented, is too stark to be realistic (in particular, as it assumes that the organization's size can contract to nothing in the limit). First, if the organization is forced to disband below a minimum size $S < 1$, then when this size is reached (i.e., for t such that $1-F(y_t) = S$) the safe policy would be adopted, but experimentation would continue until t under the conditions of Proposition 1. Second, all of our analysis is unchanged if the population is growing over time—for example, if at time t there are $e^{\alpha t}$ agents, with priors drawn from the density f, for some $\alpha > 0$. This assumption may be appropriate for countries undergoing political reforms, and is also applicable to startups, which may reach more and more potential employees and investors with each round of hiring and fundraising. In both cases, population growth may mask the effects of exit on size, at least temporarily.

3.3 Welfare

It is instructive to consider how the equilibrium and its welfare properties change as we vary the quality of the outside option, a. As a welfare benchmark, we focus on the equilibrium utility of the initial pivotal agent, m_0, net of the utility she could obtain if she controlled the policy at all times, $\max_T V_T(m_0)$. Figure 3 plots this quantity as a function

[16]This force is related to the cause of under-experimentation in Strulovici (2010) in that, in both cases, agents under-experiment to avoid a loss of control over future decisions. Similar concerns about slippery slopes are the focus of the clubs literature (Bai and Lagunoff (2011), Acemoglu et. al. (2015)).

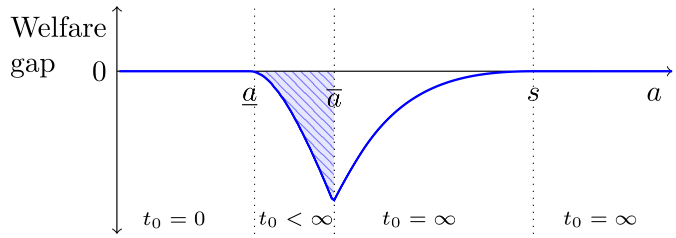

Figure 3: m_0's equilibrium welfare loss relative to her first-best stopping time

of a.[17] (Note that m_0 is itself a function of a.) The shaded region represents the range of welfare outcomes that obtain when multiple equilibria exist.

For $a \geq s$, the organization experiments forever, but this outcome is in fact optimal, as no agent is interested in the organization's safe policy. For $a \in [\overline{a}, s]$, we are in the world of Proposition 1: there is perpetual experimentation, which is excessive for all agents, in particular m_0. The lower a is, the earlier m_0 would want to halt experimentation, and the larger her welfare loss from over-experimentation. ($\overline{a}$ is defined such that $V(m_0) = \frac{s}{\gamma}$.)

For $a < \overline{a}$, m_0 will not tolerate perpetual experimentation, so the (multiple) equilibria feature finite experimentation. In this example, as τ is increasing, every $t_0 \in [0, \tau(0)]$ is an equilibrium stopping time. This range includes m_0's ideal stopping time, as well as extremes—0 and $\tau(0)$—that yield $\frac{s}{\gamma}$. Thus m_0's welfare loss can range from 0 to $\max_T V_T(m_0) - \frac{s}{\gamma}$. For low a, $\max_T V_T(m_0)$ is low, so the maximal welfare loss is lower as well. Finally, for $a \leq \underline{a}$, $\tau(t) \equiv t$, and nothing prevents m_0 from obtaining her optimal outcome by stopping right away. Thus, the welfare gap in the worst equilibrium is highest for intermediate values of a.

3.4 Public News

Finally, in a minor extension of the baseline model, we show that the organization can have a *perverse* response to information: it can, paradoxically, experiment *more* in the face of bad news. To see why, suppose that, at time 0, a preliminary test of the risky policy generates a binary public signal $\sigma \in \{0, 1\}$, where $1 > P[\sigma = 1|G] > P[\sigma = 1|B] > 0$. Agents enter or exit in response, and the organization decides whether to continue experimenting. Thereafter the game continues as in the baseline model.

Proposition 5. *If there are public news at time* 0*, there exist parameters for which the organization stops experimenting at a finite time after seeing* $\sigma = 1$*, but never stops after* $\sigma = 0$*—and, as a result, uses the risky policy more in expectation when it is bad than when it is good.*

[17] Here $s = 1.725$, $h = 1$, $\lambda = \gamma = 6$, and $f(x) \propto \frac{1}{x^2}$ for all $x \geq x_0$ with x_0 small, which guarantees that $p_t(m_t)$ is constant, τ is increasing, and Proposition 4.(iii) applies.

The intuition is simple: though a positive signal encourages experimentation, it also attracts skeptics who now favor the risky policy slightly over their outside option, but still rank the safe policy as the best choice. The latter effect dominates if f is high in a left-neighborhood of the initial marginal member, $\frac{a}{g}$. In that case, a measure of success can paradoxically lead the organization to turn away from the risky strategies that brought that very success. A salient version of this phenomenon is when the success of an innovative company invites an acquisition by a parent corporation that, not having fully understood what it bought, then begins to meddle in the company's affairs and dilutes its strategy.[18]

4 Extensions

In this Section, we extend our model to allow for heterogeneous payoffs across agents, as well as for continuous intensity of membership. Further extensions, discussed in the Conclusion, are relegated to Appendix B.

4.1 Heterogeneous Outcomes

The baseline model concerns groups and organizations that take action and distribute payoffs collectively. A related but distinct situation is when decisions are collective but payoffs are ex post heterogeneous. For example, agents may have hidden types: some may be "winners", destined to reap the eventual benefits from the risky policy if it is used, and others may be "losers", who will get nothing—but these types are only learned through experimentation. This setup, considered by Strulovici (2010), is a natural model of political or economic reforms: for example, when switching from capitalism to communism or from protectionism to free trade, citizens expect that some will benefit and others will suffer, but cannot predict who ex ante.

To adapt our model to this case, we now assume a population divided into $2K+1$ groups, each with unit mass, for some $K \in \mathbb{N}$. As before, individual agents can enter or leave the organization, the outside option pays a, and the safe policy pays s. But instead of the risky policy being good or bad for all agents, it is now either good or bad for each group i ($\vartheta_i = G, B$). Types are independent across groups; success realizations are independent across groups, but *common within each group*. That is, if group i is a "winner" from the risky policy then, while this policy is being used, the group experiences successes at rate λ, with each giving a lump sum h to *all* group-i agents in the organization. These assumptions mean

[18]For example, Pixar's success with boldly creative movies led to an acquisition by Disney, which then pressured Pixar to pivot to a "safer" strategy focused on sequels and franchises (Orr 2017).

that groups do not learn from each other but learning is perfectly shared within groups.[19]

The population of each group i is distributed according to a (common) density f with support $[0, 1]$. An agent's position now represents prior beliefs as follows: an agent $x \in [0, 1]$ in group i believes that $\vartheta_j = G$ with probability x for each j. (What matters is that x believes $\vartheta_i = G$ with probability x; agents' beliefs about other groups matter little.) We say group i is a "sure winner" if it has experienced a success.

An equilibrium can be described by a set of stopping states $\mathcal{T} \subseteq \mathbb{R} \times \mathbb{N}_0$, where $(t, k) \in \mathcal{T}$ means the organization switches to the safe policy at time t if there are k sure-winner groups at that time. We say there is perpetual experimentation if $\mathcal{T} = \emptyset$, i.e., experimentation never stops under *any* circumstances. Of course, experimentation must go on forever in *some* histories: for instance, if $k \geq K + 1$, a majority of sure winners will force experimentation on all other agents.

This model reduces to our baseline model when $K = 0$. It instead coincides with Strulovici (2010) when $a = 0$ and the distribution of priors is degenerate with $x \equiv p_0$ for all agents. A central result of that paper is that, for large K, the stopping time is finite, and approximately such that the unsure voters' posterior is $\frac{s}{g}$—in other words, fears of loss of control completely discourage experimenting for option value. Proposition 6 shows that adding heterogeneous beliefs and exit to Strulovici's model can dramatically change its results, making over-experimentation at least as likely as in our baseline model.

Proposition 6. *Perpetual experimentation is an equilibrium (and the only equilibrium) for the exact same parameter values as in our baseline model.*

The logic behind the result is as follows. Because some agents from each group always choose to remain and experiment, it is always possible for outsiders to learn about their group's type. Pessimists then leave when their own-group posteriors cross $\frac{a}{g}$, as in the baseline model. And, if perpetual experimentation is expected, any agent's continuation value $V(y)$ from experimentation is exactly the same as in the baseline model. The prior of the marginal and pivotal agents, y_t and m_t, is the same as in the baseline model if there are no sure winners, i.e., in state $(t, 0)$. When there are sure winners, all agents from those groups join and support the risky policy forever, pushing the beliefs of the median member upward. Thus the case $(t, 0)$ is the tightest, and the conditions for the baseline model, relevant for that case, also guarantee that experimentation will continue with any number of sure winners. The uniqueness result follows from a more involved version of the argument for Proposition 1, unraveling from the case of $K + 1$ sure-winner groups.

[19] An alternative way of modeling heterogeneous payoffs would give all individual agents independent types and successes, so the agent can only learn from herself, if she is a member. The results in that case track more closely with the no re-entry version of the model covered in Appendix B.

Even if the conditions for perpetual experimentation do not hold, the existence of sure winners can shift the balance of power further in favor of experimentation, as optimists can join forces with sure winners. For instance, in any history with even one sure-winner group, experimentation will never stop after $t^* = \frac{1}{\lambda} \ln \left((2K-1)\frac{g-a}{a} \right)$, assuming a non-increasing f. After this time, the sure-winner group will forever outnumber the remaining members from all unsure groups. Thus, if $V(p_t(m_t)) > \frac{s}{\gamma}$ for all $t \leq t^*$, there is infinite experimentation *with high probability* (that is, as long as any winners are revealed before t^*). In particular, there can be infinite experimentation with high probability even if the support of f is bounded away from 1, unlike in the baseline model.

4.2 Continuous Membership and Tradable Shares

Our baseline model highlights the effects of selection on experimentation under three important assumptions: membership is binary; the organization's size is flexible; and there are no property rights over the organization's future payoffs. These assumptions are appropriate for modeling political parties, social movements, or even firms in which the members with *de facto* influence over decision-making are its employees (e.g., a close-knit start-up).

In this section we present a variant of the model more applicable to a publicly-traded firm that is controlled by its shareholders. This model differs from the one in Section 2 in three respects. First, the "intensity" of membership is adjustable: agents may have ownership stakes of varying size. Second, entering or leaving the organization involves trading shares, and may entail capital gains and losses. Third, the size of the firm's operations (i.e., how much capital or labor it employs, how large its payoffs) is not a direct result of entry and exit, as investors trading on the secondary market cannot create or destroy shares.

We assume that voting power is proportional to stakes, so experimentation continues if a share-weighted majority desires it. To make the problem non-trivial, we assume that the agents are risk-averse, with CRRA utility functions.[20] To isolate the effects of continuous membership, we start with the assumption that the size of the firm is fixed.

The analysis yields two insights. First, continuous membership introduces another avenue for self-selection: even among those who own shares, more optimistic members want a larger stake. This effect intensifies as bad news arrive. Thus, the pivotal agent may become ex ante more optimistic over time even when the size of the firm is fixed, which could not happen in the baseline model. Second, selection forces are not strong enough in this setting to support perpetual experimentation. Paradoxically, this is true *even if the firm's size is flexible* as in

[20]Risk-neutrality leads to implausible results: the most optimistic agent would buy the entire firm. However, adding slight risk aversion to our baseline model would not qualitatively change the results, so it is informative to compare the results from Section 2 with the ones from this section.

the baseline model. The reason is that a success generates large capital gains in addition to the initial lump-sum payoff. This is an irreducible risky payoff which, in the limit, cannot be held by a vanishing share of the population.

In Appendix B, we show that perpetual experimentation is still possible if membership is continuous and agents are risk-averse, but members can enter and exit for free and are entitled only to current flow payoffs, as in the baseline model. In that case, there is, if anything, more selection than in the baseline model, due to selection on the intensive margin. The takeaway is that, while continuous membership gives even more power to optimists, capital markets have a corrective effect on selection forces, and may curb excessive experimentation.

The model is as follows. There is a firm, as before, and a continuum of agents distributed on $[0, 1]$ with density f. The firm's ownership is split into a unit mass of shares. There is a homogeneous good which can be consumed or used as capital. The firm owns a stock $\frac{a}{\gamma}$ of capital, and chooses at each time between a risky policy and a safe policy. Given this stock of capital, the safe policy generates a constant return of s. If the risky policy is good, it generates successes of size h at rate λ, where $g = \lambda h$. If it is bad, it never succeeds. (We are assuming that the firm's size—the amount of capital employed—is fixed. With a generic capital stock k, successes under the risky policy would pay $\frac{k\gamma}{a}h$ and the safe policy would pay $\frac{k\gamma}{a}s$.) These payoffs are distributed in proportion to shares. To simplify the analysis, we assume that, *after* the first success, the firm can offer its owners a constant flow payoff g rather than a stream of Poisson lump sums, e.g., by contracting with a risk-neutral insurer. (Without this assumption, the post-success share price would fluctuate due to wealth effects.)

Each agent starts with an endowment W_0 of the good.[21] An agent who consumes at a rate c obtains a flow utility $u(c) = \frac{c^{1-\theta}}{1-\theta}$, where $\theta \in (0, 1]$ is the agents' relative risk aversion.[22] Agents can lend or borrow the good at an interest rate γ, the same as their discount factor (possibly to or from unmodeled agents).[23] Agents can also buy or sell shares in a secondary market. Let ρ_t be the equilibrium price of a share, and let $q_t(x)$, $c_t(x)$, $W_t(x)$ be the demand for shares, consumption, and wealth of an agent with prior x, all at time t, under the assumption that the risky policy has been used until then with no success. Let $\overline{\rho} = \frac{g}{\gamma}$ be the equilibrium share price after a success, and $\underline{\rho} = \frac{s}{\gamma}$ the price after a switch to the safe policy. Let $c_t(x; \mathrm{succ})$ be x's (constant) consumption level after a success that occurred at time t.

An equilibrium is given by functions $(q_t(x), c_t(x), c_t(x; \mathrm{succ}), W_t(x), \rho_t)$ and a set $\mathcal{T}$ of stopping times such that the agents' consumption paths and share demands are utility-

[21] Or we could assume that each agent starts with an endowment W' and one share, and let $W_0 = W' + \rho_0$.

[22] In particular, we cover the case of $\theta > 0$ small, which approximates risk neutrality, and the case $\theta = 1$ of logarithmic preferences. The case $\theta > 1$ presents some technical differences but yields similar results.

[23] Note that if an agent could pull out her share of the firm's capital and consume a constant flow from it by lending, this would yield a consumption stream of size a.

maximizing given the price path, policy path, and budget constraints; the market for shares clears, that is, $\int_0^1 q_t(x)f(x)dx = 1$ for all t; and a majority at time t weakly prefers to switch to the safe policy if and only if t is a stopping time.

Proposition 7 provides a partial equilibrium characterization for this model.

Proposition 7.

(i) There is no equilibrium with perpetual experimentation.

(ii) In any equilibrium, $q_t(x)$ is weakly increasing in x for all t.

(ii) Moreover, if $\theta = 1$, then $q_t(x)$ is MLRP-increasing in t, x.

Part (ii) of Proposition 7 shows that optimists select into the organization, and part (iii) shows that this effect intensifies as time passes, in the case of logarithmic preferences.[24] This happens even though the firm is *not* shrinking operations as time passes to accommodate the shrinking number of optimists, as in the baseline model; instead, it is purely the result of selection on the intensive margin. An intuition is that share demands scale with each agent's posterior belief, $p_t(x)$, and more optimistic agents' beliefs are more resistant to bad news, that is, $\frac{p_t(x)}{p_t(x')}$ is increasing in t for $x > x'$.

Yet, per part (i) of Proposition 7, perpetual experimentation is impossible in this model: selection at the intensive margin has limits. A partial explanation is that, because the firm's size is constant, the per-capita share demands of ex post optimists must increase very quickly over time if they are to retain control forever. More precisely, at each time t, a population mass of approximate size $e^{-\lambda t}$ must hold a majority of all shares. Due to risk aversion, even optimists have diminishing returns from holding so many shares, as additional shares only pay off when these agents are already rich. Then the optimists' share demands can only be this high if shares are so cheap that less optimistic agents also want to hold some—and they then become the majority.

We might wonder, then, what happens if the firm did scale down in response to news as in Section 2. Formally, suppose that the firm could employ any capital stock $k \leq \frac{a}{\gamma}$, and it chose, at time t, to employ only a stock $0 < k_t < \frac{a}{\gamma}$ and return the rest to shareholders, with the intention of recapturing it (e.g., with a public offering) after a success or switch to the safe policy. Naïvely, we might think that if k_t decreases quickly enough, perpetual experimentation might result, as the total amount of risky payoffs to be held by each optimistic player could be kept bounded. But this intuition is incorrect.

[24] The logic is the same for all $\theta < 1$, but in the general case, the path of share demands is complex due to income effects: optimists want more shares proportional to their wealth, but they also over-consume in anticipation of a success, reducing their wealth in the long run.

Corollary 2. *For any path of the firm's capital stock* $(k_t)_t$ *such that* $k_t \xrightarrow[t\to\infty]{} 0$*, there is no equilibrium with perpetual experimentation.*

The reason is that, while the payoff generated by the *first* success may shrink, the fact remains that *after* a success, the firm would bounce back to full size, and the value of its shares would shoot up. Just these capital gains are enough to sustain the logic of Proposition 7.(i): the key is that voting power is tied to ownership not just of flow payoffs but also *future* payoffs. In contrast, in an organization run by members rather than owners, membership only entails exposure to current payoffs, and perpetual experimentation is possible even with continuous membership. (See Proposition 15 in Appendix B.)

Proposition 7.(i) can also be overturned if the population of agents is assumed to grow exponentially, at rate at least λ, over time. (That way, the total population of ex post optimists remains stable in the long run.) This assumption may plausibly model the growth phase of startups, in which they are continuously advertising and fundraising from broader pools of investors.

5 Conclusion

In this paper we have laid out a theory of learning and decision-making in organizations with endogenous membership. The most general principle emerging from our analysis is that self-selection of agents dampens and may even reverse the effect of news on the organization's collective beliefs, as well as its policy. The co-determination of policy and membership can induce path-dependence: firms in the same sector, or political parties with similar goals, may adopt different approaches which attract sets of members with diverging beliefs, giving rise to what may be seen as heterogeneous cultures. Culture thus defined may cause performance differences, and it may be persistent: unlike individual agents, two organizations that differ in their collective priors may fail to converge towards one another as information arrives.

As we have seen, the effects of self-selection are more severe the more feasible it is to exit. Capture by experimenters becomes even easier if ex post payoffs are heterogeneous, as optimists and sure winners can join forces. And if membership is a continuous choice, further selection occurs at the intensive margin. However, capture by a minority becomes more difficult when the controlling members are owners who must accept exposure to all future payoffs, as in the case of publicly traded firms.

In Appendix B, we show that our results are robust to several modifications of the model. Briefly, the analysis extends in straightforward fashion to more general voting rules, with supermajority requirements making perpetual experimentation even more likely. Results are similar if good news are imperfectly informative, i.e., if the bad risky policy also produces successes at a positive but lower rate. That case also yields an analogous result to Proposition 5: a streak of good news can paradoxically cause the risky policy to be abandoned. Perpetual experimentation can also obtain in a model of bad news—though this is less surprising, as even a single agent may want to experiment forever in a bad news environment. Our results also do not qualitatively change if the organization's payoffs, or its learning rate, are size-dependent—e.g., if there are (dis)economies of scale—or if agents differ in their valuations of the risky policy's output rather than their priors. If quitters cannot reenter, perpetual experimentation is still the equilibrium outcome, albeit for a smaller range of parameters.

Still other important extensions lie beyond the scope of the paper. For instance, organizations often choose between multiple risky policies. The same forces in our model may cause such an organization to switch too infrequently, or never, from an unsuccessful policy to an alternative, where a single agent would switch frequently to more promising policies. The general point, then, is more about rigidity than over-experimentation *per se*. Indeed, some famous examples of rigid decision-making are firms such as Blockbuster or Xerox that kept doubling down on an apparently "safe" policy that became increasingly unviable.[25]

Organizations also often compete with each other. As a result, the population they draw members from is also selected to be pessimistic about what *other* organizations are doing. When the strategies of competing organizations are in opposition, beliefs within an organization will be even more skewed towards optimism.

Finally, power is often in the hands of leaders and managers, even when they represent the interests of members. Such leaders ought to be cognizant of how success might attract "bandwagoners", and how a period of decline may render the organization increasingly sclerotic. The same dynamics, of course, affect the leaders' own ability to stay in power. An important question is under what conditions a leader would have incentives to encourage selection-induced inertia (as exemplified by the Curley effect) or to try and limit it.

Relatedly, we may ask how an organization could be *designed* to limit selection and policy inertia. To counteract inertia directly, supermajority rules should be avoided. On the contrary, it may be desirable to give greater weight to minorities in favor of policy changes. Alternatively, the organization could stabilize the voter base by making exit costly (e.g., with back-loaded pay, coercion, or by barring reentry), insulating (some) agents' payoffs from the outcome of its policy, or granting more voting power to senior members. One takeaway of Section 4.2 is that ownership by shareholders would also help curb selection effects, relative to a cooperative structure.

[25] Steve Jobs famously blamed the decline of Xerox on selection forces: namely, its focus on the copier market led to "product people", those with the sensibility to create new products, being "driven out of decision-making forums" and replaced by "toner-heads" who saw no need for innovation, even as the early PC market was booming (Tweedie 2014).

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.

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 and Byron Criddle, "The Decline of the French Communist Party," *British Journal of Political Science*, 1989, *19* (4), 515–536.

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.

Bishin, Benjamin G and Casey A Klofstad, "The Political Incorporation of Cuban Americans: Why Won't Little Havana Turn Blue?," *Political Research Quarterly*, 2012, *65* (3), 586–599.

Carreyrou, John, *Bad Blood: Secrets and Lies in a Silicon Valley Startup*, Penguin Random House, 5 2018.

Chen, Carolyn, "When Your Job Fills In for Your Faith, That's a Problem," 2022. `https://www.nytimes.com/2022/05/24/opinion/work-religion-silicon-valley.html`.

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.

Dong, Miaomiao, "Strategic Experimentation with Asymmetric Information," 2021.

Dowty, Alan, "The New Serfdom: Contemporary Control of Emigration and Expulsion," *Humboldt Journal of Social Relations*, 1988, *15* (2), 1–27.

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

Garde, Damian, "Ego, Ambition, and Turmoil: Inside One of Biotech's Most Secretive Startups," 2016. `https://www.statnews.com/2016/09/13/moderna-therapeutics-biotech-mrna/`.

Gieczewski, Germán, "Policy Persistence and Drift in Organizations," *Econometrica*, 2021, *89* (1), 251–279.

Glaeser, Edward L. and Andrei Shleifer, "The Curley Effect: The Economics of Shaping the Electorate," *Journal of Law, Economics, and Organization*, 2005, *21* (1), 1–19.

Hazareesingh, Sudhir, *Intellectuals and the French Communist Party: Disillusion and Decline*, Oxford, UK: Clarendon Press, 1991.

Hirschman, Albert O., *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*, Cambridge, Massachusetts: Harvard University Press, 1970.

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.

Meredith, Martin, *Our Votes, Our Guns: Robert Mugabe and the Tragedy of Zimbabwe*, New York: Public Affairs, 2002.

Orr, Christopher, "How Pixar Lost Its Way," 2017. `https://www.theatlantic.com/magazine/archive/2017/06/how-pixar-lost-its-way/524484/`.

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

Ross, George, "Party Decline and Changing Party Systems: France and the French Communist Party," *Comparative Politics*, 1992, *25* (1), 43–61.

Sellars, Emily A, "Emigration and Collective Action," *The Journal of Politics*, 2019, *81* (4), 1210–1222.

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

Tweedie, Steven, "Here's Why Innovation Dies At Big Tech Companies, According To Steve Jobs," 2014. `https://www.businessinsider.com/steve-jobs-on-why-innovation-dies-at-tech-monopolies-2014-11`.

A Proofs

We begin with a few preliminaries regarding the evolution of the agents' beliefs, their quitting times, and the shape of the value functions $V(y)$ and $V_T(y)$.

Lemma 1. *Let $t^z(y)$ denote the time it takes for an agent's posterior belief to go from y to z under unsuccessful experimentation. In particular, let $t(y) = t^{\frac{a}{g}}(y)$ be the time it will take for an agent with current belief y to quit. Then*

$$t^z(y) = \frac{1}{\lambda}\ln\left(\frac{1-z}{z}\frac{y}{1-y}\right) \qquad t(y) = \frac{1}{\lambda}\ln\left(\frac{g-a}{a}\frac{y}{1-y}\right)$$

If $y = \frac{2a}{g+a}$, then $e^{-\lambda t(y)} = \frac{1}{2}$. If $y = \frac{a}{\eta g+(1-\eta)a}$, then $e^{-\lambda t(y)} = \eta$.

Proof of Lemma 1. Solving $p_t(y) = \frac{ye^{-\lambda t}}{ye^{-\lambda t}+1-y} = z$ for t, we obtain $e^{-\lambda t^z(y)} = \frac{z}{1-z}\frac{1-y}{y}$ or, equivalently, $t^z(y) = \frac{1}{\lambda}\ln\left(\frac{1-z}{z}\frac{y}{1-y}\right)$. The rest are special cases. ■

Lemma 2. *The value functions $V_T(y)$, $V(y)$ satisfy the following equations:*

$$V_T(y) = y\left[\frac{g}{\gamma} - e^{-(\lambda+\gamma)T}\frac{g-s}{\gamma}\right] + (1-y)e^{-\gamma T}\frac{s}{\gamma} \qquad \text{if } T \leq t(y). \tag{2}$$

$$\begin{aligned} V_T(y) =& y\left[\frac{g}{\gamma} - \frac{g-a}{\lambda+\gamma}e^{-(\lambda+\gamma)t(y)} + \left(\frac{s-g}{\gamma} + \frac{g-a}{\lambda+\gamma}\right)e^{-(\lambda+\gamma)T}\right] + \\ &+ (1-y)\left[\frac{a}{\gamma}e^{-\gamma t(y)} + \frac{s-a}{\gamma}e^{-\gamma T}\right] \qquad \text{if } T > t(y). \end{aligned} \tag{3}$$

$$V(y) = y\left[\frac{g}{\gamma} - \frac{g-a}{\lambda+\gamma}e^{-(\lambda+\gamma)t(y)}\right] + (1-y)\frac{a}{\gamma}e^{-\gamma t(y)}. \tag{4}$$

Proof of Lemma 2. If $T \leq t(y)$, the agent never leaves the organization. Then

$$V_T(y) = y\left[\int_0^T ge^{-\gamma t}dt + \int_T^\infty \left(e^{-\lambda T}s + \left(1-e^{-\lambda T}\right)g\right)e^{-\gamma t}dt\right] + (1-y)\int_T^\infty se^{-\gamma t}dt.$$

The first term is the agent's utility conditional on the risky policy being good. Between 0 and T, she collects an expected flow payoff g. At time T, there is a probability $e^{-\lambda T}$ that no successes have occurred, in which case the safe policy is chosen and the agent receives s thereafter. With probability $1-e^{-\lambda T}$, a success has occurred, so the risky policy is retained forever and the agent receives g. The second term is the agent's utility in the bad state of the world: a flow payoff s after the switch to the safe policy. Simplifying, we obtain Equation 2.

If $T > t(y)$, the agent leaves before the switch to the safe policy. Then

$$\begin{aligned}V_T(y) =& y\left[\int_0^{t(y)} ge^{-\gamma t}dt + \int_{t(y)}^{T} \left(e^{-\lambda t}a + \left(1 - e^{-\lambda t}\right) g\right) e^{-\gamma t}dt + \right. \\ & \left. \int_T^{\infty} \left(e^{-\lambda T}s + \left(1 - e^{-\lambda T}\right) g\right) e^{-\gamma t}dt\right] + (1-y)\left[\int_{t(y)}^{T} ae^{-\gamma t}dt + \int_T^{\infty} se^{-\gamma t}dt\right].\end{aligned}$$

The only difference is that, between $t(y)$ and T, the agent receives a if there have been no successes. Simplifying yields Equation 3. Finally, we can obtain Equation 4 by taking the limit of $V_T(y)$ as $T \to \infty$. ∎

Lemma 3. *(i) $V_T(y)$ and $V(y)$ are continuous and strictly increasing in y, and differentiable at all $T \neq t(y)$.*

(ii) $V_0(y) = \frac{s}{\gamma}$ and $\left.\frac{\partial_+ V_T(y)}{\partial T}\right|_{T=0} = \max\{yg, a\} - s + y\frac{\lambda(g-s)}{\gamma}$.

(iii) Letting $T^* = \operatorname{argmax}_T V_T(y)$, $T \mapsto V_T(y)$ is strictly increasing for $T \in [0, T^*]$ and strictly decreasing for $T > T^*$.

(iv) If $V(y) > \frac{s}{\gamma}$, then $V_T(y) > \frac{s}{\gamma}$ for all $T > 0$.

Proof of Lemma 3. The continuity and differentiability of $V_T(y)$, $V(y)$ are immediate consequences of Lemma 2. That these functions are increasing in y can be proved directly, by differentiating Equations 2–4 with respect to y, but it is also conceptually obvious, as an agent with a higher prior can copy the behavior of one with a lower prior and still obtain a higher expected payoff.

For part (ii), that $V_0(y) = \frac{s}{\gamma}$ follows from the definition. For the rest of part (ii), note that if $yg > a$ then $t(y) > 0$, so $\left.\frac{\partial_+ V_T(y)}{\partial T}\right|_{T=0} = yg - s + y\frac{\lambda(g-s)}{\gamma}$ can be obtained by differentiating Equation 2 with respect to T at $T = 0$. If $yg < a$ then $t(y) < 0$, and $\left.\frac{\partial_+ V_T(y)}{\partial T}\right|_{T=0} = a - s + y\frac{\lambda(g-s)}{\gamma}$ follows analogously from Equation 3. Note in particular that $\left.\frac{\partial_+ V_T(y)}{\partial T}\right|_{T=0}$ is strictly increasing in y.

For part (iii), we will relate the values of $V_T(y)$ for different values of T and y as follows. Fix $T_0 \geq 0$ and $\epsilon > 0$. Then

$$V_{T_0+\epsilon}(y) - V_{T_0}(y) = e^{-\gamma T_0}(ye^{-\lambda T_0} + 1 - y)\left(V_\epsilon(p_{T_0}(y)) - \frac{s}{\gamma}\right),$$

since $V_{T_0+\epsilon}(y)$ and $V_{T_0}(y)$ only differ in the event that T_0 is reached with no successes—an event with probability $(ye^{-\lambda T_0} + 1 - y)$—and, in this scenario, they yield the respective continuation values $V_\epsilon(p_{T_0}(y))$ and $\frac{s}{\gamma}$, starting from T_0. Taking the limit as $\epsilon \to 0$,

$$\left.\frac{\partial V_T(y)}{\partial T}\right|_{T=T_0} = e^{-\gamma T_0}(ye^{-\lambda T_0} + 1 - y)\left.\frac{\partial V_T(p_{T_0}(y))}{\partial T}\right|_{T=0}.$$

This implies that $\left.\frac{\partial V_T(y)}{\partial T}\right|_{T=T_0}$ is positive (negative) whenever $\left.\frac{\partial V_T(p_{T_0}(y))}{\partial T}\right|_{T=0}$ is positive (negative). In addition, we know that $T_0 \mapsto p_{T_0}(y)$ is decreasing by Equation 1, and $z \mapsto \left.\frac{\partial V_T(z)}{\partial T}\right|_{T=0}$ is increasing by part (i). Moreover, $\left.\frac{\partial V_T(p_{T_0}(y))}{\partial T}\right|_{T=0}$ is negative for large enough T_0, as $p_{T_0}(y)$ tends to zero for all $y < 1$. Thus $\left.\frac{\partial V_T(y)}{\partial T}\right|_{T=T_0}$ is either always negative or changes signs once from positive to negative. In the first case, $T^* = 0$. In the second, T^* is the unique solution to $\left.\frac{\partial V_T(y)}{\partial T}\right|_{T=T^*} = 0$. Either way, $V_T(y)$ is single-peaked in T. Intuitively, the higher is T_0, the more pessimistic the agent is at the stopping time, and the less she wants to prolong experimentation at the margin.

Hence, if $T^* > 0$, $V_T(y) > V_0(y) = \frac{s}{\gamma}$ for $T \in (0, T^*]$ because $T \mapsto V_T(y)$ is increasing over this region. For $T > T^*$, $V_T(y) \geq \lim_{T\to\infty} V_T(y) = V(y)$ because $T \mapsto V_T(x)$ is decreasing over this region. Then, if $V(y) > \frac{s}{\gamma}$, $V_T(y) > \frac{s}{\gamma}$ for $T > T^*$, and of course $T^* > 0$, so part (iv) follows. ■

Proof of Proposition 1. Suppose that all pivotal agents expect perpetual experimentation in equilibrium ($\mathcal{T} = \emptyset$). Then, when m_t is pivotal, she expects a payoff $V(p_t(m_t))$ from continuing to experiment and a payoff $\frac{s}{\gamma}$ from stopping. If $V(p_t(m_t)) \geq \frac{s}{\gamma}$ for all t, then it is weakly optimal for all pivotal agents to continue, and hence $\mathcal{T} = \emptyset$ is an equilibrium. Conversely, if $V(p_t(m_t)) < \frac{s}{\gamma}$ for some t, $\mathcal{T} = \emptyset$ cannot be an equilibrium as m_t would deviate to the safe policy.

As for the uniqueness, if $V(p_t(m_t)) \geq \frac{s}{\gamma}$ for all t with equality for some t, we can also construct an equilibrium with stopping at t and nowhere else. It is left to prove that, if $V(p_t(m_t)) > \frac{s}{\gamma}$ for all t, then there are no other equilibria besides perpetual experimentation. Suppose for the sake of contradiction that there is an equilibrium $\mathcal{T} \neq \emptyset$. Choose any $t \in \mathcal{T}$, and let $t' = \inf\{\tilde{t} \in \mathcal{T} : \tilde{t} > t\}$ be the "next" stopping time after t. If $t' = \infty$, then m_t would receive $V(p_t(m_t))$ from continuing and $\frac{s}{\gamma}$ from stopping, so she strictly prefers to continue, a contradiction. If $t' > t$ is finite, we similarly have a contradiction because, by Lemma 3.(iv), $V(p_t(m_t)) > \frac{s}{\gamma}$ implies $V_{t'-t}(p_t(m_t)) > \frac{s}{\gamma}$. Finally, if $t' = t$, then m_t's payoffs from continuing and stopping coincide. However, by Lemma 3.(iv), $V(p_t(m_t)) > \frac{s}{\gamma}$ implies $V_\epsilon(p_t(m_t)) > \frac{s}{\gamma}$ for all $\epsilon > 0$, whence m_t must choose to continue by Condition (ii), a contradiction. ■

Proof of Proposition 2.

We prove each inequality in two steps. First, we note that the median posterior belief,

$p_t(m_t)$, is uniformly bounded below for all t, with different bounds depending on the density f. For any f with full support, $p_t(m_t) \geq p_t(y_t) = \frac{a}{g}$. When f is uniform, $p_t(m_t) \searrow \frac{2a}{g+a}$, as shown in the text. More generally, for any $\omega > 0$, if $f(x) = f_\omega(x) \equiv (\omega+1)(1-x)^\omega$ then $p_t(m_t) \searrow \frac{a}{\eta g+(1-\eta)a}$, where $\eta = 2^{-\frac{1}{\omega+1}}$. This is a consequence of the following claim:

Claim 1. Suppose that the distribution of priors is f_ω for some $\omega \geq 0$. Then

$$p_t(m_t) = \frac{a + (1-\eta)(g-a)e^{-\lambda t}}{\eta(g-a) + a + (1-\eta)(g-a)e^{-\lambda t}}.$$

Proof of Claim 1. As shown in the text, $y_t = \frac{a}{a+(g-a)e^{-\lambda t}}$. The median m_t is such that $2\int_{m_t}^1 f_\omega(x)dx = \int_{y_t}^1 f_\omega(x)dx$, so that $2(1-m_t)^{\omega+1} = (1-y_t)^{\omega+1}$. Hence $1-m_t = \eta(1-y_t)$, which implies that

$$m_t = 1-\eta+\eta y_t = 1-\eta+\eta\frac{a}{a+(g-a)e^{-\lambda t}} = \frac{a+(1-\eta)(g-a)e^{-\lambda t}}{a+(g-a)e^{-\lambda t}}.$$

Substituting this expression into Equation 1 yields the result. ■

It is then immediate that $p_t(m_t) \searrow \frac{a}{\eta g+(1-\eta)a}$ when $f = f_\omega$.

Second, we observe that, since $V(y)$ is strictly increasing and continuous in y (by Lemma 3.(i)), 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.

We begin with part (ii). To calculate $V\left(\frac{a}{\eta g+(1-\eta)a}\right)$, we combine the results of Lemma 1 and Lemma 2. Substituting $y = \frac{a}{\eta g+(1-\eta)a}$ and $e^{-\lambda t(y)} = \eta$ into Equation 4, we obtain

$$\begin{aligned} V\left(\frac{a}{\eta g+(1-\eta)a}\right) &= \frac{a}{\eta g+(1-\eta)a}\left[\frac{g}{\gamma} - \frac{g-a}{\lambda+\gamma}\eta^{1+\frac{\gamma}{\lambda}}\right] + \frac{\eta(g-a)}{\eta g+(1-\eta)a}\frac{a}{\gamma}\eta^{\frac{\gamma}{\lambda}} \\ &= \frac{a}{\eta g+(1-\eta)a}\frac{g}{\gamma} + \eta^{1+\frac{\gamma}{\lambda}}(g-a)\frac{a}{\eta g+(1-\eta)a}\left[-\frac{1}{\lambda+\gamma}+\frac{1}{\gamma}\right] \end{aligned}$$

which, after rearranging and multiplying both sides by γ, yields the equation from part (ii). For part (i), calculating $V\left(\frac{2a}{g+a}\right)$ is a special case of part (ii), with $\omega = 0$ and hence $\eta = \frac{1}{2}$. For part (iii), we substitute $y = \frac{a}{g}$ and $t(y) = 0$ into Equation 4 to obtain the value of $V\left(\frac{a}{g}\right)$.

The only thing left to do is show that part (i) holds for all non-decreasing f, not just when f is uniform. 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. We will show that $\inf_t p_t(\tilde{m}_t) = \inf_t p_t(m_t) = \frac{2a}{g+a}$, which of course implies that $\inf_t V(p_t(\tilde{m}_t)) = \inf_t V(p_t(m_t))$, as desired. To do this, we will need three auxiliary results.

Lemma 4. *Suppose that $\tilde{f}$ MLRP-dominates f, i.e., $\frac{\tilde{f}(x)}{f(x)}$ is non-decreasing in x. Let $\tilde{m}_t$ and m_t be the median members at t under each respective density. Then $\tilde{m}_t \geq m_t$ for all t.*

Proof of Lemma 4. Note that y_t, the prior of the indifferent agent at time t, is independent of the distribution of priors. By definition, $\int_{y_t}^{m_t} f(x)dx = \int_{m_t}^{1} f(x)dx$. Suppose that $\tilde{m}_t < m_t$ for some t. This is equivalent to

$$\int_{y_t}^{m_t} \frac{\tilde{f}(x)}{f(x)} f(x)dx = \int_{y_t}^{m_t} \tilde{f}(x)dx > \int_{m_t}^{1} \tilde{f}(x)dx = \int_{m_t}^{1} \frac{\tilde{f}(x)}{f(x)} f(x)dx.$$

Since $\frac{\tilde{f}(x)}{f(x)}$ is weakly increasing,

$$\int_{y_t}^{m_t} \frac{\tilde{f}(m_t)}{f(m_t)} f(x)dx \geq \int_{y_t}^{m_t} \frac{\tilde{f}(x)}{f(x)} f(x)dx > \int_{m_t}^{1} \frac{\tilde{f}(x)}{f(x)} f(x)dx \geq \int_{m_t}^{1} \frac{\tilde{f}(m_t)}{f(m_t)} f(x)dx$$

which is a contradiction. ■

Lemma 5. *Suppose $\tilde{f}$ is a non-decreasing density, and f is the uniform density over* $[0, 1]$. *Then $\frac{1-\tilde{m}_t}{1-m_t} \to 1$ as $t \to \infty$.*

Proof of Lemma 5.

Let $\tilde{f}_{0t} = \tilde{f}(y_t)$ and $\tilde{f}_1 = \tilde{f}(1)$. Suppose $\tilde{f}$ is continuous at 1. (If not, redefine $\tilde{f}(1)$ as $\sup_{x\in[0,1)} \tilde{f}(x)$, which does not alter $\tilde{m}_t$.) By the same logic as in Lemma 4, we have $m_t \leq \tilde{m}_t \leq \hat{m}_t$, where $\hat{m}_t$ is the median corresponding to a density $\hat{f}$ such that $\hat{f}(x) = \tilde{f}_{0t}$ for $x \in [y_t, \hat{m}_t]$ and $\hat{f}(x) = \tilde{f}_1$ for $x \in [\hat{m}_t, 1]$. Then $\frac{1-\hat{m}_t}{1-m_t} \leq \frac{1-\tilde{m}_t}{1-m_t} \leq 1$, so it is enough to show that $\frac{1-\hat{m}_t}{1-m_t} \to 1$.

By construction, because $\hat{m}_t$ is the median, we have $\tilde{f}_{0t}(\hat{m}_t - y_t) = \tilde{f}_1(1 - \hat{m}_t)$, so $\hat{m}_t = \frac{\tilde{f}_{0t}y_t+\tilde{f}_1}{\tilde{f}_{0t}+\tilde{f}_1}$. Thus $1 - \hat{m}_t = \frac{\tilde{f}_{0t}(1-y_t)}{\tilde{f}_{0t}+\tilde{f}_1}$ and, because $m_t = \frac{y_t+1}{2}$ and $1 - m_t = \frac{1-y_t}{2}$, we have $\frac{1-\hat{m}_t}{1-m_t} = \frac{2\tilde{f}_{0t}}{\tilde{f}_{0t}+\tilde{f}_1}$.

Since $\tilde{f}$ is continuous at 1, $\tilde{f}(x) \to \tilde{f}(1)$ as $x \to 1$. Then, as $t \to \infty$, $y_t \to 1$, $\tilde{f}_{0t} = f(y_t) \to \tilde{f}_1$ and $\frac{1-\hat{m}_t}{1-m_t} \to 1$. ■

Lemma 6. *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 6.

Applying Equation 1, we obtain

$$\frac{p_t(\tilde{x}_t)}{p_t(x_t)} = \frac{\tilde{x}_t e^{-\lambda t}}{\tilde{x}_t e^{-\lambda t} + (1 - \tilde{x}_t)} \frac{x_t e^{-\lambda t} + (1 - x_t)}{x_t e^{-\lambda t}} = \frac{\tilde{x}_t}{x_t} \frac{x_t + (1 - x_t)e^{\lambda t}}{\tilde{x}_t + (1 - \tilde{x}_t)e^{\lambda t}}.$$

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 $\frac{1-x_t}{1-\tilde{x}_t} \to 1$, $\frac{(1-x_t)e^{\lambda t}}{(1-\tilde{x}_t)e^{\lambda t}} \to 1$. The result then follows, as

$$1 \leftarrow \min\left\{\frac{x_t}{\tilde{x}_t}, \frac{(1-x_t)e^{\lambda t}}{(1-\tilde{x}_t)e^{\lambda t}}\right\} \leq \frac{x_t + (1-x_t)e^{\lambda t}}{\tilde{x}_t + (1-\tilde{x}_t)e^{\lambda t}} \leq \max\left\{\frac{x_t}{\tilde{x}_t}, \frac{(1-x_t)e^{\lambda t}}{(1-\tilde{x}_t)e^{\lambda t}}\right\} \to 1.$$

■

Lemma 5 shows that $\frac{1-\tilde{m}_t}{1-m_t} \to 1$, while Lemma 4 shows that $\tilde{m}_t \geq m_t$ for all t. And, of course, $m_t \to 1$ as $t \to \infty$. Then Lemma 6 applies to the sequences $\tilde{m}_t$ and m_t, guaranteeing that $\frac{p_t(\tilde{m}_t)}{p_t(m_t)} \to 1$ and hence $p_t(\tilde{m}_t) \to \frac{2a}{g+a}$. In particular, $\inf_t p_t(\tilde{m}_t) \leq \frac{2a}{g+a}$. Since $p_t(\tilde{m}_t) \geq p_t(m_t)$ for all t by Lemma 4, $\inf_t p_t(\tilde{m}_t) \geq \inf_t p_t(m_t) = \frac{2a}{g+a}$, concluding the proof. ■

Proof of Corollary 1. This follows from Proposition 2.(iii): if f has full support and $a \in \left(\frac{s}{1+\frac{g-s}{g}\frac{\lambda}{\gamma+\lambda}}\right]$, then

$$\gamma \inf_{t\geq 0} V(p_t(m_t)) \geq \gamma V\left(\frac{a}{g}\right) = a + \frac{a(g-a)}{g}\frac{\lambda}{\gamma+\lambda} \geq a + \frac{a(g-s)}{g}\frac{\lambda}{\gamma+\lambda} > s.$$

■

Proof of Proposition 3. Lemma 4 implies that an MLRP-increase in f increases $\inf_t V(p_t(m_t))$. An increase in γ decreases $\gamma V(y)$ by reducing the agent's option value from experimentation, while leaving s unchanged. We can verify this by differentiating Equation 4 with respect to γ:

$$\begin{aligned}\frac{\partial\left[\gamma V(y)\right]}{\partial\gamma} &= y(g-a)\frac{\gamma}{\lambda+\gamma}e^{-(\lambda+\gamma)t(y)}t(y) - \frac{y(g-a)\lambda}{(\lambda+\gamma)^2}e^{-(\lambda+\gamma)t(y)} - (1-y)ae^{-\gamma t(y)}t(y)\\ &= (1-y)a\left[\frac{\gamma}{\lambda+\gamma} - 1\right]e^{-\gamma t(y)}t(y) - \frac{y(g-a)\lambda}{(\lambda+\gamma)^2}e^{-(\lambda+\gamma)t(y)} < 0,\end{aligned}$$

where we have used that $e^{-\lambda t(y)} = \frac{a}{g-a}\frac{1-y}{y}$ by Lemma 1. An increase in λ with a proportional decrease in h (so g remains unchanged) is formally equivalent to a decrease in γ up to a relabeling of the time variable, so it has the same effects.

An increase in a increases $V(y)$ for each y (this can be proved by differentiating Equation 4), and also increases y_t, and hence m_t, for each t. The effect of a change in s is straightforward since it has no impact on $V(y)$. ■

Proof of Proposition 4. We first note some properties of τ. Let t be the current time and t^* be the time at which m_t would choose to stop experimenting if she had complete control over the policy. In other words, $t^* = \operatorname{argmax}_T V_{T-t}(x)$.

If $t^* = t$ then, by Lemma 3, $V_{T-t}(x) < \frac{s}{\gamma}$ for all $T > t$, and $\tau(t) = t$. If $t^* > t$ and

$V(p_t(m_t)) < \frac{s}{\gamma}$, then, by the same lemma, $V_{T-t}(p_t(m_t))$ crosses $\frac{s}{\gamma}$ only once, at a value of $T > t^*$ equal to $\tau(t)$. Finally, if $t^* > t$ and $V(p_t(m_t)) \geq \frac{s}{\gamma}$, then Lemma 3 implies that $V_{T-t}(p_t(m_t)) > \frac{s}{\gamma}$ for all $T > t$, so $\tau(t) = \infty$.

Next, we argue 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 $V_{\tau(t)-t}(p_t(m_t)) = \frac{s}{\gamma}$, where $p_t(m_t)$ is differentiable in t, and $V_T(x)$ is differentiable in (T, x) at $(T, x) = (\tau(t), p_t(m_t))$ (by Lemma 2) and strictly decreasing in T (by Lemma 3), so the continuity of τ follows from the Implicit Function Theorem. The case $\tau(t_0) = t_0$ is similar. τ is also continuous at ∞ if we take the one-point compactification topology on $[0, \infty]$.

Consider a pure strategy equilibrium with finite experimentation, $\mathcal{T} \neq \emptyset$. Let $t_0 = \inf \mathcal{T}$ be the stopping time on the equilibrium path. Clearly $t_0 \leq \tau(0)$, as otherwise m_0 would switch to the safe policy at time 0.

Suppose $t_0 \in \mathcal{T}$. Consider what happens at time t_0 if m_{t_0} deviates and continues experimenting. Suppose first that $\tau(t_0) \in (t_0, \infty)$. Let $t_1 = \inf (\mathcal{T} \cap (t_0, \infty))$ be the time when experimentation stops in this continuation. We claim that t_1 must equal $\tau(t_0)$. To see why, suppose that $t_1 > \tau(t_0)$. In this case, for all $\epsilon > 0$ sufficiently small, $m_{t_0+\epsilon}$ would strictly prefer to stop experimenting, which contradicts the assumption that $t_1 > \tau(t_0) > t_0$ was the first stopping time after t_0. On the other hand, if $t_1 < \tau(t_0)$, then m_{t_0} would strictly prefer to deviate from the equilibrium path and not stop. (If $t_1 = t_0$, m_{t_0} would still deviate and not stop by Condition (ii).)

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

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

Finally, if $t_0 \notin \mathcal{T}$, then must be a sequence $(t^k)_k \subseteq \mathcal{T}$ such that $t^k \searrow t_0$. Applying the previous argument to m_{t^k}'s stopping decision, we conclude that $\tau(t^k) \leq t^{k-1}$ (else m_{t^k} would deviate). Taking the limit yields $\tau(t_0) = t_0$, so m_{t_0} stops no matter the continuation by Condition (ii), i.e., $t_0 \in \mathcal{T}$, a contradiction.

We can iterate this argument to show that $t_1 = \tau(t_0) \in \mathcal{T}$ is the second stopping time, $\tau(t_1) \in \mathcal{T}$ is the third, and so on.

Next, we show that if τ is increasing and $t \in [0, \tau(0)]$, then $\mathcal{T} = (t, \tau(t), \tau(\tau(t)), \ldots)$ constitutes an equilibrium. Our construction already shows that m_{t_n} is indifferent about switching to the safe policy at time $t_n = \tau^n(t_0)$. What is left is to show that for $t \notin \mathcal{T}$, m_t weakly prefers to continue experimenting. Fix $t \in (t_n, t_{n+1})$. Since $t > t_n$ and τ is increasing, $\tau(t) \geq \tau(t_n) = t_{n+1}$. Hence the definition of $\tau(t)$ and the fact that $T \mapsto V_T(x)$ is single-peaked by Lemma 3 imply that $V_{t_{n+1}-t}(p_t(m_t)) \geq \frac{s}{\gamma}$, as we wanted. This proves part (iii).

Next, 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 $t \in (t_k, t_{k+1})$ for which $\tau(t) < t_{k+1}$. 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.[26] 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 backward construction in the previous paragraph. If there is no such ϵ, then the previous argument works.[27] 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 $\tau(t) > t$ 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 same backward construction. This finishes the proof of part (ii). Finally, part (iv) is proved with the same logic as the uniqueness in Proposition 1. More generally, if $\tau(t) = \infty$, then $t \notin \mathcal{T}$ for any equilibrium $\mathcal{T}$. ∎

Proof of Proposition 5.

Let $P[\sigma = 1|G] = \overline{\pi}$ and $P[\sigma = 1|B] = \underline{\pi}$. By Equation 1, the indifferent agent after $\sigma = 1$ has prior $x_* = \frac{a}{a+(g-a)\frac{\overline{\pi}}{\underline{\pi}}} < \frac{a}{g}$, while the indifferent agent after $\sigma = 0$ has prior $x^* = \frac{a}{a+(g-a)\frac{1-\overline{\pi}}{1-\underline{\pi}}} > \frac{a}{g}$.

[26]Since τ is continuous, and $\tau(t) < \infty$ for all t, 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, $k > l$.

[27]If there is $\epsilon > 0$ with the required 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 $\bar{t}$.

Given any values of a, λ, h, γ, choose s such that $V\left(\frac{a}{g}\right) < \frac{s}{\gamma} < V\left(\frac{2a}{g+a}\right)$, and take f as follows: $f(x) = 0$ for $x \in [0, x_*]$; $f(x) = \frac{1}{2\epsilon}$ for $x \in (x_*, x_* + \epsilon)$; and $f(x) = \frac{1}{2(1-x_*-\epsilon)}$ for $x \in [x_* + \epsilon, 1]$, for $\epsilon > 0$ small enough. (The essence of the construction is simply that f takes high enough values within $[x_*, x^*)$. Of course, it can be perturbed to make f continuous.) Then, after bad news, the set of potential members during experimentation is contained in $[x^*, 1]$. As f is uniform over this interval, the condition $\frac{s}{\gamma} < V\left(\frac{2a}{g+a}\right)$ guarantees perpetual experimentation by Proposition 2. After good news, the median member is $x_* + \epsilon$, whose posterior is arbitrarily close to $\frac{a}{g}$ for ϵ small enough. Then the condition $V\left(\frac{a}{g}\right) < \frac{s}{\gamma}$ guarantees finite experimentation in equilibrium. Moreover, for ϵ small enough, y_t crosses $x^* + \epsilon$ after an arbitrarily short time, after which no stopping is possible, by the logic of Proposition 4.(iv). So the equilibrium stopping time after $\sigma = 1$ must be arbitrarily close to 0, meaning that the time the risky policy is used for is determined almost entirely by σ, and hence negatively correlated with the state. ■

Proof of Proposition 6. Let $V(y)$, $V_T(y)$, y_t denote the same functions as in the baseline model. As for pivotal agents, note that if k groups have been revealed as winners, there is a mass k of members always in favor of experimentation. Of the remaining $2K+1-k$ groups, only agents with $p_t(x) \geq y_t$ will be members at time t. Then the pivotal agent, $m_{t,k}$, satisfies $(2K+1-k)\left[F(m_{t,k}) - F(y_t)\right] = k + (2K+1-k)\left[1 - F(m_{t,k})\right]$. Clearly $m_{t,0} = m_t$ from the baseline model, and $m_{t,k}$ is strictly increasing in k.

By the same logic as in Proposition 1, perpetual experimentation is an equilibrium if and only if $V(p_t(m_{t,k})) \geq \frac{s}{\gamma}$ for all t, k. Because V and $p_t(\cdot)$ are increasing functions, and $m_{t,k}$ is increasing in k, this holds if and only if $V(p_t(m_{t,0})) \geq \frac{s}{\gamma}$ for all t, which is the same condition from Proposition 1.

As for the uniqueness, if $V(p_t(m_t)) \geq \frac{s}{\gamma}$ for all t with equality for some t, we can obviously construct an equilibrium with stopping in state $(t, 0)$ and nowhere else. The opposite implication is more involved. Suppose that $V(p_t(m_t)) > \frac{s}{\gamma}$ for all t, and there is an equilibrium $\mathcal{T} \neq \emptyset$. As noted in the text, any $(t, k) \in \mathcal{T}$ must have $k \leq K$.

Suppose that there exists t_0 such that $(t_0, K) \in \mathcal{T}$. Note that starting at (t_0, K), if any additional group is revealed as a winner, experimentation is locked in forever after, as there are $K+1$ sure-winner groups. There are two cases: either $\mathcal{T} \cap \{(t, K) : t > t_0\}$ is empty, or not. In the first case, if $m_{t_0,K}$ deviates and continues experimentation, it will never stop. Then her equilibrium action contradicts the condition $V(p_{t_0}(m_{t_0,K})) > V(p_{t_0}(m_{t_0})) > \frac{s}{\gamma}$. In the second case, experimentation next stops, if no more winner groups are revealed, at some time $t_1 \geq t_0$. Then $m_{t_0,K}$'s continuation value from experimentation is a convex combination of $V(p_{t_0}(m_{t_0,K}))$ (which she receives conditional on another group succeeding for the first time before t_1) and $V_{t_1-t_0}(p_{t_0}(m_{t_0,K}))$ (the complementary case). By Lemma 3, and because

$m_{t_0,K} > m_{t_0}$, the condition $V(p_{t_0}(m_{t_0})) > \frac{s}{\gamma}$ implies that $V(p_{t_0}(m_{t_0,K})), V_{t_1-t_0}(p_{t_0}(m_{t_0,K})) > \frac{s}{\gamma}$ for all $t_1 > t_0$. Then $m_{t_0,K}$ strictly prefers to experiment, a contradiction. (If $t_1 = t_0$, Condition (ii) applies.)

Thus there is no t for which $(t, K) \in \mathcal{T}$, i.e., experimentation never stops after K groups are revealed winners. But then the same argument applies to histories of the form $(t, K-1)$, etc. Repeating the argument leads to the conclusion $\mathcal{T} = \emptyset$, a contradiction. ■

Proof of Proposition 7. We first characterize the agents' equilibrium share demands and wealth and consumption paths given an expected path of prices $(\rho_t)_t$ and an expected stopping time $t_0 \in [0, \infty]$.

An agent's per-share gain after the risky policy first succeeds is $h + \overline{\rho} - \rho_t$, if this success occurs at time t. In addition, the instantaneous cost of holding a share through time t, assuming no success, is $\gamma\rho_t - \rho'_t$: ρ'_t is the agent's net capital gain, and $\gamma\rho_t$ the opportunity cost of not lending the funds invested in the share.

Let $Q_t(x) = q_t(x)(h + \overline{\rho} - \rho_t)$ be an agent x's gain from success at time t, and $\xi_t = \frac{\gamma\rho_t - \rho'_t}{h+\overline{\rho}-\rho_t}$ the flow cost of increasing $Q_t(x)$ by 1. Let $V_t(W, x)$ be the continuation utility of an agent x starting at time t, if her wealth at time t is W and there have been no successes, and let $U_t(W, x)$ be the same but assuming a success has occurred. Then the solution to the agent's consumption and investment problem must satisfy the following FOCs:

$$0 = \gamma u'(c_t(x)) - \frac{\partial V_t(W_t(x), x)}{\partial W} \tag{5}$$

$$0 = \gamma u'(c_t(x; \text{succ})) - \frac{\partial U_t(W_t(x) + Q_t(x), x)}{\partial W} \tag{6}$$

$$0 \geq \lambda p_t(x) \frac{\partial U_t(W_t(x) + Q_t(x), x)}{\partial W} - \xi_t \frac{\partial V_t(W_t(x), x)}{\partial W} \quad (= \text{ if } Q_t(x) > 0\,) \tag{7}$$

$$-\frac{\partial u'(c_t(x))}{\partial t} = \lambda p_t(x)(u'(c_t(x; \text{succ})) - u'(c_t(x))) \tag{8}$$

These FOCs, which follow from the Hamilton-Jacobi-Bellman equation for the agent's optimization problem, reflect the following tradeoffs. The agent must be indifferent at the margin between consuming and saving at time t, if there has been no success (Equation 5), and between consuming and saving, immediately after a success that occurred at time t (Equation 6).[28] She must not want to buy any more shares at time t, and must be indifferent at the margin between saving and buying shares at time t if she buys a positive amount (Equation 7). In addition, her (expected) consumption path must satisfy the Euler equation (Equation 8).

[28] Of course the agent must remain indifferent for all $s > t$, but this condition simply leads to the consumption path after a success being constant.

Substituting Equations 5 and 6 into Equation 7, and using that $u'(c) = c^{-\theta}$, we obtain

$$\lambda p_t(x) u'(c_t(x;\text{succ})) \leq \xi_t u'(c_t(x)) \Longleftrightarrow c_t(x;\text{succ}) \geq c_t(x) \left[\frac{\lambda p_t(x)}{\xi_t}\right]^{\frac{1}{\theta}}, \tag{9}$$

again with equality if $Q_t(x) > 0$. Relatedly, $Q_t(x) > 0$ if and only if $\frac{\gamma W_t(x)}{c_t(x)} < \left[\frac{\lambda p_t(x)}{\xi_t}\right]^{\frac{1}{\theta}}$.

We can characterize for the agent's path of choices as follows. Suppose that the agent is holding some shares at time t, so $c_t(x;\text{succ}) = c_t(x) \left[\frac{\lambda p_t(x)}{\xi_t}\right]^{\frac{1}{\theta}}$. Denote $\hat{h} := \frac{\frac{\partial h}{\partial t}}{h}$. Using that $u'(c) = c^{-\theta}$, and hence $\hat{u}'(c_t(x)) = -\theta \hat{c}_t(x)$, and substituting Equation 9 into Equation 8 yields that

$$\begin{aligned}
\theta \hat{c}_t(x) &= -\hat{u}'(c_t(x)) = \lambda p_t(x) \left(\frac{u'(c_t(x;\text{succ}))}{u'(c_t(x))} - 1 \right) = \xi_t - \lambda p_t(x) \\
\Longrightarrow c_t'(x) &= \frac{c_t}{\theta} \left[\xi_t - \lambda p_t(x)\right] = \frac{c_t \xi_t}{\theta} \left(1 - \frac{\lambda p_t(x)}{\xi_t}\right).
\end{aligned} \tag{10}$$

Differentiating Equation 9 with respect to t, substituting in Equation 10 and using the functional form of $p_t(x)$ (in particular, $\frac{\partial p_t(x)}{\partial t} = -\lambda p_t(x)(1 - p_t(x))$) yields

$$\begin{aligned}
\hat{c}_t(x;\text{succ}) &= \hat{c}_t(x) + \frac{1}{\theta}\left[\hat{p}_t(x) - \hat{\xi}_t\right] = \frac{1}{\theta}\left[\xi_t - \lambda p_t(x)\right] + \frac{1}{\theta}\left[-\lambda(1 - p_t(x)) - \hat{\xi}_t\right] \\
&= \frac{1}{\theta}\left[-\lambda + \xi_t - \hat{\xi}_t\right] =: \Gamma_t.
\end{aligned} \tag{11}$$

The rate of change of $c_t(x;\text{succ})$ is thus equal for all agents who are holding shares. An intuition is that, while optimistic agents want to hold more shares over time, they also consume more of their wealth in anticipation of a success, and these two effects cancel out.

The agent must also satisfy the following budget constraints:

$$W_t'(x) = \gamma W_t(x) - c_t(x) - Q_t(x)\xi_t \tag{12}$$

$$c_t(x;\text{succ}) = \gamma W_t(x) + \gamma Q_t(x) \tag{13}$$

Equation 12 is the agent's budget constraint before a success, while Equation 13 reflects that the optimal consumption path after a success is constant. Combining these two equations with Equation 9,

$$
\begin{aligned}
W_t'(x) &= \gamma W_t(x) - c_t(x) - \xi_t \left(\frac{c_t(x)}{\gamma} \left[\frac{\lambda p_t(x)}{\xi_t} \right]^{\frac{1}{\theta}} - W_t(x) \right) \\
\implies (\gamma W_t(x))' &= (\gamma + \xi_t)(\gamma W_t(x) - c_t(x)) + c_t(x)\xi_t \left(1 - \left[\frac{\lambda p_t(x)}{\xi_t} \right]^{\frac{1}{\theta}} \right). \qquad (14)
\end{aligned}
$$

Equations 10 and 14 characterize the evolution of $W_t(x)$ and $c_t(x)$ when share demand is positive. Suppose now instead that $Q_t(x) = 0$. Plugging this into Equation 12, and $c_t(x;\text{succ}) \equiv \gamma W_t(x)$ into Equation 8, we obtain

$$
(\gamma W_t(x))' = \gamma(\gamma W_t(x) - c_t(x)) \qquad (15)
$$

$$
c_t'(x) = \frac{c_t(x)}{\theta} \lambda p_t(x) \left(\left[\frac{c_t(x)}{\gamma W_t(x)} \right]^{\frac{1}{\theta}} - 1 \right). \qquad (16)
$$

We will now show the following:

Claim 2. Set $\theta < 1$. For all t, $\gamma W_t(x) \geq c_t(x)$. If $Q_{t'}(x) > 0$ for some $t' > t$, then $\gamma W_t(x) > c_t(x)$.

Proof. Suppose that $\gamma W_t(x) < c_t(x)$ for some t. If the agent is not holding shares at t, then, from Equations 15 and 16, $c_t'(x) > 0$ and hence $\gamma W_t'(x) - c_t'(x) < \gamma(\gamma W_t(x) - c_t(x))$. If instead $Q_t(x) > 0$, note that $1 - y^{\frac{1}{\theta}} < \frac{1-y}{\theta}$ for any $y \neq 1$ and $\theta < 1$, so

$$
c_t(x)\xi_t \left(1 - \left[\frac{\lambda p_t(x)}{\xi_t} \right]^{\frac{1}{\theta}} \right) \leq \frac{c_t \xi_t}{\theta} \left(1 - \frac{\lambda p_t(x)}{\xi_t} \right) = c_t'(x).
$$

Then, from Equations 10 and 14, $(\gamma W_t(x) - c_t(x))' \leq (\gamma + \xi_t)(\gamma W_t(x) - c_t(x)) < \gamma(\gamma W_t(x) - c_t(x))$. By Grönwall's inequality, $\gamma W_{t'}(x) - c_{t'}(x) \leq (\gamma W_t(x) - c_t(x)) e^{\gamma(t'-t)}$ for all $t' > t$, which goes to $-\infty$. Since $W_t'(x) \leq \gamma W_t(x) - c_t(x)$ by Equation 12, $W_t(x)$ eventually becomes negative, a contradiction.

Next, suppose $\gamma W_t(x) = c_t(x)$ for some t. If the agent is not holding shares at time t, from Equations 15 and 16, $W_t(x)' = c_t(x)' = 0$. If instead $Q_t(x) > 0$, then $\lambda p_t(x) > \xi_t$. Then $\gamma W_t'(x) < c_t'(x)$ by Equation 14, so $\gamma W_t(x) < c_t(x)$ in a right-neighborhood of t, leading to the same contradiction. Hence either $\gamma W_t(x) > c_t(x)$ or $\gamma W_{t'}(x) = c_{t'}(x)$ for all $t' > t$ and the agent never holds shares after t. ■

Note that at the last time $t(x)$ when an agent x ever holds shares, $\lambda p_{t(x)} = \xi_{t(x)}$. For other times $t < t(x)$ when the agent starts or stops holding shares ($Q_t(x) = 0$ but $Q_{t'}(x) > 0$ for t'

arbitrarily close to t), we must have $\left[\frac{\lambda p_t(x)}{\xi_t}\right]^{\frac{1}{\theta}} = \frac{\gamma W_t(x)}{c_t(x)} > 1$. Hence $\frac{\lambda p_t(x)}{\xi_t} > 1$ is a necessary (but not sufficient) condition for x to hold shares. It then follows that $c_t'(x) < 0$ for all $t < t(x)$: if the agent holds shares at t, then this follows from Equation 10 since $\frac{\lambda p_t(x)}{\xi_t} > 1$, and if not, it follows from Equation 16 and Claim 2. Finally, note that Equations 10, 14, 15 and 16 allow us to solve backwards for the agent's choices starting from $t(x)$, given a value of $W_{t(x)}(x)$.[29]

If $\theta = 1$, by analogous arguments, $\gamma W_t(x) \equiv c_t(x)$ and $Q_t(x) > 0$ if and only if $\lambda p_t(x) > \xi_t$.

We now prove part (i). Suppose there is an equilibrium with perpetual experimentation. By Equations 9 and 13, and the fact that $c_t(x) \leq c_0(x) \leq \gamma W_0$, we have that for any x holding shares at time t, $\gamma q_t(x)(h + \overline{\rho} - \rho_t) \leq \gamma W_0 \left[\frac{\lambda p_t(x)}{\xi_t}\right]^{\frac{1}{\theta}}$. Letting $\overline{f} = \max_{x \in [0,1]} f(x)$, and bounding $h + \overline{\rho} - \rho_t \geq h$, it follows that

$$h = \int_0^1 q_t(x) h f(x) dx \leq \frac{\lambda^{\frac{1}{\theta}} W_0}{\xi_t^{\frac{1}{\theta}}} \int_0^1 (p_t(x))^{\frac{1}{\theta}} \overline{f} dx \leq \frac{\lambda^{\frac{1}{\theta}} W_0 \overline{f}}{\xi_t^{\frac{1}{\theta}}} \int_0^1 p_t(x) dx.$$

Note that $p_t(x)^{\frac{1}{\theta}} \leq p_t(x)$ because $p_t(x), \theta \leq 1$. Since $\int_0^1 \frac{x e^{-\lambda t}}{x e^{-\lambda t} + 1 - x} dx = \frac{e^{-\lambda t} \lambda t}{(1 - e^{-\lambda t})^2} - \frac{e^{-\lambda t}}{1 - e^{-\lambda t}} \leq 2\lambda t e^{-\lambda t}$ for t away from 0, there is $M > 0$ such that $\xi_t \leq M e^{-\lambda \theta t} t^{\theta}$ for all t away from 0. In particular $\xi_t \to 0$, so $\rho_t \to 0$.[30] Because ξ_t and $p_t(x)$ go to zero exponentially, Equations 10 and 16 imply that $c_t(x) \searrow c(x)$ for some limit $c(x) > 0$.

We will now show that optimists eventually lose control, i.e., $p_t(m_t) \to 0$. Suppose instead that $p_t(m_t) \geq p > 0$ for arbitrarily high t (say, for a sequence $(t_n)_n$ going to ∞). Note that $m_t \geq \frac{p}{p + (1-p)e^{-\lambda t}}$ and $1 - m_t \leq \frac{(1-p)e^{-\lambda t}}{p + (1-p)e^{-\lambda t}} \leq \frac{(1-p)e^{-\lambda t}}{p}$ for all $t = t_n$.

From Equations 9, 10 and 16, $c_t'(x) \leq \frac{c_t(x)}{\theta}(\xi_t - \lambda p_t(x))$, with equality when $q_t(x) > 0$. Because ξ_t goes to zero exponentially, and $p_t(x)$ goes to 1 exponentially as t decreases, $\int_0^\infty \xi_t$ and $\int_{-\infty}^t (1 - p_z(x)) dz$ are finite. (Moreover, the latter integral is uniformly bounded for all x, t such that $p_t(x) \geq p$.) And of course $c_0(x) \leq \gamma W_0$. Then there is M' such that $c_t(x) \leq M' e^{-\frac{\lambda}{\theta} t}$ for all x and t such that $p_t(x) \geq p$. Then, for all $t = t_n$,

$$\frac{(1-p)e^{-\lambda t}}{p} M' e^{-\frac{\lambda}{\theta} t} \frac{\lambda^{\frac{1}{\theta}} \overline{f}}{\xi_t^{\frac{1}{\theta}}} \geq \int_{m_t}^1 c_t(x) \left[\frac{\lambda p_t(x)}{\xi_t}\right]^{\frac{1}{\theta}} f(x) dx \geq \int_{m_t}^1 c_t(x) \left[\frac{\lambda p_t(x)}{\xi_t}\right]^{\frac{1}{\theta}} \mathbb{1}_{q_t(x)>0} f(x) dx =$$

$$= \int_{m_t}^1 c_t(x; \text{succ}) \mathbb{1}_{q_t(x)>0} f(x) dx \geq \int_{m_t}^1 \gamma q_t(x) h f(x) dx = \frac{\gamma h}{2}.$$

[29] This value can be normalized to 1 and at the end the solution can be scaled to satisfy $W_0(x) = W_0$, since preferences are homothetic.

[30] Otherwise, along a sequence of local maxima of ρ_t converging to $\limsup \rho_t$, or a sequence going monotonically to ρ_t with ρ_t' going to zero, we must have $\limsup_t \xi_t \geq \frac{\gamma \limsup \rho_t}{h + \frac{g}{\gamma}} > 0$.

Then there is $M'' > 0$ such that $\xi_t \leq M'' e^{-\lambda(\theta+1)t}$ for all $t = t_n$. Thus $\frac{\lambda p_t(x)}{\xi_t} \geq \frac{\lambda x e^{\lambda\theta t}}{M''}$ for all x, $t = t_n$, and $\gamma W_t(x) + \gamma Q_t(x) = c_t(x;\text{succ}) \geq c_t(x)\left[\frac{\lambda x}{M''}\right]^{\frac{1}{\theta}} e^{\lambda t}$, whence $W_t(x) \geq \frac{1}{\gamma}\left[\frac{\lambda x}{M''}\right]^{\frac{1}{\theta}} e^{\lambda t} c(x) - Q_t(x)$, for all x, $t = t_n$. Fixing an $\epsilon > 0$, there is $\tilde{M} > 0$ such that $W_t(x) \geq \tilde{M} e^{\lambda t} - Q_t(x)$ for all $x \in [\epsilon, 1-\epsilon]$, $t = t_n$. Assume WLOG that $t_n \geq n$ for all n. Then $W_{t_n}(x) \xrightarrow[n\to\infty]{} \infty$ a.s. in $[\epsilon, 1-\epsilon]$.[31] Since this works for any ϵ, $W_{t_n}(x) \xrightarrow[n\to\infty]{} \infty$ a.s. in $[0,1]$. Finally, Equation 12 then implies that there is x for whom $W_t(x) \geq Ce^{\gamma t}$ for all t, which contradicts the agent's transversality constraint.[32]

Thus, for any p, the fraction of shares held by agents with posterior at least p eventually goes below $\frac{1}{2}$ forever. By the same argument, for any $p, z \in (0,1)$, the fraction held by agents with posterior at least p eventually dips under z forever. Using this result, we will show that there cannot be a majority in favor of experimentation at all t.

If a deviation to the safe policy happens at time t, each x then consumes $\gamma W_t(x) + \gamma q_t(x)\left(\frac{s}{\gamma} - \rho_t\right)$ forever. Under perpetual experimentation, we bound the agent's continuation utility starting at t as follows. The agent would be weakly better off if she kept her equilibrium share demands $(q_{t'}(x))_{t'\geq t}$ but paid zero for them. If so, the expected present value of her consumption stream in the continuation would be $W_t(x) + p_t(x)\int_t^\infty e^{-\gamma(t'-t)} q_{t'}(x)(h + \overline{\rho} - \rho_{t'})\lambda e^{-\lambda(t'-t)} dt'$. Her certainty equivalent is lower, as she is risk-averse. Then, for any agent in favor of experimentation, and for any t large enough that $\rho_t < \frac{s}{2\gamma}$,

$$q_t(x)\frac{s}{2\gamma} \leq p_t(x)\int_t^\infty e^{-(\gamma+\lambda)(t'-t)} q_{t'}(x)\lambda\left(h + \frac{g}{\gamma}\right) dt'.$$

Let $B_t \subseteq [0,1]$ be the set in favor of experimentation at time t. By assumption, $\int_{B_t} q_t(x) f(x) dx \geq \frac{1}{2}$ for all t. Then, for all t,

$$\begin{aligned}\frac{s}{4\gamma} \leq \int_{B_t} q_t(x)\frac{s}{2\gamma} f(x)dx \leq \int_{B_t}\left[p_t(x)\int_t^\infty e^{-(\gamma+\lambda)(t'-t)} q_{t'}(x)\lambda\left(h+\frac{g}{\gamma}\right)dt'\right] f(x)dx \leq \\ \leq \int_t^\infty e^{-\gamma(t'-t)}\left[\int_0^1 p_{t'}(x) q_{t'}(x)\lambda\left(h+\frac{g}{\gamma}\right) f(x)dx\right] dt',\end{aligned}$$

where in the last step we have used that $p_{t'}(x) \geq p_t(x) e^{-\lambda(t'-t)}$ for $t' > t$. Clearly this inequality cannot hold for all t if $\int_0^1 p_{t'}(x) q_{t'}(x) f(x) dx$ goes to 0 as $t' \to \infty$. But of course

[31] Suppose not, i.e., there is $A \subseteq [\epsilon, 1-\epsilon]$ with positive measure and $C > 0$ such that, for every $x \in A$, $W_{t_n}(x) \leq C$ for arbitrarily high n. But then $A \subseteq \cup_{n\geq n_0} A_n = \{x : W_{t_n}(x) \leq C\}$ for all n_0, and $|A_n| \leq \frac{h+\frac{g}{\gamma}}{\tilde{M}e^{\lambda n} - C}$ which goes to zero exponentially, a contradiction.

[32] Recall that $c_t(x) \leq \gamma W_0$ and $\xi_t \leq \hat{M} e^{-\frac{\lambda}{2}\theta t}$ for all t. Take x such that $W_{t_n}(x) \to \infty$ and $Q(x) := \int_0^\infty Q_t(x)\xi_t dt \leq \int_0^1 Q(x) f(x) dx \leq \int_0^\infty (h + \frac{g}{\gamma})\xi_t dt < \infty$; such x must exist if the market-clearing constraint is not violated. We can then show that, for $t \geq t_n$, $W_t(x) \geq (W_{t_n}(x) - Q(x) - \gamma W_0)e^{\gamma(t-t_n)} + \gamma W_0$, so taking n large enough that $W_{t_n}(x) > Q(x) + \gamma W_0$ yields the result.

$\int_0^1 q_{t'}(x)f(x)dx \equiv 1$, and $p_{t'}(x)$ goes to zero pointwise as $t' \to \infty$. For any p, z, take t such that $\int_{p_{t'}(x)\geq p} q_{t'}(x)f(x)dx < z$ for all $t' \geq t$. Then

$$\int_0^1 p_{t'}(x)q_{t'}(x)f(x)dx = \int_0^{p_{t'}(x)\leq p} p_{t'}(x)q_{t'}(x)f(x)dx + \int_{p_{t'}(x)\geq p}^1 p_{t'}(x)q_{t'}(x)f(x)dx < p + z$$

for all $t' \geq t$. Taking p, z low enough yields a contradiction.

For part (ii), we give a formula for $Q_t(x)$. Denote $\int_0^t \xi_s ds = \zeta_t$, and suppose $Q_{\tilde{t}}(x) > 0$ for all $\tilde{t} < t(x)$ and $Q_{\tilde{t}}(x) = 0$ for $\tilde{t} > t(x)$ for some $t(x) > t$. Using Equations 11 and 13,

$$c_t(x;\text{succ}) = e^{\int_0^t \Gamma_{\tilde{t}} d\tilde{t}} c_0(x;\text{succ}) = e^{-\frac{\lambda t}{\theta} + \frac{\zeta_t}{\theta}} \left(\frac{\xi_0}{\xi_t}\right)^{\frac{1}{\theta}} c_0(x;\text{succ}) \tag{17}$$

$$Q_t(x) = e^{\int_0^t \Gamma_{\tilde{t}} d\tilde{t}} \frac{c_0(x;\text{succ})}{\gamma} - W_t(x). \tag{18}$$

Substituting Equations 9, 17 and 18 into Equation 12 yields

$$\begin{aligned} W_t'(x) &= \gamma W_t(x) - e^{\int_0^t \Gamma_{\tilde{t}} d\tilde{t}} c_0(x;\text{succ}) \left[\frac{\xi_t}{\lambda p_t(x)}\right]^{\frac{1}{\theta}} - \xi_t \left(e^{\int_0^t \Gamma_{\tilde{t}} d\tilde{t}} \frac{c_0(x;\text{succ})}{\gamma} - W_t(x)\right) \\ &= (\gamma + \xi_t) W_t(x) - e^{\int_0^t \Gamma_{\tilde{t}} d\tilde{t}} c_0(x;\text{succ}) \left[\left(\frac{\xi_t}{\lambda p_t(x)}\right)^{\frac{1}{\theta}} + \frac{\xi_t}{\gamma}\right]. \end{aligned}$$

Using the method of variation of parameters, for some C_0,

$$W_t(x) = C_0 e^{\gamma t + \zeta_t} - c_0(x;\text{succ}) e^{\gamma t + \zeta_t} \int_0^t e^{-\frac{\lambda z}{\theta} + \frac{\zeta_z}{\theta} - \gamma z - \zeta_z} \left(\frac{\xi_0}{\xi_z}\right)^{\frac{1}{\theta}} \left[\left(\frac{\xi_z}{\lambda p_z(x)}\right)^{\frac{1}{\theta}} + \frac{\xi_z}{\gamma}\right] dz.$$

Plugging in $t = 0$ yields $C_0 = W_0$. Denoting the factor multiplying $c_0(x;\text{succ})$ by Z_t, and $\psi_t = -\frac{\lambda t}{\theta} + \frac{\zeta_t}{\theta} - \gamma t - \zeta_t$,

$$\begin{aligned} W_{t(x)}(x) &= W_0 e^{\gamma t + \zeta_t} - c_0(x;\text{succ}) Z_{t(x)} = \frac{1}{\gamma} c_{t(x)}(x) = \frac{1}{\gamma} c_0(x,\text{succ}) e^{\int_0^{t(x)} \Gamma_t dt} \left[\frac{\xi_{t(x)}}{\lambda p_{t(x)}(x)}\right]^{\frac{1}{\theta}} \\ c_0(x,\text{succ}) &= \frac{W_0 e^{\gamma t(x) + \zeta_{t(x)}}}{Z_{t(x)} + \frac{1}{\gamma} e^{\int_0^{t(x)} \Gamma_t dt} \left[\frac{\xi_{t(x)}}{\lambda p_{t(x)}(x)}\right]^{\frac{1}{\theta}}} \\ &= \frac{W_0}{\int_0^{t(x)} e^{\psi_z} \left(\frac{\xi_0}{\xi_z}\right)^{\frac{1}{\theta}} \left[\left(\frac{\xi_z}{\lambda p_z(x)}\right)^{\frac{1}{\theta}} + \frac{\xi_z}{\gamma}\right] dz + \frac{1}{\gamma} e^{\psi_{t(x)}} \left(\frac{\xi_0}{\lambda p_{t(x)}(x)}\right)^{\frac{1}{\theta}}}. \end{aligned}$$

Substituting this value of $c_0(x;\text{succ})$ into the previous equations,

$$c_t(x;\text{succ}) = \frac{W_0 e^{-\frac{\lambda t}{\theta}+\frac{\zeta_t}{\theta}}\left(\frac{1}{\xi_t}\right)^{\frac{1}{\theta}}}{\int_0^{t(x)} e^{\psi_z}\left(\frac{1}{\xi_z}\right)^{\frac{1}{\theta}}\left[\left(\frac{\xi_z}{\lambda p_z(x)}\right)^{\frac{1}{\theta}}+\frac{\xi_z}{\gamma}\right]dz + \frac{1}{\gamma}e^{\psi_{t(x)}}\left(\frac{1}{\lambda p_{t(x)}(x)}\right)^{\frac{1}{\theta}}}$$

$$c_t(x) = \frac{W_0 e^{-\frac{\lambda t}{\theta}+\frac{\zeta_t}{\theta}}\left(\frac{1}{\lambda p_t(x)}\right)^{\frac{1}{\theta}}}{\int_0^{t(x)} e^{\psi_z}\left(\frac{1}{\xi_z}\right)^{\frac{1}{\theta}}\left[\left(\frac{\xi_z}{\lambda p_z(x)}\right)^{\frac{1}{\theta}}+\frac{\xi_z}{\gamma}\right]dz + \frac{1}{\gamma}e^{\psi_{t(x)}}\left(\frac{1}{\lambda p_{t(x)}(x)}\right)^{\frac{1}{\theta}}}$$

$$W_t(x) = W_0 e^{\gamma t+\zeta_t}\frac{\int_t^{t(x)} e^{\psi_z}\left(\frac{1}{\xi_z}\right)^{\frac{1}{\theta}}\left[\left(\frac{\xi_z}{\lambda p_z(x)}\right)^{\frac{1}{\theta}}+\frac{\xi_z}{\gamma}\right]dz + \frac{1}{\gamma}e^{\psi_{t(x)}}\left(\frac{1}{\lambda p_{t(x)}(x)}\right)^{\frac{1}{\theta}}}{\int_0^{t(x)} e^{\psi_z}\left(\frac{1}{\xi_z}\right)^{\frac{1}{\theta}}\left[\left(\frac{\xi_z}{\lambda p_z(x)}\right)^{\frac{1}{\theta}}+\frac{\xi_z}{\gamma}\right]dz + \frac{1}{\gamma}e^{\psi_{t(x)}}\left(\frac{1}{\lambda p_{t(x)}(x)}\right)^{\frac{1}{\theta}}}$$

$$Q_t(x) = W_0 e^{\gamma t+\zeta_t}\frac{\frac{1}{\gamma}e^{\psi_t}\left(\frac{1}{\xi_t}\right)^{\frac{1}{\theta}} - \int_t^{t(x)} e^{\psi_z}\left(\frac{1}{\xi_z}\right)^{\frac{1}{\theta}}\left[\left(\frac{\xi_z}{\lambda p_z(x)}\right)^{\frac{1}{\theta}}+\frac{\xi_z}{\gamma}\right]dz - \frac{1}{\gamma}e^{\psi_{t(x)}}\left(\frac{1}{\lambda p_{t(x)}(x)}\right)^{\frac{1}{\theta}}}{\int_0^{t(x)} e^{\psi_z}\left(\frac{1}{\xi_z}\right)^{\frac{1}{\theta}}\left[\left(\frac{\xi_z}{\lambda p_z(x)}\right)^{\frac{1}{\theta}}+\frac{\xi_z}{\gamma}\right]dz + \frac{1}{\gamma}e^{\psi_{t(x)}}\left(\frac{1}{\lambda p_{t(x)}(x)}\right)^{\frac{1}{\theta}}}$$

If the agent holds positive shares all the way up to the firm's stopping time t_0, the same equations apply, writing t_0 in place of $t(x)$.[33] From the last equation, part (ii) is immediate if comparing two agents $x < x'$ with the same quitting time ($t(x) = t(x')$), or if both hold shares until t_0. This argument extends to the general case.[34]

For part (iii), set $\theta = 1$. Recall that, in this case, $q_t(x) > 0$ if and only if $\frac{\lambda p_t(x)}{\xi_t} \geq 1$. Our expression for $Q_t(x)$ simplifies to

$$\begin{aligned} Q_t(x) &= W_0 e^{\gamma t+\zeta_t}\frac{\frac{1}{\gamma}e^{-\lambda t-\gamma t}\frac{1}{\xi_t} - \int_t^{t(x)} e^{-\lambda z-\gamma z}\left[\frac{xe^{-\lambda z}+1-x}{\lambda xe^{-\lambda z}}+\frac{1}{\gamma}\right]dz - \frac{1}{\gamma}e^{-\lambda t(x)-\gamma t(x)}\frac{xe^{-\lambda t(x)}+1-x}{\lambda xe^{-\lambda t(x)}}}{\int_0^{t(x)} e^{-\lambda z-\gamma z}\left[\frac{xe^{-\lambda z}+1-x}{\lambda xe^{-\lambda z}}+\frac{1}{\gamma}\right]dz + \frac{1}{\gamma}e^{-\lambda t(x)-\gamma t(x)}\frac{xe^{-\lambda t(x)}+1-x}{\lambda xe^{-\lambda t(x)}}} \\ &= W_0 e^{\zeta_t}\left[x\left(e^{-\lambda t}\frac{\lambda}{\xi_t} - e^{-\lambda t}+1\right)-1\right]. \end{aligned}$$

[33] If the safe policy is adopted at t_0, this affects share prices, as $\rho_t \xrightarrow[t\to t_0]{} \frac{s}{\gamma}$, but it has no impact on ξ_t or any other aspect of the solution: the windfall of switching to the safe policy is baked into share prices. If agents are assumed to initially hold shares, this increases their initial wealth, but there are no other changes.

[34] Briefly, applying Equations 14 and 15, we can show that, if facing two price paths $(\xi_t)_t$, $(\tilde{\xi}_t)_t$ such that $\tilde{\xi}_t < \xi_t$ for t in some set A and $\tilde{\xi}_t = \xi_t$ elsewhere, then $\frac{\tilde{c}_t(x')}{\tilde{W}_t(x')} \leq \frac{c_t(x')}{W_t(x')}$, $\hat{\tilde{c}}_t(x) \leq \hat{c}_t(x)$, and $\tilde{c}_t(x) \leq c_t(x)$ for all $t \leq \inf A$, whence $\tilde{Q}_t(x') \leq Q_t(x')$ for all $t \notin A$, as $Q_t(x') = \max\left\{c_t(x')\left[\frac{1}{\gamma}\left(\frac{\lambda p_t(x')}{\xi_t}\right)^{\frac{1}{\theta}} - \frac{W_t(x')}{c_t(x')}\right], 0\right\}$. Then we can replace the path ξ_t with $\tilde{\xi}_t = \min\left(\xi_t, \lambda p_t(x)\left(\frac{c_t(x)}{\gamma W_t(x)}\right)^{\theta}\right)$. By construction, $\tilde{Q}_t(x) \equiv Q_t(x)$, $\tilde{Q}_t(x') \leq Q_t(x)$ for all $t \notin A$, and our formula applies to $\tilde{Q}_t(x)$, $\tilde{Q}_t(x')$ since both agents weakly want to hold shares at all times.

In general, $Q_t(x) = \max\left\{W_0 e^{\zeta_t}\left[x\left(e^{-\lambda t}\frac{\lambda}{\xi_t} - e^{-\lambda t} + 1\right) - 1\right], 0\right\}$.[35] This is MLRP-increasing in t if and only if $A(t) = \frac{\lambda}{\xi_t}e^{-\lambda t} - e^{-\lambda t} + 1$ is decreasing in t. The market-clearing constraint is

$$W_0 e^{\zeta_t}\int_0^1 \max\{xA(t) - 1, 0\} f(x)dx = h + \overline{\rho} - \rho_t.$$

The log-derivative of e^{ζ_t} with respect to t is $\xi_t = \frac{\gamma\rho_t - \rho_t'}{h+\overline{\rho}-\rho_t}$, while the log-derivative of the right-hand side is $\frac{-\rho_t'}{h+\overline{\rho}-\rho_t}$, a lower value. Hence $A(t)$ is decreasing in t, as we wanted.

■

Proof of Corollary 2. In this case, the instantaneous cost of a share is $\gamma\rho_t - \rho_t' + k_t'$, the gain from a success is $\frac{k_t\gamma}{a}h + \frac{g}{\gamma} - \rho_t - \left(\frac{a}{\gamma} - k_t\right)$, and the windfall from switching to the safe policy is $\frac{s}{\gamma} - \rho_t - \left(\frac{a}{\gamma} - k_t\right)$. Redefine $\xi_t = \frac{\gamma\rho_t - \rho_t' + k_t'}{\frac{k_t\gamma}{a}h + \frac{g-a}{\gamma} - \rho_t + k_t}$, $Q_t(x) = q_t(x)\left(\frac{k_t\gamma}{a}h + \frac{g-a}{\gamma} - \rho_t + k_t\right)$. The same proof of Proposition 7.(i) applies, so long as $\left(\frac{k_t\gamma}{a}h + \frac{g-a}{\gamma} - \rho_t + k_t\right)$ and $\left(\frac{s-a}{\gamma} - \rho_t + k_t\right)$ are bounded away from zero for all t large enough.

For the sake of contradiction, suppose $\liminf_{t\to\infty}\left(\frac{k_t\gamma}{a}h + \frac{g-a}{\gamma} - \rho_t + k_t\right) = 0$.[36] Equivalently, $\limsup \rho_t = \frac{g-a}{\gamma}$. We will argue that then $\limsup \xi_t = \infty$. Indeed, if $\rho_t - k_t$ has local maxima arbitrarily close to $\frac{g-a}{\gamma}$, ξ_t goes to infinity along a sequence of such maxima. If $\rho_t - k_t$ has no local maxima for t greater than some t_0, it must instead converge monotonically to $\frac{g-a}{\gamma}$, and $\rho_t' - k_t'$ must be arbitrarily close to zero for large values of t, with the same result. But then there is t for which $\xi_t > \lambda$, whence $\frac{\lambda p_t(x)}{\xi_t} < 1$ for all agents and, as shown in Proposition 7, no one holds shares, a contradiction.

Because the gain from a success is bounded away from zero, there is $M > 0$ such that $\xi_t \leq Me^{-\lambda\theta t}t^{\theta}$ for all t, as shown in Proposition 7.(i), and as $\xi_t \to 0$, $\rho_t \to 0$. (Note that these partial results did not require the gains from the safe policy to be bounded away from zero.) Then the windfall from switching to the safe policy is also bounded away from zero, and the rest of the proof goes through. ■

[35]This expression is correct even if the agent's share demand switches multiple times between positive and zero in the future: in fact, it does not depend on the agent's future choice set at all.

[36]Naturally this expression can never become negative, or no one would hold shares at that time.

B Additional Extensions (Online Appendix)

B.1 Other Learning Processes

The baseline model 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 and develop 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—a more organic version of Proposition 5. In the case of perfectly informative bad news, in contrast, there is typically under-experimentation.

A Model of Bad News

We consider the same model as in Section 2, except that the risky policy now generates different flow payoffs: if the risky policy is good, it generates a guaranteed flow payoff g. If it is bad, it generates a guaranteed flow payoff g but also experiences failures, which arrive according to a Poisson process with rate λ. Each failure lowers the payoffs of all members by h. Thus, as in the baseline model, the expected flow payoff from the risky policy is g 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. (The switch to the safe technology must happen upon observing a failure but may happen even if no failures are observed.)

As before, m_t is the median member at time t provided that the risky policy has been used up to time t, with no failures. $p_t(m_t)$ 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 $\underline{t}$ be the earliest time when an agent with $V(p_t(m_t)) < \frac{s}{\gamma}$ is pivotal. Proposition 8 provides an equilibrium characterization for this model.

Proposition 8.

(i) If $V(p_t(m_t)) > \frac{s}{\gamma}$ *for all* t*, then there is a unique equilibrium. In it, the organization experiments forever.*

(ii) If $V(p_t(m_t)) < \frac{s}{\gamma}$ *for some* t*, then in any equilibrium the organization stops experimenting at a finite time* $T < \underline{t}$.[37]

Proposition 8 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 halt experimentation, the organization switches to the safe policy even *before* any of these pessimists become pivotal. Note that, 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, consider first 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 the risky policy produces no failures. The more optimistic she becomes, the more she wants to use the risky policy. Hence the agent wants to experiment either forever or not at all.

Two implications follow. First, pessimistic agents with $V(p_t(m_t)) < \frac{s}{\gamma}$ always switch to the safe policy when they are pivotal: they prefer no experimentation to perpetual experimentation, and thus also to any other continuation. Second, optimistic agents with $V(p_t(m_t)) > \frac{s}{\gamma}$ 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 the policy. For them, current and future experimentation are strategic complements.

This reasoning underpins part (ii) of the Proposition. Indeed, agents with $V(p_t(m_t)) > \frac{s}{\gamma}$ 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, even if optimistic, they may prefer to stop experimenting rather than experience "success frustration".[38] 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 is finite. These results stand in stark contrast to those of the baseline model. They depend on a special feature of the perfectly informative bad news

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

[38] This effect is qualitatively similar to Strulovici (2010)'s *winner frustration*, i.e., the phenomenon of a sure winner being unable to capitalize on her learning because too many others oppose the risky policy.

learning process: bad news create common knowledge that the risky policy is bad. There is then no room for organizational capture by optimists who would disagree with the majority.

Proofs

Lemma 7. *In the bad news setting, the value function of an agent with current belief y who is in the organization ($y \geq \frac{a}{g}$) and expects the organization to experiment forever unless a failure is observed is*

$$V(y) = (yg + (1-y)s)\frac{1}{\gamma} - (1-y)s\frac{1}{\gamma+\lambda}.$$

If she expects experimentation to end after a length of time T, her continuation value is

$$V_T(y) = (yg + (1-y)s)\frac{1-e^{-\gamma T}}{\gamma} - (1-y)s\frac{1-e^{-(\gamma+\lambda)T}}{\gamma+\lambda} + e^{-\gamma T}\frac{s}{\gamma}.$$

Proof of Lemma 7. If the risky technology is good, it never experiences a failure, and is never abandoned. The agent then receives an expected flow payoff of g forever. If the technology is bad, it experiences a failure by time t with probability $1 - e^{-\lambda t}$. The agent receives 0 in expectation before the failure, and s after, as the safe policy is adopted. Then

$$V(y) = \int_0^\infty \left(yg + (1-y)\left(1-e^{-\lambda t}\right)s\right)e^{-\gamma t}dt = (yg + (1-y)s)\frac{1}{\gamma} - (1-y)s\frac{1}{\gamma+\lambda}.$$

Similarly, in the case of finite experimentation,

$$\begin{aligned} V_T(y) &= \int_0^T \left(yg + (1-y)(1-e^{-\lambda t})s\right)e^{-\gamma t}dt + \int_T^\infty se^{-\gamma t}dt \\ &= (yg + (1-y)s)\frac{1-e^{-\gamma T}}{\gamma} - (1-y)s\frac{1-e^{-(\gamma+\lambda)T}}{\gamma+\lambda} + e^{-\gamma T}\frac{s}{\gamma}. \end{aligned}$$

■

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

Assumption 1 guarantees that the agents' value functions are well-behaved: that is, for each t', the function $t \mapsto V_{t'-t}(p_t(m_t))$ crosses the threshold $\frac{s}{\gamma}$ finitely many times, and is never tangent to it. Under this assumption, Proposition 9 characterizes the equilibrium of this model.

Proposition 9. *Under Assumption 1, there is a unique equilibrium $\mathcal{T}$. $\mathcal{T}$ is the union of a finite, possibly empty collection of intervals $I_0 = [t_0, t_1]$, $I_1 = [t_2, t_3]$, $\dots$, I_n such that $t_0 < t_1 < t_2 < \dots$. Conditional on the risky policy having reen 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 9. We first argue that there exists T such that for all $t \geq T$, if no failures have been observed during $[0, t]$, then $V(p_t(m_t)) > \frac{s}{\gamma}$ and $p_t(m_t)g > s$. Note that, because in a model of bad news agents never exit, we have $\liminf_{t\to\infty} m_t > 0$. Moreover, $\lim_{t\to\infty} e^{-\lambda t} = 0$. This implies that $\lim_{t\to\infty} p_t(m_t) = \lim_{t\to\infty} \frac{m_t}{m_t + e^{-\lambda t}(1-m_t)} = 1$, so $\lim_{t\to\infty} p_t(m_t)g = g > s$. 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 $V(1) = \frac{g}{\gamma} > \frac{s}{\gamma}$.

Next, we argue that these agents will always experiment.

Claim 3. If $p_t(m_t)g > s$, then in any equilibrium m_t continues experimenting.

Proof of Claim 3. Suppose not. Let $t + t_+$ denote the first time after t when the equilibrium prescribes a switch to the safe policy.[39] Then m_t's payoff if she does not stop experimentation is $V_{t_+}(p_t(m_t))$. From Lemma 7, it follows that if $yg > s$ then $V_T(y) > \frac{s}{\gamma}$ for all $T > 0$. In particular, $V_{t_+}(p_t(m_t)) > \frac{s}{\gamma}$. Then m_t strictly prefers to continue experimenting, a contradiction. ■

We can now already deal with one important case: if $V(p_t(m_t)) > \frac{s}{\gamma}$ 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 3, so $\mathcal{T} \subseteq [0, T)$. Assume $\mathcal{T}$ is nonempty. Let $t^* = \sup \mathcal{T}$. If $t^* \in \mathcal{T}$, then m_{t^*} stops experimenting even though $V(p_t(m_t)) > \frac{s}{\gamma}$ and m_{t^*} gets perpetual experimentation by continuing, a contradiction. If $t^* \notin \mathcal{T}$, a similar argument can be made leveraging Condition (ii).

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

Claim 4. Suppose that in some equilibrium $\mathcal{T}$, m_{t_0} stops experimenting ($t_0 \in \mathcal{T}$). If $p_t(m_t)g < s$ for all $t \in [\underline{t}, t_0)$, then $[\underline{t}, t_0) \subseteq \mathcal{T}$.

Proof of claim 4. Suppose not. 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.[40]

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

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

m_τ's payoff from continuing experimentation is $V_\epsilon(p_\tau(m_\tau))$, which, by Lemma 7, is of the form $\frac{s}{\gamma} + (p_\tau(m_\tau)g - s)\epsilon + \mathcal{O}(\epsilon^2)$. The payoff from stopping is $\frac{s}{\gamma}$. Then, since $p_\tau(m_\tau)g < s$ 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 (ii). ■

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

Let $T_1 = \{t \le t_{2n+1} : p_t(m_t)g \le s\}$ and $T_2 = \{t \le t_{2n+1} : p_t(m_t)g > s\}$. 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 $p_t(m_t)g \le s$ for all $t < t_{2n+1}$. In this case, by claim 4, for all $t \le t_{2n+1}$, 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 $p_t(m_t)g > s$. Set $t_{2n} = \sup\{t < t_{2n+1} : p_t(m_t)g > s\}$. Since the distribution of priors is continuous, $t \mapsto p_t(m_t)$ is continuous, which implies that $p_{t_{2n}}(m_{t_{2n}})g - s = 0$. Then claim 4 implies that for all $t \in [t_{2n}, t_{2n+1}]$, m_t stops experimentation. Note also that $t_{2n} < t_{2n+1}$ as $s = \gamma V(p_{t_{2n+1}}(m_{t_{2n+1}})) > p_{t_{2n+1}}(m_{t_{2n+1}})g$.

Let us conjecture a continuation equilibrium path on which, starting at t, the organization experiments until t_{2n}. We then let $t_{2n-1} = \sup\left\{t < t_{2n} : V_{t_{2n}-t}(p_t(m_t)) \le \frac{s}{\gamma}\right\}$. By construction, for $t \in (t_{2n-1}, t_{2n})$ we have $V_{t_{2n}-t}(p_t(m_t)) > \frac{s}{\gamma}$, so the median m_t continues experimenting for all $t \in (t_{2n-1}, t_{2n})$.

Since the map $t \mapsto V_{t_{2n}-t}(p_t(m_t))$ is continuous (by continuity of the prior distribution plus Lemma 7), we must have $t_{2n-1} = \max\left\{t < t_{2n} : V_{t_{2n}-t}(p_t(m_t)) \le \frac{s}{\gamma}\right\}$. Note that it is then consistent with equilibrium for the median $m_{t_{2n}}$ to stop experimenting.

Now note that if $V_{t_{2n}-t_{2n-1}}(p_{t_{2n-1}}(m_{t_{2n-1}})) = \frac{s}{\gamma}$, then $p_{t_{2n-1}}(m_{t_{2n-1}})g < s$. 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 $\underline{t}_i = \min\{t < t_{2n-1} : p_t(m_t)g \le s\}$).

Set $t_{2n-2} = \underline{t}_i$. Because $p_t(m_t)g \le s$ for all $t \in [t_{2n-2}, t_{2n-1}]$, Claim 3 implies that, for all $t \in [t_{2n-2}, t_{2n-1}]$, m_t stops experimenting.

We then proceed inductively in the same manner, finding the largest t strictly less than t_{2n-2} such that $V_{t_{2n-2}-t}(p_t(m_t)) \le \frac{s}{\gamma}$. 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 $V_{t_{2k+2}}(p_{t_{2k+1}}(m_{t_{2k+1}})) = \frac{s}{\gamma}$ but also $\frac{\partial}{\partial t}V_{t_{2k+2}-t}(p_t(m_t))|_{t=t_{2k+1}} > 0$, that is, $V_{t_{2k+2}-t}(p_t(m_t)) < \frac{s}{\gamma}$ 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 $p_{t_{2k}}(m_{t_{2k}})g - s = 0$ but also $\frac{\partial}{\partial t}p_t(m_t)|_{t=t_{2k}} < 0$, that is, $p_t(m_t)g - s > 0$ 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 8. Part 1 is proved as part of Proposition 9. Part 2 follows from the characterization given in Proposition 9, in particular, from the observation that $t_{2n} < t_{2n+1}$. ■

A Model of Imperfectly Informative (Good) News

The case of imperfectly informative news allows for richer dynamics than the baseline model: agents' beliefs, rather than decreasing monotonically or jumping 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 2 except for the payoffs generated by the risky policy. If the risky policy is good, it generates successes of size h according to a Poisson process with rate λ. If it is bad, successes instead arrive at a rate $\lambda' < \lambda$. We denote $g = \lambda h$ and $b = \lambda' h$, and assume that $g > s > a > b > 0$.

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. The posterior belief of an agent with prior x must then be

$$\frac{x(\lambda t)^k e^{-\lambda t} k!}{x(\lambda t)^k e^{-\lambda t} k! + (1-x)(\lambda' t)^k e^{-\lambda' t} k!} = \frac{x}{x + (1-x)L(k,t)},$$

where

$$L(k,t) = \left(\frac{\lambda'}{\lambda}\right)^k e^{(\lambda - \lambda')t}.$$

We will suppress the dependence of $L(k,t)$ on k and t and use L to denote our sufficient statistic, and focus on Markov equilibria with L as the state variable. Note that high L indicates bad news about the risky policy.

Let $p_L(x)$ be the posterior belief of an agent with prior x in informational state L. Let $V(x,L)$ be the value function of an agent with prior x given that the informational state is L and the organization experiments forever in the continuation. In addition, denote x's ex ante utility under perpetual experimentation, $V(x,1)$, by $V(x)$. (Of course, $V(x,L) = V(p_L(x))$.) The next Proposition shows that, as in Section 3, experimentation can continue forever regardless of how badly the risky policy performs.

Proposition 10. *If $V(m(L), L) > \frac{s}{\gamma}$ for all L, then there is a unique equilibrium. In this* *equilibrium, experimentation never stops, no matter the outcome.*

Moreover, if f is non-decreasing, then $V\left(m(L), L\right) \geq V\left(\frac{2(a-b)}{(g-b)+(a-b)}\right) \geq \frac{1}{\gamma}\frac{(g-a)b+2(a-b)g}{(g-b)+(a-b)}$, *so there exist parameter values such that* $V\left(m(L), L\right) > \frac{s}{\gamma}$ *for all* L.

The first part of Proposition 10 mirrors Proposition 1, based on similar logic. The second part is an abbreviated analog of Proposition 2 (of course, similar bounds can be given for other families of densities, as in Proposition 2.(ii), (iii)). It implies that perpetual experimentation obtains, for instance, if g is high enough and $2a - b > s$. The lower bound we provide for V is based only on the expected flow payoff of the risky policy; it is hard to give a closed form expression for the option value of experimentation in this case.

The following result shows that, under imperfectly informative news, the organization may respond perversely to information, as anticipated in Proposition 5.

Proposition 11. *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.*

The intuition for the result in Proposition 11 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 pivotal agent is most pessimistic when $L = L^*$.[41] Moreover, s must be such that $V_{m(L^*)}(L^*) = \frac{s}{\gamma}$, 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 $L^* < 1$, 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.

Proofs

Proof of Proposition 10. The proof is largely analogous to the proofs of Propositions 1 and 2 for the baseline model. If $V(m(L), L) > \frac{s}{\gamma}$ for all L, perpetual experimentation is clearly an equilibrium, as each pivotal agent $m(L)$ has a choice between $V(m(L), L)$ and $\frac{s}{\gamma}$, and strictly prefers the former. The equilibrium is unique by the following argument.

[41] 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 $m(L)$.

Suppose for the sake of contradiction that there is another equilibrium in which experimentation stops whenever $L \in \mathcal{L} \neq \varnothing$. Let $V_{\mathcal{L}}(x)$ denote the continuation utility of an agent with current belief x in this equilibrium.

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, $V_{\mathcal{L}}(x) \geq \frac{s}{\gamma}$ 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 $L_1 = \inf \mathcal{L}$. In analogous fashion to Lemma 3.(iv), we argue that, if $m(L_1)$ would rather experiment forever than not at all, she would also prefer to experiment until L hits $\mathcal{L}$. That is, if $V(p(L_1, m(L_1))) > \frac{s}{\gamma}$ then $V_{\mathcal{L}}(p(L_1, m(L_1))) \geq \frac{s}{\gamma}$. To see why, suppose that $V_{\mathcal{L}}(p(L_1, m(L_1))) < \frac{s}{\gamma}$. Note that $V(p(L_1, m(L_1)))$ and $V_{\mathcal{L}}(p(L_1, m(L_1)))$ differ only in that, at histories when any element L of $\mathcal{L}$ is first hit, the agent's continuation value is $V(p(L, m(L_1)))$ in the first case and $\frac{s}{\gamma}$ in the second. By construction, $V(p(L, m(L_1))) \leq V(p(L_1, m(L_1)))$ for all $L \in \mathcal{L}$. Letting B be the discounted flow payoffs obtained in these continuations and A be the flow payoffs from all other histories, there is then $\rho \in [0, 1)$ such that $V(p(L_1, m(L_1))) = \rho A + (1-\rho)B$, $V_{\mathcal{L}}(p(L_1, m(L_1))) = \rho A + (1-\rho)\frac{s}{\gamma}$, and $B \leq V(p(L_1, m(L_1)))$.[42] Then $\frac{s}{\gamma} < V(p(L_1, m(L_1))) \leq A$ and hence $V_{\mathcal{L}}(p(L_1, m(L_1))) \geq \frac{s}{\gamma}$. Moreover, there is equality only if $\rho = 0$, in which case $m(L_1)$ would still experiment by Condition (ii). The same argument extends to all L in a neighborhood of L_1, as even a single success pushes L_t strictly left of L_1. But then $\inf \mathcal{L}$ must be strictly greater than L_1, a contradiction.

This proves the first statement. Next, we provide an explicit bound on V when f is non-decreasing. Note that in this model, an agent with belief y expects a flow payoff $yg + (1-y)b$ from the risky policy and a from the outside option. She would then choose to join if and only if $y \geq \frac{a-b}{g-b}$. In particular, the marginal member $y(L)$ in state L satisfies $\frac{y(L)}{y(L)+(1-y(L))L} = p(L, y(L)) = \frac{a-b}{g-b}$. Rearranging yields $y(L) = \frac{a-b}{a-b+(g-a)\frac{1}{L}}$. Then, as f is non-decreasing, $m(L) \geq \frac{1+y(L)}{2} = \frac{2(a-b)L+g-a}{2(a-b)L+2(g-a)}$, and $p(L, m(L)) = \frac{m(L)}{m(L)+(1-m(L))L} = \frac{2(a-b)L+g-a}{2L(a-b)+L(g-a)+g-a} \searrow \frac{2(a-b)}{2(a-b)+(g-a)}$ as L goes to infinity. Finally, it is obvious that $V(y) \geq \frac{yg+(1-y)b}{\gamma}$, as the agent could obtain this payoff by never leaving; substituting $y = \frac{2(a-b)}{2(a-b)+(g-a)}$ yields the inequality in the Proposition. ∎

Lemma 8. *There exists a density f for which some interior state L^* is the unique minimizer of $p(L, m(L))$; $p(L^*, m(L^*)) < \lim_{L\to\infty} p(L, m(L))$; and $L \mapsto p(L, m(L))$ has a kink at L^*, with strictly positive right-derivative.*

Proof of Lemma 8. A construction similar to the one given in Lemma 5 works. Take f as follows: $f(x) = 0$ for $x \in [0, x^*]$; $f(x) = \frac{\rho}{\epsilon}$ for $x \in (x^*, x^* + \epsilon)$; and $f(x) = \frac{1-\rho}{1-x^*-\epsilon}$ for

[42] Technically, this is an application of Lemma 11.

$x \in [x^* + \epsilon, 1]$, for $x^*, \rho \in (0,1)$, and $\epsilon > 0$ small enough. Let L^* be such that $y(L^*) = x^*$ and L^{**} be such that $y(L^{**}) = x^* + \epsilon$. For $L \leq L^*$, $y(L)$ is constant, hence so is $m(L)$, and $L \mapsto p(L, m(L))$ is decreasing. As in Proposition 2, we can show that $L \mapsto p(L, m(L))$ is decreasing in L under a uniform density, which implies that $L \mapsto p(L, m(L))$ is decreasing for $L \geq L^{**}$. For $L \in (L^*, L^{**})$, it can be shown that $L \mapsto p(L, m(L))$ is strictly increasing, with positive right-derivative at L^*, if ϵ is small enough. In addition, taking ϵ small enough and $\rho \geq \frac{1}{2}$ close enough to 1 makes $m(L^*)$ arbitrarily close to $y(L^*)$, so $p(L^*, m(L^*))$ can be made arbitrarily close to $p(L^*, y(L^*)) = \frac{a-b}{g-b}$, which is less than $\frac{2(a-b)}{2(a-b)+(g-a)} = \lim_{L\to\infty} p(L, m(L))$ (as shown in Proposition 10). ■

Proposition 12. *There exist λ, λ', h, s, a, γ, 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 12. For convenience, we multiply all the value functions in this proof by γ. Let $V^\epsilon(x, L)$ denote the value function of agent x given that the state is L and the behavior on the equilibrium path is as described in the Proposition. Note that $V^0(x, L) = V(x, L)$.

Assume that f satisfies the conditions laid out in Lemma 8. Because $y \mapsto V(y)$ is smoothly increasing, it follows that $L \mapsto V^0(m(L), L)$ is uniquely minimized at L^*, with a kink at L^*, positive right-derivative at L^*, and lower value at L^* than in the limit as $L \to \infty$.

Note that both $V^0(m(L), L)$ and the density f constructed in the proof of Lemma 8 are independent of s. Then we can choose s such that $V^0(m(L^*), L^*) = s$.

Because L^* is the unique minimizer of $L \mapsto V^0(m(L), L)$, $V^0(m(L), L) > s$ for all $L \neq L^*$. We aim to show that the conditions $V^\epsilon(m(L^*), L^*) = s$ and $V^\epsilon(m(L), L) \geq s$ for all $L \neq L^*$ still hold for $\epsilon > 0$ small enough.

As in the proof of Proposition 10, we will want to write the difference between $V^\epsilon(m(L), L)$ and $V^0(m(L), L)$ recursively. Note that the paths of play underlying these value functions differ only when the state hits L^* and the ϵ-probability event of $m(L^*)$ stopping is triggered in the former path. Then, for any L, there is some $\rho(L, \epsilon) \in (0,1)$ such that

$$V^0(m(L), L) - V^\epsilon(m(L), L) = \rho(L, \epsilon) \left(V^0(m(L), L^*) - s \right). \tag{19}$$

Here $\rho(L, \epsilon)$ is the expected time (discounted and weighted by probability, based on $m(L)$'s beliefs) that the organization will devote to the safe policy in the conjectured equilibrium with stopping at L^*.[43] Formally, $\rho(L, \epsilon) = \int_0^\infty \gamma e^{-\gamma t} \Pr^\epsilon_{m(L)}(\text{safe policy used at } t) dt$.

[43] Again, this is an application of Lemma 11.

Next, we argue that $\rho(L,\epsilon)$ converges to 0 uniformly as a function of L when $\epsilon \to 0$. That is, denoting $\rho(\epsilon) = \rho(\cdot,\epsilon)$, we want to show $\|\rho(\epsilon)\|_\infty \xrightarrow[\epsilon \to 0]{} 0$.

For the path of play underlying the value function V^0 (that is, when no agent ever stops experimenting), let $N_{t,L}$ be a random variable equal to the number of times $t' \leq t$ for which $L_{t'} = L^*$, according to $m(L)$'s beliefs. Clearly, $\Pr^{\epsilon}_{m(L)}(\text{safe policy used at } t) = E\left(1-(1-\epsilon)^{N_{t,L}}\right) \leq \epsilon E(N_{t,L})$. ($E(N_{t,L})$ is finite, and in fact bounded by a linear function of t, since any two consecutive times t', t'' for which $L_{t'} = L_{t''} = L^*$ must differ by at least $\frac{\ln(\lambda)-\ln(\lambda')}{\lambda-\lambda'}$.) must Denoting $\tilde{\rho}(L) = \int_0^\infty \gamma e^{-\gamma t} E(N_{t,L})dt$, then, $\tilde{\rho}(L)$ is a bounded function, and $\rho(L,\epsilon) \leq \epsilon\tilde{\rho}(L)$, which yields the result.

We are now ready to show that, if $\epsilon > 0$ is small enough, then $V^\epsilon(m(L), L) \geq s$ for all L, with equality at L^*. The case $L = L^*$ is trivial by Equation 19. The case $L < L^*$ is also straightforward: since $V^0(m(L), L^*) < V^0(m(L), L)$, Equation 19 yields

$$\begin{aligned} V^\epsilon(m(L), L) &= V^0(m(L), L) - \rho(L,\epsilon)\left(V^0(m(L), L^*) - s\right) \\ &\geq (1-\rho(L,\epsilon))V^0(m(L), L) + \rho(L,\epsilon)s > s. \end{aligned}$$

Consider now the case $L > L^*$. Because $V^0(m(L^*), L^*) < V^0(m(L), L)$ for all $L > L^*$; $V^0(m(L^*), L^*) < \lim_{L\to\infty} V^0(m(L), L)$; and $\frac{\partial_+ V^0(m(L),L)}{\partial L}|_{L=L^*} > 0$, there are constants $k_1 > 0$, $k_2 > s$ such that $V^0(m(L), L) \geq \min(s + k_1(L-L^*), k_2)$. Indeed, we can take $k_1 = \frac{1}{2}\frac{\partial_+ V^0(m(L),L)}{\partial L}|_{L=L^*}$, a radius $\delta > 0$ such that $V^0(m(L), L) \geq s + k_1(L-L^*)$ for $L \in [L^*, L^*+\delta)$, and $k_2 = \inf_{L \geq L^*+\delta} V^0(m(L), L)$.

Note that $V^0(m(L), L^*) \leq g$. Take ϵ_1 small enough that, if $\epsilon < \epsilon_1$, then $k_1 > \|\rho(\epsilon)\| \sup \frac{\partial V^0(x,L')}{\partial L'}$. Then, by Equation 19, $V^\epsilon(m(L), L) > s$ for all $L \in (L^*, L^*+\delta)$. Take ϵ_2 small enough that, if $\epsilon < \epsilon_2$, then $k_2 - \|\rho(\epsilon)\|_\infty(g-s) > s$. Then, by Equation 19, $V^\epsilon(m(L), L) > s$ for all $L \geq L^* + \delta$. Thus, we can take any $\epsilon \leq \min(\epsilon_1, \epsilon_2)$. ■

Proof of Proposition 11. Take the equilibrium constructed in Proposition 12, and assume that $L_0 > L^*$.[44] Let $P_\vartheta(L_0)$ be the probability that, conditional on starting at L_0 and the policy type being $\vartheta \in \{G, B\}$, the organization stops experimenting at any time $t < \infty$. We will show that $P_G(L_0) > P_B(L_0)$ for L_0 large enough, by proving a stronger result: 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 $Q_\vartheta(L_0, L^*)$ denote the probability that there is a $t < \infty$ such that $L_t \leq L^*$ (i.e., the probability that L_t ever crosses to the left of L^*) when $\vartheta = G, B$. 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 $l(k,t) = \ln L(k,t) = k(\ln(\lambda') - \ln(\lambda)) + (\lambda - \lambda')t$. Let $l_0 = \ln(L_0)$.

[44]Though our definition of L_t requires that $L_0 = 1$, starting the game in state $(1, f)$ ($L_0 = 1$, distribution of beliefs f) is equivalent to starting in state $(L, \tilde{f})$ for a modified density $\tilde{f}$, defined such that, if the distribution of priors is given by $\tilde{f}$, then the distribution of posteriors in state L is given by f.

When $\vartheta = G$, we then have $l_t = l_0 + (\lambda - \lambda')t - [\ln(\lambda) - \ln(\lambda')]\, N(t)$, where $(N(t))_t$ is a Poisson process with rate λ, that is, $N(t) \sim P(\lambda t)$. This can be written as a random walk: for integer values of t, $l_t - l_0 = \sum_{i=0}^{t} S_i$, where $S_i = \lambda - \lambda' - [\ln(\lambda) - \ln(\lambda')]\, N_i$, and $N_i \sim P(\lambda)$ are iid. Note that $E[S_i] = \lambda - \lambda' - \lambda\,(\ln(\lambda) - \ln(\lambda')) < 0$.[45] 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 $\vartheta = B$, we have $(l_t)_t = l_0 + (\lambda - \lambda')t - [\ln(\lambda) - \ln(\lambda')]N(t)$, where $(N(t))_t$ is a Poisson process with rate λ'. This can be written as a random walk with positive drift: $l_t - l_0 = \sum_{i=0}^{t} S_i$, where $S_i = \lambda - \lambda' - [\ln(\lambda) - \ln(\lambda)]\, N_i$, $N_i \sim P(\lambda')$, and $E[S_i] = \lambda - \lambda' - \lambda'\,(\ln(\lambda) - \ln(\lambda')) > 0$. As above, by the strong law of large numbers, we have $l_t \xrightarrow[t\to\infty]{} \infty$ a.s.

Note that $Q_B\left(L, \frac{\lambda'}{\lambda} L\right) = q$ is independent of L because $(l_t)_t$ follows a random walk. Suppose for the sake of contradiction that $\limsup_{L\to\infty} Q_B(L, L^*) > 0$. We claim that this implies $q = 1$. Suppose instead that $q < 1$. Fix $J \in \mathbb{N}$. Then, for L_0 large enough that $\left(\frac{\lambda'}{\lambda}\right)^{2J+1} L_0 > L^*$,

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

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, $P_G(L_0) \geq Q_G(L_0, L^*) \inf_{L\in\left(\left(\frac{\lambda'}{\lambda}\right)L^*, L^*\right]} P_G(L) > 0$. 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 $L_{t'} \in \left(\frac{\lambda'}{\lambda} L^*, L^*\right)$, 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 $\inf_{L\in\left(\left(\frac{\lambda'}{\lambda}\right)L^*, L^*\right]} P_G(L) > 0$ because it is equal to $P_G\left(\left(\frac{\lambda'}{\lambda}\right) L^*\right)$. ■

B.2 General Voting Rules

We assume throughout the paper that the median member of the organization is pivotal. Our analysis, however, extends to other voting rules under which the agent at the z-th

[45] Let $\frac{\lambda}{\lambda'} = 1 + x$. Then $E[S_i] = \lambda'(x - (1 + x)\ln(1 + x))$, where $x - (1 + x)\ln(1 + x)$ is negative for all $x > 0$. Similarly, $\lambda - \lambda' - \lambda'\,(\ln(\lambda) - \ln(\lambda')) = \lambda'(x - \ln(1 + x))$, where $x - \ln(1 + x)$ is positive for all $x > 0$.

percentile is pivotal. (Formally, the pivotal agent z_t at time t is such that $\frac{\int_{z_t}^1 f(x)dx}{\int_{y_t}^1 f(x)dx} = z$.)

Clearly, z_t and $p_t(z_t)$ are increasing in z for all t. To illustrate, assume that f is uniform. Then, as $t \to \infty$, the posterior belief of the pivotal agent converges to $\frac{a}{za+(1-z)g}$ rather than $\frac{2a}{g+a}$. Clearly, $z > \frac{1}{2}$ makes over-experimentation more likely and vice versa.

More generally, everything goes through if the z-th percentile agent is pivotal when the risky policy is in use, and the z'-th percentile agent is pivotal when the safe policy is in use, so long as $z \geq z'$. In particular, this allows us to model supermajority requirements, by taking $z = q$, $z' = 1 - q$ for $q > \frac{1}{2}$. When the risky policy is in use, stringent supermajority requirements are functionally equivalent to more optimistic leadership of the organization, and make it easier to sustain excessive experimentation.

B.3 No Re-entry

Our assumption of free entry and exit has two purposes: it captures the notion of fluid membership, a core premise of our argument, in its most ideal form; and it keeps the model—in particular, membership decisions—simple. But for many organizations it is not descriptively accurate: many firms, and some political parties, would balk at rehiring someone who quit or was fired.

As an alternative, we can consider the case of no re-entry: agents are free to quit, but cannot come back if they do. The model is otherwise identical. The following Proposition summarizes our results for this case.

Proposition 13. *If perpetual experimentation is an equilibrium in the model with no re-entry, it is an equilibrium in the baseline model. Perpetual experimentation is still the unique equilibrium under no re-entry for some parameter values—in particular, if f decreases no faster than a power law,*[46] *and a is close enough to s.*

In other words, in the case of no re-entry, perpetual experimentation obtains only under more stringent parameter conditions than in the baseline model. The logic is as follows: the value functions $V(x)$, $V_T(x)$ take lower values than in the baseline model, because the agent's option value from experimentation is lower when she cannot reenter after a success. In addition, agents exit later than in the baseline model, because remaining a member now confers some option value. As a result, y_t, and hence m_t, are lower than in the baseline model. Both forces make the pivotal agent less willing to experiment.

And yet perpetual experimentation is again the unique equilibrium under familiar conditions: if exit is tempting enough and the number of initial optimists is not too small.

[46] That is, if f MLRP-dominates f_ω for some $\omega > 0$.

The proof also yields explicit parameter conditions analogous to those in Proposition 2. (In this variant, there is also a parameter region with positive measure in which equilibria with perpetual and finite experimentation coexist. The reason is that a finite stopping time encourages pessimists to hold off quitting until the safe policy arrives, thus creating its own base of support.)

Proof of Proposition 13.

The first part is trivial. To prove the second part requires a full solution for this variant of the model.

We first solve for exit decisions. The agent's problem of choosing an exit time, if experimentation is perpetual (or too prolonged for the agent to consider staying until it ends), is a single-agent bandit problem in which exit stops experimentation (as far as she is concerned). The problem is well-known in the literature. Briefly, if the agent's current belief is z and she is indifferent, her marginal payoff from staying in the organization for another instant is

$$zg - a + z\frac{\lambda(g-a)}{\gamma}.$$

$yg - a$ is her net flow payoff from the change, λ her learning rate, and $\frac{(g-a)}{\gamma}$ her net gain from learning that the risky policy is good. Then her belief must be $\frac{a}{g+\frac{\lambda(g-a)}{\gamma}}$. Using Lemma 1, we obtain

$$t(y) = \frac{1}{\lambda}\ln\left(\frac{g-a}{a}\frac{\gamma+\lambda}{\gamma}\frac{y}{1-y}\right).$$

Assume that experimentation stops at time T in equilibrium. An agent y's utility if she stays until T is

$$\begin{aligned}
V_T^{\text{stay}}(y) &= y\int_0^T ge^{-\gamma t}dt + y(1-e^{-\lambda T})\int_T^\infty ge^{-\gamma t}dt + (1-y+ye^{-\lambda T})\int_T^\infty se^{-\gamma t}dt \\
&= \frac{yg}{\gamma} + (1-y)\frac{s}{\gamma}e^{-\gamma T} - \frac{yg-ys}{\gamma}e^{-(\lambda+\gamma)T}.
\end{aligned}$$

If she quits, it is optimal to quit at time $t(y)$. Then she receives

$$\begin{aligned}
V(y) &= y\int_0^{t(y)} ge^{-\gamma t}dt + y(1-e^{-\lambda t(y)})\int_{t(y)}^\infty ge^{-\gamma t}dt + (1-y+ye^{-\lambda t(y)})\int_{t(y)}^\infty ae^{-\gamma t}dt \\
&= \frac{yg}{\gamma} + (1-y)\frac{a}{\gamma}e^{-\gamma t(y)} - \frac{yg-ya}{\gamma}e^{-(\lambda+\gamma)t(y)} \\
&= \frac{yg}{\gamma} + \frac{(1-y)a}{\gamma}\frac{\lambda}{\gamma+\lambda}\left(\frac{a}{g-a}\frac{\gamma}{\gamma+\lambda}\frac{1-y}{y}\right)^{\frac{\gamma}{\lambda}}.
\end{aligned}$$

Note that this is the agent's value function regardless of T, so long as it is optimal to exit before T. In particular, $V(y)$ is the agent's value function under perpetual experimentation.

It is easy to show, as in Lemma 2, that $V_T^{\text{stay}}(y)$ is single-peaked in T, with a peak $T^*(y) < t(y)$. In addition, $V_{t(y)}^{\text{stay}}(y) > V(y)$. Then there is $\hat{T}(y) > t(y)$ such that the agent leaves (at time $t(y)$) if $T \geq \hat{T}(y)$, and stays until T otherwise. In fact, because any agent y's posterior is the same at time $t(y)$, we must have $\hat{T}(y) = t(y) + T_0$ for a fixed $T_0 > 0$. To find T_0, we can consider the case $y = \frac{a}{g+\frac{\lambda(g-a)}{\gamma}}$, for which $t(y) = 0$, and set $V(y) = \frac{a}{\gamma}$, yielding

$$\begin{aligned}
&\frac{1}{\gamma}\frac{ag}{g+\frac{\lambda(g-a)}{\gamma}} + \frac{(g-a)\frac{\gamma+\lambda}{\gamma}}{g+\frac{\lambda(g-a)}{\gamma}}\frac{s}{\gamma}e^{-\gamma T_0} - \frac{a}{g+\frac{\lambda(g-a)}{\gamma}}\frac{g-s}{\gamma}e^{-(\lambda+\gamma)T_0} = \frac{a}{\gamma} \\
&\qquad\qquad \iff (g-a)s\frac{\gamma+\lambda}{\gamma}e^{-\gamma T_0} - (g-s)ae^{-(\lambda+\gamma)T_0} = (g-a)a\frac{\lambda}{\gamma}.
\end{aligned}$$

The equation has a unique solution because the left-hand side is greater than the right for $T_0 = 0$, smaller for large T_0, and decreasing everywhere.

Quitting decisions are then as follows. If perpetual experimentation is expected, each agent x quits at time $t(x)$, and the marginal agent at time t is y_t such that $p_t(y_t) = \frac{a}{g+\frac{\lambda(g-a)}{\gamma}}$. The pivotal agent is m_t, the median of $[y_t, 1]$. If experimentation stops at time T, each agent $x \in [0, y_{T-T_0})$ quits at time $t(x)$, but agents in $[y_{T-T_0}, 1]$ never quit. The marginal agent, $y_{T,t}$, is then equal to $\max(y_t, y_{T-T_0})$. In particular, the marginal agent at time T is y_{T-T_0}, and the pivotal agent is m_{T-T_0}.

By the same argument as in Proposition 1, perpetual experimentation is *an* equilibrium if and only if $V(p_t(m_t)) \geq \frac{s}{\gamma}$ for all t. To guarantee that it is the *unique* equilibrium, we must rule out equilibria with a finite stopping time T. Suppose one exists, and let T' be the next (off-path) stopping time if experimentation does not stop at T. ($T' = T$ if the set of stopping times accumulates at T from above.) Then it must be that $V(p_t(m_t)) \geq \frac{s}{\gamma}$ for all $t \leq T - T_0$, and m_{T-T_0}'s continuation value at T (either $V_{T'-T}^{\text{stay}}(p_T(m_{T-T_0}))$ if $T' - T \leq t(p_T(m_{T-T_0})) + T_0$, or $V(p_T(m_{T-T_0}))$ otherwise) must be weakly less than $\frac{s}{\gamma}$.

In particular, if $V(p_T(m_{T-T_0})) > \frac{s}{\gamma}$, then this cannot be an equilibrium, as m_{T-T_0} would deviate and continue experimenting at T no matter what continuation she expects. (Recall that $V_{T'-T}^{\text{stay}}(p_T(m_{T-T_0}))$ is single-peaked in T', $V_0^{\text{stay}}(p_T(m_{T-T_0})) = \frac{s}{\gamma}$, and $V_{T'-T}^{\text{stay}}(p_T(m_{T-T_0})) = V(p_T(m_{T-T_0}))$ at the switching point, so $V(p_T(m_{T-T_0})) > \frac{s}{\gamma}$ implies $V_{T'-T}^{\text{stay}}(p_T(m_{T-T_0})) > \frac{s}{\gamma}$ for all $T' - T \in (0, t(p_T(m_{T-T_0})) + T_0]$. If $T' = T$, then Condition (ii) applies.)

Thus, $V(p_T(m_{T-T_0})) > \frac{s}{\gamma}$ for all T is a sufficient condition for perpetual experimentation to be the unique equilibrium.

From Equation 1, $y_t = \frac{a}{a+(g-a)\frac{\lambda+\gamma}{\gamma}e^{-\lambda t}}$ for all t. Suppose $f = f_\omega$ as in Proposition 2. Then, as in Claim 1,

$$
\begin{aligned}
m_t = 1-\eta+\eta y_t &= \frac{a+(1-\eta)(g-a)\frac{\lambda+\gamma}{\gamma}e^{-\lambda t}}{a+(g-a)\frac{\lambda+\gamma}{\gamma}e^{-\lambda t}} \\
\implies p_t(m_t) &= \frac{a+(1-\eta)(g-a)\frac{\lambda+\gamma}{\gamma}e^{-\lambda t}}{a+(1-\eta)(g-a)\frac{\lambda+\gamma}{\gamma}e^{-\lambda t}+\eta(g-a)\frac{\lambda+\gamma}{\gamma}} \searrow \frac{a}{a+\eta(g-a)\frac{\lambda+\gamma}{\gamma}} \\
p_{t+T_0}(m_t) &= \frac{a+(1-\eta)(g-a)\frac{\lambda+\gamma}{\gamma}e^{-\lambda t}}{a+(1-\eta)(g-a)\frac{\lambda+\gamma}{\gamma}e^{-\lambda t}+\eta(g-a)\frac{\lambda+\gamma}{\gamma}e^{\lambda T_0}} \searrow \frac{a}{a+\eta(g-a)\frac{\lambda+\gamma}{\gamma}e^{\lambda T_0}}.
\end{aligned}
$$

Thus, for this family of prior densities, $V\left(\frac{a}{a+\eta(g-a)\frac{\lambda+\gamma}{\gamma}}\right) > \frac{s}{\gamma}$ guarantees that perpetual experimentation is an equilibrium; $V\left(\frac{a}{a+\eta(g-a)\frac{\lambda+\gamma}{\gamma}e^{\lambda T_0}}\right) > \frac{s}{\gamma}$ guarantees that it is the only one. The same argument as in Proposition 3 implies that these are also sufficient conditions for any f that MLRP-dominates f_ω. We can substitute back into our expression for V to obtain a transparent condition on the parameters a, s, λ, h, γ.

Finally, we will show that if $f = f_\omega$, and leaving all parameters but a fixed, if $a < s$ is close enough to s, then $V\left(\frac{a}{a+\eta(g-a)\frac{\lambda+\gamma}{\gamma}e^{\lambda T_0}}\right) > \frac{s}{\gamma}$.

It is enough to prove this for the case $a = s$; our claim then follows by continuity of V and T_0. Note that, if $a = s$, then $T_0 = 0$.[47] Then we need to verify that $\gamma V\left(\frac{s}{s+\eta(g-s)\frac{\lambda+\gamma}{\gamma}}\right) > s$. Plugging in this belief and $a = s$ into our expression for V, we obtain

$$
\gamma V\left(\frac{s}{s+\eta(g-s)\frac{\lambda+\gamma}{\gamma}}\right) = \frac{sg}{s+\eta(g-s)\frac{\lambda+\gamma}{\gamma}} + s\frac{\eta(g-s)\frac{\lambda}{\gamma}}{s+\eta(g-s)\frac{\lambda+\gamma}{\gamma}}\eta^{\frac{\gamma}{\lambda}}.
$$

This is greater than s if and only if

$$
\begin{aligned}
\frac{g}{s+\eta(g-s)\frac{\lambda+\gamma}{\gamma}} + \frac{\eta(g-s)\frac{\lambda}{\gamma}}{s+\eta(g-s)\frac{\lambda+\gamma}{\gamma}}\eta^{\frac{\gamma}{\lambda}} &> 1 \\
\iff 1+\frac{\lambda}{\gamma}\eta^{\frac{\gamma+\lambda}{\lambda}} &> \eta\frac{\lambda+\gamma}{\gamma}.
\end{aligned}
$$

This inequality is of the form $1+\frac{x^{a+1}}{a} - x\frac{a+1}{a} > 0$ for some $a > 0$. It holds for all $x \in [0,1)$, as it holds at $x = 0$, becomes an equality at $x = 1$, and its derivative is negative for all $x \in (0,1)$. This finishes the proof as $\eta \in (0,1)$ by definition.

■

[47] Intuitively, in this case an agent gains nothing from waiting beyond $t(y)$, as a switch to the safe policy is no better than quitting.

B.4 Size-Dependent Payoffs

In some settings the payoffs that a policy generates may depend on the organization's 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 $z(\mu)s$, the good risky policy yields instantaneous payoffs of size $z(\mu)h$ generated at rate λ, and the bad risky policy yields zero. We assume that $z(1) = 1$, so that $g = \lambda h$, s 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, $z(\mu)$ is increasing in μ, so there are economies of scale. For the second type, $z(\mu)$ 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 contracts, higher flow payoffs increase the benefits from experimentation, which makes experimentation more attractive.[48] 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 contracts at a lower speed, which begets an ex ante less optimistic pivotal agent. We call this the *control effect*. With economies of scale, both effects are reversed.

With economies of scale, the membership stage of the game may have multiple equilibria, since the more members there are, the higher their payoffs. For simplicity, we assume parameters such that the set of members is uniquely determined.[49] It is sufficient to assume that z does not increase too fast.

The following Proposition presents our first result.

Proposition 14. *Suppose that $f = f_\omega$.*[50] *Let $\overline{z} = \lim_{\mu \to 0} z(\mu)$, and let $V_{z,t}(p_t(m_t))$ denote the utility of the pivotal agent at time t if she expects experimentation to continue forever. If*

$$\eta^{\frac{\gamma}{\lambda}} a \frac{\lambda}{\gamma + \lambda} + \frac{a}{\eta} \frac{\gamma}{\gamma + \lambda} > g \frac{\lambda}{\gamma + \lambda} + a \frac{\gamma}{\gamma + \lambda}$$

then $\lim_{t \to \infty} V_{z,t}(p_t(m_t))$ is strictly increasing in $\overline{z}$ for all $\overline{z} \in [a, \infty)$. In this case, perpetual

[48] 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.

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

[50] Recall that $f_\omega(x)$ is a density with support $[0, 1]$ equal to $(\omega + 1)(1 - x)^\omega$ for $x \in [0, 1]$, and $\eta = 2^{\frac{-1}{\omega + 1}}$.

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 $\lim_{t\to\infty} V_{z,t}(p_t(m_t))$ *is strictly decreasing in* $\overline{z}$ *for all* $\overline{z} \in [a, \infty)$*.*

The intuition for the Proposition is as follows. By the same argument as in the baseline model a sufficient condition for perpetual experimentation is that $V_{z,t}(p_t(m_t)) \geq \frac{s}{\gamma}$ for all t. While it is difficult to calculate $V_{z,t}(p_t(m_t))$ 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 $\overline{z}$ rather than the entire function z. Moreover, it is a hyperbola in $\overline{z}$, so it is either increasing or decreasing in $\overline{z}$ 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 $t \to \infty$, conditional on the risky policy having reen 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.[51]

The inequality in the Proposition determines which case we are in. Because $g > \eta^{\frac{\gamma}{\lambda}} a$ (note that $\eta \in (0,1)$) and $\frac{a}{\eta} > a$, if λ is large enough relative to γ, then over-experimentation is more likely with economies of scale and less likely with a congestion effect, relative to the baseline model. The opposite happens if γ is large relative to λ. The logic 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 $\frac{\lambda}{\gamma}$ is large, so that successes are frequent or the agent is patient, the first success is expected to be one of many, so the cost of a small first success is minor—whereas, if $\frac{\lambda}{\gamma}$ is small, further successes are 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 in which we operationalize size-dependent payoffs 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 $\lambda\mu$. Each success pays a total of h, which is split evenly among members, so that each member gets $\frac{h}{\mu}$. 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 denoting by V_t the

[51] When congestion effects make experimentation less likely in the limit, they may not do so for all t.

continuation utility under perpetual experimentation starting at time t,[52]

$$\gamma \inf V_t(p_t(m_t)) = \gamma \lim_{t\to\infty} V_t\left(\frac{2a}{g+a}\right) = \frac{2ga}{g+a}.$$

In other words, the asymptotic median's expected payoff is simply the flow payoff of the risky policy; the option value of experimentation vanishes as the learning rate goes 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 $\frac{2ga}{g+a} > s$. Note that, in this case, additional members increase the learning rate—a positive externality on other players 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 risky policy is in use—and that the risky policy is used for too long.

Proof of Proposition 14. Fix an equilibrium candidate with perpetual experimentation. Let μ_t be the size of the organization at time t on the equilibrium path. Let $z_t = z(\mu_t)$. The *first success* that happens at time t yields a per-capita payoff $z_t h$, and all further successes pay h (because all agents enter after the first success).

Recall that, if the risky policy is good, a success arrives by time t with probability $1-e^{-\lambda t}$. An agent x with belief x who expects perpetual experimentation has utility

$$\begin{aligned} V_{(z_t)_t}(x) =& x\int_0^{t^*} e^{-\gamma t}\left(\left(1-e^{-\lambda t}\right)g + e^{-\lambda t} z_t g\right)dt + \\ & x\int_{t^*}^{\infty} e^{-\gamma t}\left(\left(1-e^{-\lambda t}\right)g + e^{-\lambda t} a\right)dt + (1-x)\int_{t^*}^{\infty} e^{-\gamma t} a dt, \end{aligned}$$

where t^* is the time at which the agent leaves, that is, when her posterior reaches $\frac{a}{z_t g}$. In particular, if $z_t = z$ for all t, then the above expression equals

$$V_g(x) = x\left(\frac{g}{\gamma} - \frac{g}{\gamma+\lambda} + \frac{zg\left(1-e^{-(\gamma+\lambda)t^*}\right)}{\gamma+\lambda} + \frac{e^{-(\gamma+\lambda)t^*}a}{\gamma+\lambda} - \frac{e^{-\gamma t^*}a}{\gamma}\right) + \frac{e^{-\gamma t^*}a}{\gamma}.$$

Suppose that $f = f_\omega$, as in Proposition 2. By the same arguments as in that Proposition, if y_t satisfies $p_t(y_t) = \frac{a}{zg}$ for all t, then $p_t(m_t) \searrow \frac{a}{\eta(zg-a)+a}$ as $t\to\infty$, and, by Lemma 1, we have $t^* = -\frac{\ln(\eta)}{\lambda}$.[53] Then

$$V_z\left(\frac{a}{\eta(zg-a)+a}\right) = \frac{a}{\eta zg + (1-\eta)a}\left(\frac{g}{\gamma} - \frac{g}{\gamma+\lambda} + \frac{zg\left(1-\eta^{\frac{\gamma+\lambda}{\lambda}}\right)}{\gamma+\lambda} + \frac{\eta^{\frac{\gamma+\lambda}{\lambda}}a}{\gamma+\lambda} - \frac{\eta^{\frac{\gamma}{\lambda}}a}{\gamma}\right) + \frac{\eta^{\frac{\gamma}{\lambda}}a}{\gamma}$$

[52] V_t depends on t because the higher t is, the smaller the organization and the lower the learning rate.
[53] Note that this is the same t^* as in the baseline model.

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

$$\lim_{z\to\infty} \gamma V_z \left(\frac{a}{\eta(zg-a)+a}\right) = \frac{a}{\eta}\left(1-\eta^{\frac{\gamma+\lambda}{\lambda}}\right)\frac{\gamma}{\gamma+\lambda} + \eta^{\frac{\gamma}{\lambda}} a = \frac{a}{\eta}\frac{\gamma}{\gamma+\lambda} + \eta^{\frac{\gamma}{\lambda}} a \frac{\lambda}{\gamma+\lambda}.$$

On the other hand, when the economies of scale are maximal, that is, as $z \to \frac{a}{g}$,[54]

$$\lim_{z\to\frac{a}{g}} \gamma V_z \left(\frac{a}{\eta(zg-a)+a}\right) = \gamma\left(\frac{g}{\gamma} - \frac{g}{\gamma+\lambda} + \frac{a\left(1-\eta^{\frac{\gamma+\lambda}{\lambda}}\right)}{\gamma+\lambda} + \frac{\eta^{\frac{\gamma+\lambda}{\lambda}} a}{\gamma+\lambda}\right) = g\frac{\lambda}{\gamma+\lambda} + a\frac{\gamma}{\gamma+\lambda}$$

Thus $\frac{a}{\eta}\frac{\gamma}{\gamma+\lambda} + \eta^{\frac{\gamma}{\lambda}} a\frac{\lambda}{\gamma+\lambda} > g\frac{\lambda}{\gamma+\lambda} + a\frac{\gamma}{\gamma+\lambda}$ iff $\lim_{z\to\infty} V_z\left(\frac{a}{\eta(zg-a)+a}\right) > \lim_{z\to\frac{a}{g}} V_z\left(\frac{a}{\eta(zg-a)+a}\right)$. Because $V_z\left(\frac{a}{\eta(zg-a)+a}\right)$ is either increasing in z for all $z > 0$ or decreasing in z for all $z > 0$, this condition implies that $V_z\left(\frac{a}{\eta(zg-a)+a}\right)$ is increasing in z for all $z > 0$. Analogously, f the inequality is reversed, $V_z\left(\frac{a}{\eta(zg-a)+a}\right)$ is decreasing in z.

In addition, note that if $V_z\left(\frac{a}{\eta(zg-a)+a}\right)$ is increasing in z, then we can guarantee that, with a congestion effect,

$$V_{(z_\tau)_{\tau\geq t}}\left(p_t(m_t)\right) > V_{z_t}\left(p_t(m_t)\right) > V_{z_t}\left(\frac{a}{\eta(z_t g-a)+a}\right) > V\left(\frac{a}{\eta(g-a)+a}\right)$$

The first inequality follows because $z_\tau \mapsto V_{z_t,\ldots,z_\tau,\ldots}(x)$ is increasing, with a congestion effect $\mu \mapsto z(\mu)$ is decreasing and $t \mapsto \mu_t$ is decreasing, so $t \mapsto z_t = z(\mu_t)$ is increasing. The second inequality follows because $x \mapsto V_{z_t}(x)$ is strictly increasing in x and $p_t(m_t) \searrow \frac{a}{\eta(zg-a)+a}$ as $t \to \infty$ if the function $z(\mu)$ is constantly equal to z . The last inequality follows because $z \mapsto V_z\left(\frac{a}{\eta(zg-a)+a}\right)$ is increasing in z, and with a congestion effect $z_t = z(\mu_t) > z(1) = 1$.

Thus the condition to obtain perpetual experimentation 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 perpetual experimentation is tighter for all t under economies of scale.[55] ■

[54] If $z < \frac{a}{g}$, we enter a degenerate case in which the organization becomes empty immediately.

[55] If $z \mapsto V_z\left(\frac{a}{\eta(zg-a)+a}\right)$ is decreasing, it is harder to give a clean comparison with the baseline model because the tightest case may be away from the limit: for example, with economies of scale, the pivotal agent for large t is more likely to want to experiment as z is low, but a pivotal agent at an early time may be dissuaded by the decreasing path of $(z_t)_t$.

B.5 Continuous (But Non-Tradable) Membership

As a stepping stone between our baseline model and our model with tradable shares from Section 4.2, we can consider a model in which membership is a continuous choice and agents are risk-averse, as in Section 4.2, but membership only confers exposure to current payoffs, and not property rights over future payoffs, as in the baseline model. In this model, perpetual experimentation is possible, and if anything made more likely by selection on the intensive margin. The structure of property rights is thus the key factor limiting experimentation in Section 4.2.

Formally, assume that, at each t, each agent x chooses a participation intensity $q_t \geq 0$. For a worker, q_t represents hours worked; in the example of a dairy cooperative, q_t represents how many cows the agent puts to work within the cooperative. The agent then receives a flow payoff $(s-a)q_t$ if the safe policy is being used, $-aq_t$ if the bad risky policy is being used, and $-aq_t$ plus lump sums of size hq_t at rate λ if the good risky policy is being used. Crucially, if an agent lowers or increases q_t, her exposure to the organization's policy changes but she does not derive any capital gains or losses. And, as in Section 4.2, we will assume that the organization is able to "smooth out" payoffs from the risky policy after the first success, and instead offer a guaranteed flow payoff g per unit of input. As in the baseline model, the organization's size adjusts to accommodate total share demands, with one exception: we will assume that the organization has a maximum capacity K, which becomes binding only after a success or a switch to the safe policy. In these scenarios, in which all agents want infinite participation, the capacity constraint will restrict them to the choice $q_t = K$. We assume K to be high enough that, in all other scenarios, the capacity constraint does not bind.

Agents differ in two dimensions: their prior beliefs x are distributed according to a density f with support $[0, 1]$, and their initial wealth levels W_0 are distributed according to a density w with support $[0, +\infty)$. The two dimensions are independently distributed.[56] As in Section 4.2, they are risk-averse, with utility function $u(c) = \frac{c^{1-\theta}}{1-\theta}$, and can lend and borrow at rate γ. Again, we assume $\theta \in (0, 1]$.

Proposition 15. *Holding all parameters but a fixed, if $a < s$ is close enough to s, there is perpetual experimentation.*

Proof. Much of our analysis from Section 4.2 carries over: we can think of this model as one with tradable shares, but with a constant share price $\rho_t = \frac{a}{\gamma}$ and no market-clearing

[56] The case in which all agents have identical initial wealth is substantively similar but poses some technical problems: it is possible there for all agents to impoverish themselves enough that no one wants to hold any shares, and so the pivotal agent is ill-defined, even though many agents want experimentation to continue.

constraint. The instantaneous cost of increasing payoffs by 1 in the event of a success is $\xi_t = \frac{\gamma \rho_t}{h} \equiv \frac{a}{h}$. The other difference is that, after a success at time t, the agent's wealth effectively goes not from $W_t(W,x)$ to $W_t(W,x) + q_t(W,x)(h + \overline{\rho} - \rho_t)$ but to $W_t(W,x) + q_t(W,x)h + K\frac{(g-a)}{\gamma}$. That is, instead of capital gains, the agent receives a fixed lump sum, and consumes $\gamma W_t(W,x) + q_t(W,x)\gamma h + K(g-a)$ forever after. Similarly, if the safe policy is adopted at time t, the agent consumes $\gamma W_t(W,x) + q_t(W,x)\gamma h + K(s-a)$ forever after.

We first provide a time limit by which an agent stops participating in the organization.

Lemma 9. *Assume an agent (W,x) stops holding shares at time $t(W,x)$. Then $t(W,x) \leq t(x)$ from Lemma 1 for all W. In particular, $p_{t(W,x)}g \geq a$ for all W, x. Moreover, $t(W,x) \xrightarrow[W\to\infty]{} t(x)$.*

Proof. Formally, let $t(W,x) = \sup\{t : q_t(W,x) > 0\}$. We first argue that $t(W,x)$ is finite for all $x < 1$. Suppose instead that $q_t(W,x) > 0$ for arbitrarily high t. For all $t > t(x)$, as $p_t(x) < \frac{a}{g}$, holding shares can only make sense if consumption is higher in the absence of a success. Denoting by $c_t(W,x)$ the consumption at time t of an agent with *initial* (not current) wealth W and prior x, by Equation 9,

$$\frac{c_t(W,x)}{\gamma W_t(W,x) + q_t(W,x)\gamma h + K(g-a)} = \frac{c_t(W,x)}{c_t(W,x;\text{succ})} = \left[\frac{a}{gp_t(x)}\right]^{\frac{1}{\theta}} > 1.$$

Then, for all $t > t(x)$ such that $q_t(W,x) > 0$, Equation 8 implies that $c_t(W,x)$ is increasing in t. Now suppose share demand is positive up to t', and then zero in some interval t'', and then positive again. By continuity, $c_{t'}(W,x) > c_{t'}(W,x;\text{succ})$, so $c_t(W,x)$ is increasing in t at t'. Moreover, because $c_{t'}(W,x) > c_{t'}(W,x;\text{succ}) > \gamma W_{t'}(W,x)$, $W_t(W,x)$ is decreasing in t at t', and hence $c_t(W,x;\text{succ})$ is decreasing in t at t', so the gap $c_t(W,x) - c_t(W,x;\text{succ})$ is increasing. Extending this argument carefully, we can show that $c_t(W,x)$ increases over all of $[t',t'']$. To summarize, there is t_0 such that $c_t(W,x)$ is increasing for all $t \geq t_0$, and $c_{t_0}(W,x) > \gamma W_{t_0}(W,x)$. This violates the agent's budget constraint, a contradiction.

Now knowing that $t(W,x)$ is finite, note that, if $t(W,x) > t(x)$, then positive share demands right before $t(W,x)$ imply $c_{t(W,x)}(W,x) > c_{t(W,x)}(W,x;\text{succ}) > \gamma W_{t(W,x)}(W,x)$. The same argument as above shows that $c_t(W,x)$ increases for all $t > t(W,x)$, violating the budget constraint.

Finally, we can normalize the problem of an agent with high W as that of an agent with $W = 1$ but for whom K is low. For $K = 0$, the agent stops holding shares at $t(x)$, since after she stops holding shares she will not receive any windfalls, and hence her consumption path must be constantly equal to $\gamma W_{t(x)}(W,x)$ whether a success occurs or not, so $\frac{c_t(W,x)}{c_t(W,x;\text{succ})} = 1$ implies $p_t(x) = \frac{a}{g}$. For very small K the result follows by a continuity argument. ■

Hence the distribution of posterior beliefs is always contained in $\left[\frac{a}{g}, 1\right]$, as in the baseline model. Moreover, because $t(W, x)$ is close to $t(x)$ for high W, the organization is never empty. Let $V(y, W)$ be the value function of an agent with current wealth W and current belief y, if she expects perpetual experimentation. Then it is enough to show that there is $\overline{a} < s$ such that, if $a \in [\overline{a}, s]$, then $V(\frac{a}{g}, W)$ is greater than the utility of consuming $\gamma W + K(s-a)$ forever (namely $\frac{1}{\gamma}\frac{(\gamma W + K(s-a))^{1-\theta}}{1-\theta}$), for all W.

We can bound $V(\frac{a}{g}, W)$ below by noting that the agent could (suboptimally) consume γW forever in the absence of a success, and hold no shares, and consume $\gamma W + K(g-a)$ forever after a success. Thus $V(\frac{a}{g}, W) \geq \frac{1}{\gamma}\frac{(\gamma W)^{1-\theta}}{1-\theta} + \frac{a}{g}\frac{\lambda}{\gamma(\gamma+\lambda)}\frac{(\gamma W + K(g-a))^{1-\theta} - (\gamma W)^{1-\theta}}{1-\theta}$. It can be shown that this expression is greater than $\frac{1}{\gamma}\frac{(\gamma W + K(s-a))^{1-\theta}}{1-\theta}$ for all W if it is greater for $W = 0$, i.e., it is sufficient to choose a close enough to s that $\frac{a}{g}\frac{\lambda}{\gamma+\lambda}(K(g-a))^{1-\theta} > (K(s-a))^{1-\theta}$. The result follows. ■

B.6 A Model with Heterogeneous Preferences

Although we assume throughout the paper that the agents' differing incentives to both experiment and remain in the organization are the result of heterogeneous priors, similar results arise in a model with common priors but (ex ante) heterogeneous payoffs from experimentation. Formally, assume that all agents have a prior p_0 that the risky policy is good, and they are distributed according to a density f with support $[0, +\infty)$, where an agent h receives a lump sum of size h each time the risky policy succeeds. As in the baseline model, the safe policy yields s for everyone, and the outside option a for everyone.

An agent h now wants to be a member at time t if and only if $p_t \lambda h = \frac{p_0 e^{-\lambda t}}{p_0 e^{-\lambda t} + 1 - p_0}\lambda h \geq a$, so the marginal agent at time t is now $h_t = \frac{a}{\lambda}\left(1 + \frac{1-p_0}{p_0}e^{\lambda t}\right)$. The pivotal agent m_t satisfies $\int_{h_t}^{m_t} f(h)dh = \int_{m_t}^{\infty} f(h)dh$. We can show that Proposition 1 holds, replacing $V(p_t(m_t))$ with $V(p_t, m_t)$, i.e., the value function of an agent with belief p_t and prize size m_t under perpetual experimentation. Moreover, $V(p_t, m_t)$ is greater than $\frac{s}{\gamma}$ if a is close enough to s and f does not decrease too steeply. For instance, if $f(x) \propto \frac{1}{x^2}$, then $m_t \equiv 2h_t$, and $V(p_t, m_t) \geq \frac{1}{\gamma}\lambda p_t m_t = \frac{1}{\gamma}2\lambda p_t h_t = 2\frac{a}{\gamma}$, so it is enough to take $a > \frac{s}{2}$.

B.7 A Model with Unrestricted Policy Changes

In this Section we present a version of the model in which membership and policy strategies are the primitives, and switching to the safe policy is reversible. We show that the equilibrium membership strategies are as postulated in Section 2, and switches to the safe policy are in fact permanent in every equilibrium. Hence our simplifying assumptions in the main text are without loss of generality.

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, is chosen by the decision-maker who is pivotal given the incumbent policy π_{t^-}. Similarly, π_{t^+} is chosen by 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 $L = e^{\lambda t}$ if there have been no successes and $L = 0$ otherwise. As in our model with imperfectly informative news, L summarizes the informational state of the game. We define a membership function β so that $\beta(x, L, \pi) = 1$ if agent x chooses to be a member given information L and policy π, and $\beta(x, L, \pi) = 0$ otherwise. We define a policy correspondence α so that $\alpha(L, \pi)$ is the set of policies that the median voter, $m(L, \pi)$, is willing to choose.[57] We emphasize that $\alpha(L, \pi)$ need *not* be the set of policies that the median voter finds optimal in the sense of maximizing her utility given the behavior of the other agents—that is, $\alpha(L, \pi)$ is not an equilibrium notion. Our notion of strategy profile summarizes the above requirements:

Definition 2. A Markov strategy profile is given by a membership function $\beta : [0, 1] \times \mathbb{R}_{\geq 0} \times \{0, 1\} \to \{0, 1\}$, a policy correspondence $\alpha : \mathbb{R}_{\geq 0} \times \{0, 1\} \to \{\{0\}, \{1\}, \{0, 1\}\}$, 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 λ or 0 if $\vartheta = G$ or B respectively; letting $\left(\tilde{L}_\tau\right)_\tau$ be given by $\tilde{L}_\tau = e^{\lambda\tau}$ if $\tilde{k}_\tau = 0$ and $\tilde{L}_\tau = 0$ otherwise; and letting $n(t) = \int_0^t \pi_{t'} dt'$ 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 $t \geq 0$.

[57] $\alpha(L, \pi)$ 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. α should be understood as a choice rule in the decision-theoretic sense.

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 $(L_{t_0}, \pi_{t_0}) = (L, \pi)$.

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

(i) $\beta(x, L, \pi) = 1$ if $s + \pi(p(L, x)g - s) > a$ and $\beta(x, L, \pi) = 0$ otherwise.

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

Part (i) of the definition says that agents make membership decisions that maximize their flow payoffs. Part (ii) says that the pivotal agent chooses her preferred policy based on her expected utility, assuming that the equilibrium strategies are played in the continuation.

As in Section 2, an additional condition is needed to rule out undesirable equilibria that arise when $V(m(L, \pi)), L, 1) = V(m(L, \pi), L, 0)$ for the trivial reason that the continuation is independent of $m(L, \pi)$'s actions. To eliminate such equilibria, we will consider 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 at time $t_0 + \epsilon$. We will impose the requirement that equilibria satisfy the following:

(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 $\alpha(L, \pi) = \{\pi'\}$.

Analysis

The results are structured as follows. Lemmas 10, 11 and 13 are technical statements. Lemma 12 shows that agents strictly prefer the risky policy after a success. Proposition 16 shows that switches to the safe policy are permanent.

Lemma 10. *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.*[58]

Proof of Lemma 10.

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—otherwise $\mathcal{T}$ would

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

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 $\hat{\pi}_{t+} = \pi_{t+} = \hat{\pi}_t$, so $(\hat{\pi}_t)_t$ is càdlàg. ■

Corollary 3. *For any strategy profile $(\beta, \alpha, (L_t, \pi_t)_t | (L, \pi, \vartheta))$ the stochastic process $(L_t, \hat{\pi}_t)_t | (L, \pi, \vartheta)$ (where $(\hat{\pi}_t)_t$ is as in Lemma 10) has càdlàg paths, satisfies Conditions (a) and (b), and induces a path of play that yields the same payoffs as the strategy.*

Lemma 11 (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 Ψ be a closed subset of Θ and define a stochastic process $(\psi_t)_t$ with a co-domain $(\Psi \cup \{\varnothing\})$ as follows: $\psi_t = \theta \in \Psi$ if there exists $t' \leq t$ such that $\theta_{t'} = \theta$ and $\theta_{t''} \notin \Psi$ for all $t'' < t'$. If this is not true for any $\theta \in \Psi$, then $\psi_t = \varnothing$.[59] *Then*

$$U(\theta_0) = \int_0^\infty e^{-\gamma t} E_{\theta_0} \left[f(\theta_t) \mathbb{1}_{\{\psi_t = \varnothing\}} \right] dt + \int_\Psi U(\theta) d\tilde{P}$$

where $\tilde{P}$ is defined as follows: P_{ψ_t} is the probability measure on $\Psi \cup \varnothing$ induced by ψ_t, and $\tilde{P} = \gamma \int_0^\infty e^{-\gamma t} P_{\psi_t} dt$.

Proof of Lemma 11.

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

So it remains to show that

$$\int_\Psi U(\theta) d\tilde{P} = \int_0^\infty e^{-\gamma t} E_{\theta_0} \left[f(\theta_t) \mathbb{1}_{\{\psi_t \in \Psi\}} \right] dt.$$

Define a random variable ξ with co-domain $(\Psi \times [0, \infty)) \cup \{\varnothing\}$ as follows: $\xi = (\theta, t)$ if $\theta_t = \theta \in \Psi$ and $\theta_{t'} \notin \Psi$ for all $t' < t$. If this is not true for any $\theta \in \Psi$ and $t \geq 0$, then $\xi = \varnothing$.[60] Let P_ξ be the probability measure on $(\Psi \times [0, \infty)) \cup \{\varnothing\}$ induced by ξ. Let $\theta(\xi)$

[59] In other words, ψ_t takes the value of the first $\theta \in \Psi$ that $(\theta_t)_t$ hits.
[60] In other words, ξ takes the value of the first $\theta \in \Psi$ that $(\theta_t)_t$ hits, and the time when it hits.

and $t(\xi)$ be the random variables equal to the first and second coordinates of ξ, conditional on $\xi \neq \varnothing$. Note that $\psi_t = \theta$ if and only if $\xi = (\theta, t')$ for some $t' \leq t$. Then we can write

$$\int_0^\infty e^{-\gamma t} E_{\theta_0}\left[f(\theta_t)\mathbb{1}_{\{\psi_t \in \Psi\}}\right] dt = \int_0^\infty e^{-\gamma t} E_{\theta_0}\left[f(\theta_t)\mathbb{1}_{\{\xi \in \Psi \times [0,\infty)\}}\mathbb{1}_{\{t \geq t(\xi)\}}\right] dt =$$
$$= \int_{\Psi \times [0,\infty)} \left(\int_{t(\xi)}^\infty e^{-\gamma t} E_{\theta_0}[f(\theta_t)|\xi] dt\right) dP_\xi = \int_{\Psi \times [0,\infty)} e^{-\gamma t(\xi)} U(\theta(\xi)) dP_\xi =$$
$$= \int_{\Psi \times [0,\infty)} \left(\int_{t(\xi)}^\infty \gamma e^{-\gamma t} dt\right) U(\theta(\xi)) dP_\xi = \int_0^\infty \int_{\Psi \times [0,\infty)} \gamma e^{-\gamma t} \mathbb{1}_{\{t \geq t(\xi)\}} U(\theta(\xi)) dP_\xi dt =$$
$$= \int_0^\infty \gamma e^{-\gamma t} \left(\int_{\Psi \times [0,\infty)} \mathbb{1}_{\{t \geq t(\xi)\}} U(\theta(\xi)) dP_\xi\right) dt = \int_0^\infty \gamma e^{-\gamma t} \left(\int_\Psi U(\theta) dP_{\psi_t}\right) dt = \int_\Psi U(\theta) d\tilde{P}$$

as desired. ■

Lemma 12. *In any equilibrium, $\alpha(0,1) = \alpha(0,0) = 1$: after a success, the risky policy is used.*

Proof of Lemma 12.

$L_{t_0} = 0$ 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 11, there is $\rho \in [0,1]$ such that

$$V(0,0) = \rho \frac{s}{\gamma} + (1-\rho) V(0,1). \tag{20}$$

It follows that there exist $\eta \in [0,1]$ and $\eta' \in [0,1]$ such that $\eta \geq \eta'$[61] and

$$V(0,0) = \eta \frac{s}{\gamma} + (1-\eta)\frac{g}{\gamma}, \qquad V(0,1) = \eta' \frac{s}{\gamma} + (1-\eta')\frac{g}{\gamma}.$$

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.

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. ■

[61] $\eta \geq \eta'$ for the following reason. $V(0,1) = \eta' \frac{s}{\gamma} + (1-\eta')\frac{g}{\gamma}$ and Equation (20) imply that $\eta \frac{s}{\gamma} + (1-\eta)\frac{g}{\gamma} = V(0,0) = \rho\frac{s}{\gamma} + (1-\rho)V(0,1) = (\rho + (1-\rho)\eta')\frac{s}{\gamma} + (1-\rho)(1-\eta')\frac{g}{\gamma}$. Then $\eta = \rho + (1-\rho)\eta'$, which implies that $\eta \geq \eta'$, as required.

Lemma 13. *For any state (L,π), there is a CDF G with support*[62] *contained in $[0,\infty]$ such that*

$$V(x,L,\pi)=\int_0^\infty V_T(p(L,x))dG(T)$$

for all $x\in[0,1]$, where $V_T(y)$ is as defined in the text. Similarly, for any state (L,π) and any $\epsilon>0$, there is a distribution G_ϵ with support contained in $[0,\infty]$ such that

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

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

Proof of Lemma 13.

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

Without loss of generality, we can assume that the distribution over future states 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 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 12, 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. ■

Proposition 16. *In any equilibrium, for any L, if $0\in\alpha(L,1)$, then $\alpha(L,0)=0$: switches to the safe policy are permanent.*

Proof of Proposition 16.

If $L=0$, then $\alpha(0,\pi)=1$ for all π by Lemma 12, 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

$$V(x,L,0)=\rho_L\frac{s}{\gamma}+(1-\rho_L)V(x,L,1) \tag{21}$$

for all x. This follows from Lemma 11, with the added observation that ρ_L (equivalently, $\tilde{P}$

[62] 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 of perpetual experimentation.

in the notation of Lemma 11) is independent of x in this case because the stochastic process governing (L, π) is independent of x if $\pi = 0$.[63] We now consider three cases.

Case 1: Suppose that $\rho_L > 0$, and that the expected amount of experimentation after switching to state $(L, 1)$ is positive. Clearly $V(x, L, 1)$ is strictly increasing in x. Then Equation (21) implies that $V(x, L, 1) - V(x, L, 0)$ is strictly increasing in x. Since $m(L, 1) > m(L, 0)$, we have $V(m(L, 1), L, 1) - V(m(L, 1), L, 0) > V(m(L, 0), L, 1) - V(m(L, 0), L, 0)$. 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 after switching to 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

$$\overline{V}(x, L, 0, \epsilon) = \rho_\epsilon \frac{s}{\gamma} + (1 - \rho_\epsilon) V(x, L, 1) \tag{22}$$

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

In the following argument, for convenience, we subtract $\frac{s}{\gamma}$ from every value function.[64] Let us calculate $\overline{V}(x, L, 1, \epsilon)$ and $V(x, L, 1)$. Let $G_\epsilon(T)$ and $G(T)$ be the corresponding CDFs from Lemma 13. By definition, for $T \in [0, \epsilon]$, $1 - G_\epsilon(T) = 1$ and for $T > \epsilon$, $1 - G_\epsilon(T) =$

[63] If (L_t, π_t) has càdlàg paths, this follows from Lemma 11. If not, then Lemma 11 cannot be applied because the stochastic process in question is not necessarily right-continuous. However, we can use Corollary 3 of Lemma 10 to obtain a payoff-equivalent path of play with càdlàg paths and then apply Lemma 11 to it.

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

$\frac{1-G(T)}{1-G(\epsilon)}$.[65] Hence for $\epsilon > 0$ sufficiently small we have

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

The third equality follows from the fact that $G_\epsilon(T) = 0$ for all $T \in [0,\epsilon]$ and $G_\epsilon(T) = \frac{G(T)-G(\epsilon)}{1-G(\epsilon)}$ for $T > \epsilon$. For the last equality, we use the fact that $\lim_{\epsilon \to 0} G(\epsilon) = 0$ since $\inf\{t' > t : \pi_{t'} = 0\} > t$ a.s. and the fact that $\left.\frac{\partial_+ V_T(x)}{\partial T}\right|_{T=0} = \max\{xg, a\} - s + x\frac{\lambda(g-s)}{\gamma}$ by part (ii) of Lemma 3.[66]

Our assumption that $1 \in \alpha(L, 0)$ implies that $V(m(L,0), L, 1) \geq 0$: else we would obtain a contradiction of Condition (iii), as Equation 22 and our calculation of $\overline{V}(x, L, 1, \epsilon)$ would imply that $\overline{V}(m(L,0), L, 1) \leq V(m(L,0), L, 1) + G(\epsilon)\mathcal{O}(\epsilon) < e^{-\gamma\epsilon} V(m(L,0), L, 1) = \overline{V}(x, L, 0, \epsilon)$ for all ϵ small enough.

It follows that, because $m(L,1) > m(L,0)$ and $x \mapsto V(x, L, 1)$ is strictly increasing, $V(m(L,1), L, 1) > 0$. Then $\overline{V}(x, L, 1, \epsilon) = \frac{V(x,L,1)}{1-G(\epsilon)} + G(\epsilon)\mathcal{O}(\epsilon)$ implies that $\overline{V}(m(L,1), L, 1, \epsilon) \geq V(m(L,1), L, 1)$. Moreover, because $V(m(L,1), L, 1) > 0$, we have $V(m(L,1), L, 1) > e^{-\gamma\epsilon} V(m(L,1), L, 1) = \overline{V}(m(L,1), L, 0, \epsilon)$.[67] Then $\overline{V}(m(L,1), L, 1, \epsilon) > \overline{V}(m(L,1), L, 0, \epsilon)$ 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 $V(x, L, 0) = V(x, L, 1) = \frac{s}{\gamma}$ for all x, and $\overline{V}(x, L, 0, \epsilon) = \frac{s}{\gamma}$ for all x and $\epsilon > 0$. Again, we subtract $\frac{s}{\gamma}$ 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 13 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

[65] This is immediate if stopping is always permanent on the equilibrium path, i.e., if $G(T)$ actually represents the probability of stopping experimentation permanently at time T. It is also true in general, though for a less obvious reason: the operation defined in Lemma 13 mapping a path of play to a cdf G, and the operation modifying a path of play to never stop for $t \in [0, \epsilon)$, commute.

[66] In greater detail, $\int_0^\epsilon V_T(p(L,x))dG(T) \approx \int_0^\epsilon (k_0 T + k_1)dG(T) \leq \int_0^\epsilon (k_0\epsilon + k_1)dG(T) = (k_0\epsilon + k_1)\int_0^\epsilon dG(T) = (k_0\epsilon + k_1)(G(\epsilon) - G(0)) = (k_0\epsilon + k_1)G(\epsilon) = k_0\epsilon G(\epsilon) = G(\epsilon)\mathcal{O}(\epsilon)$ where we have used the fact that we subtracted $\frac{s}{\gamma}$ from every value function to get rid of the constant k_1.

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

G (that is, $G_\epsilon(T) \xrightarrow[\epsilon\to 0]{} G(T)$ for all $T \geq 0$). Then

$$\overline{V}(x, L, 1, \epsilon) \xrightarrow[\epsilon\to 0]{} \int_0^\infty V_T(p(L, x))dG(T)$$

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,[68] whence $\lim_{\epsilon\to 0} \overline{V}(m(L, 0), L, 1, \epsilon) \geq 0$.

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 as $E_G[T] > 0$. But then $\overline{V}(m(L, 1), L, 1, \epsilon) > 0$ for all $\epsilon > 0$ sufficiently small, which implies that $\alpha(L, 1) = 1$ by Condition (iii), a contradiction.

Second, if $E_G[T] = 0$, then we can employ a similar argument using the fact that, by Lemma 3, $\left.\frac{\partial V_\epsilon(p(L,x))}{\partial \epsilon}\right|_{\epsilon=0}$ is strictly increasing in x, and we have

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

■

The remaining results of the paper can then be proved in this model.

[68]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 $\frac{s}{\gamma}$ 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).