

# OUR DISTRUST IS VERY EXPENSIVE

RAHUL DEB<sup>§</sup>, MATTHEW MITCHELL<sup>†</sup>, AND MALLESH M. PAI<sup>‡</sup>

MARCH 12, 2019

**ABSTRACT:** Motivated by reputation management in a variety of different markets for “expertise” (such as online content providers and experts in organizations), we develop a novel repeated-game framework in which a principal screens a strategic agent whose type determines the rate at which he privately receives payoff relevant information. The stage game is a bandit setting, where the principal chooses whether or not to experiment with a risky arm which is controlled by an agent who privately knows its type. Irrespective of type, the agent strategically chooses output from the arm to maximize the duration of experimentation. Experimentation is only potentially valuable to the principal if the arm is of the high type. Our main insight is that reputational incentives can be exceedingly strong: the agent makes inefficient output choices in all equilibria (subject to a mild refinement) and that this can result in market breakdown even when the uncertainty about the agent’s type is arbitrarily small. We show that (one-sided) transfers do not prevent this inefficiency and we suggest alternate ways to improve the functioning of these markets.

**KEYWORDS:** reputation, repeated games with imperfect public monitoring, relational contracting, strategic experimentation, markets for expertise, media.

**JEL CLASSIFICATION:** D82, D83, D86.

## 1. INTRODUCTION

Attention is money for much of the advertising driven Internet. Typically, consumers do not pay for content and instead revenue is generated by their continued attention in the form of clicks. A consequence is that content providers need to sustain continued interest as consumers can freely withdraw their attention at any time. This creates a dilemma: since genuine content (which varies in quality) can only be generated periodically, how do content providers balance the quality and the frequency of the new content they provide with the aim of retaining both interest and trust?

---

<sup>§</sup>DEPARTMENT OF ECONOMICS, UNIVERSITY OF TORONTO, [RAHUL.DEB@UTORONTO.CA](mailto:RAHUL.DEB@UTORONTO.CA)

<sup>†</sup>GRADUATE DEPARTMENT OF MANAGEMENT, UNIVERSITY OF TORONTO, [MATTHEW.MITCHELL@UTORONTO.CA](mailto:MATTHEW.MITCHELL@UTORONTO.CA)

<sup>‡</sup>DEPARTMENT OF ECONOMICS, RICE UNIVERSITY, [MALLESH.PAI@RICE.EDU](mailto:MALLESH.PAI@RICE.EDU)

The title is a quote from the essay “Man the Reformer” by Ralph Waldo Emerson. We would like to thank Heski Bar-Isaac, Dirk Bergemann, Jernej Copic, Martin Cripps, Joyee Deb, Mehmet Ekmekci, Jeff Ely, Drew Fudenberg, Marina Halac, Johannes Hörner, Matt Jackson, Thomas Mariotti, Maher Said, Colin Stewart, Juuso Välimäki, Alex Wolitzky and audiences at numerous seminars for their time and insightful comments. Deb would like to thank Boston University, the Einaudi Institute and MIT for their hospitality and the SSHRC for their continued and generous financial support. Pai gratefully acknowledges financial support from NSF grant CCF-1763349.

Specifically, how do they manage reputation in an environment where “fake” content can be generated at will? And how does this revenue model (as opposed to traditional payment by subscription) affect market functioning?

Although it is a very different context, similar incentives are also faced by “experts” employed in organizations. Specifically, consider a scientist working at a pharmaceutical company. The scientist’s ability determines the frequency at which she receives research ideas that may ultimately lead to new products after successful drug trials. Even a talented scientist under pressure to keep his job could choose to recommend a drug trial for a new compound which is very unlikely to succeed but may help create the impression that he is generating new ideas. How does the company’s decision of when to fire the scientist and their bonus structure (for successful trials) balance their dual goals of giving a talented scientist a sufficient amount of time to succeed (and reveal himself to be of high ability) but also prevent costly failed trials?

What is common to these (and numerous other) seemingly disparate examples is that they are environments in which a principal wants to dynamically screen an agent whose type (good or bad) determines the *rate* at which she receives information that is payoff relevant for the principal and who acts based on this information in an effort to manage his *reputation*. We develop and study a novel repeated game to analyze such environments. Our main insight is that reputational incentives can be exceedingly strong in such environments: the need to establish reputation forces even the good type agent to act inefficiently in *every* equilibrium (subject to a mild refinement). This can, in turn, result in large surplus losses due to market breakdown even when the principal knows that the agent is almost surely the good type.

### 1.1. *Summary of Model and Results*

Our framework is perhaps easiest to describe as a *bandit model with a strategic arm*. We use this metaphor throughout the paper. In each period, the principal (online content consumer, pharmaceutical company) chooses between experimenting (visiting the website, extending employment) with a costly risky arm of unknown type and a costless safe arm. The arm’s type is privately known by the agent (content provider, scientist) who also controls its output. If the arm is the good type, it receives private information (news stories, research ideas) of varying quality (the accuracy of reporting, the scientific basis for efficacy of the compound) at a Poisson rate; the bad type never receives any information. If the principal decides to experiment, the agent chooses whether or not to publicly act (publish a story, run a drug trial) based on the received information (if any). Acting in turn generates a public success or failure (the veracity of the story once cross checked by other news outlets, the outcome of the drug trial), the probability of which depends on the quality of information. In our model, acting based on high quality information always results in a success; acting on low quality information sometimes results in a success (and sometimes in a failure), while acting without information always generates a failure. The principal wants to simultaneously maximize number of successes (true stories, successful drug trials) and minimize failures, whereas the agent wants to maximize the duration of experimentation by the principal (number of website visits, period of employment).

Since actions are costless to the agent, there are no frictions in our environment if the agent is publicly known to the good type. In this “first-best” benchmark (Theorem 1), the principal-optimal Nash equilibrium strategy is for her to always experiment and, in response, the good type agent (who is indifferent between all strategies) acts efficiently, that is, only when he receives high quality information. To make the model interesting, we assume that acting on low quality (and, by definition, without) information is inefficient in that it generates a negative expected payoff for the principal. This Nash equilibrium is also the unique Pareto-efficient outcome. Conversely, if the agent is known to be the bad type, there is a unique Nash equilibrium outcome in which the principal never experiments. This is because experimentation is costly and the bad type can never generate positive payoffs for the principal.

Despite the lack of frictions, there are a multiplicity of Nash equilibria even when the agent is known to be the good type. Indeed, there is a Nash equilibrium in which there is complete market breakdown: the agent’s strategy is to never act. In response, the principal never experiments. We view such inefficient Nash equilibria to be unrealistic as they do not survive even the weakest notions of renegotiation-proofness. Thus, we impose a refinement which rules out these, and only these, types of Nash equilibria. Specifically, our refinement requires that both players play the efficient or first-best strategies at all on-path histories where the principal’s belief assigns probability 1 to the agent being the good type. In what follows, this refinement is implicit whenever we refer to equilibrium without using the additional “Nash” qualifier. We consider the refinement mild since it only restricts on-path equilibrium strategies at precisely one belief. In this sense, it is less restrictive than standard refinements used in the literature such as Markov perfection.

When there is type uncertainty, the principal experiments in order to give the agent a chance to reveal himself as the good type by generating a success. Of course, whether or not experimentation is worthwhile depends on how many failures the principal must suffer along the way. To make this tradeoff between the agent’s action strategy and the principal’s decision to experiment explicit, we consider a second “static” benchmark. Here, the principal experiments for a single period and stops experimenting if no success is generated; conversely, if a success is generated then the refinement bites (since, the principal’s belief jumps to 1) and the first-best continuation play results. Since only a success guarantees positive continuation value, the good type strictly best responds by acting on both high and low quality information; being indifferent, we assume he does not act in the absence of information.<sup>1</sup> The sign of the principal’s payoff in this benchmark determines her tolerance for what we call a lack of *quality control*; that is, when the good type agent always acts on low-quality information. When positive, these strategies constitute an equilibrium and, hence, the principal is willing to experiment even when the good type agent always acts on low-quality information.<sup>2</sup> Conversely, when the payoff in this benchmark is negative, *quality control is necessary* for experimentation. In our applications, this corresponds to an unwillingness to visit a website with a consistently low standard of reporting or hiring a scientist

<sup>1</sup>As we argue, this provides an upper bound for the principal’s payoff in this static benchmark.

<sup>2</sup>The bad type must also act with positive probability as otherwise acting alone will cause the principal’s belief to jump to one and so she will continue to experiment even after a failure. However, because our refinement only bites at belief one, this probability can be taken to be arbitrarily small so that the principal’s loss from the bad type is correspondingly small.

who is always willing to conduct speculative drug trials. In this case, the principal has to provide dynamic incentives to police the agent’s actions in order to receive an overall positive value from experimentation.

Our first result (part 1 of [Theorem 3](#)) shows that the principal can never completely prevent the agent from choosing inefficient actions in any equilibrium. Specifically, in every equilibrium where the principal experiments, there are on-path histories where the agent acts both on low-quality information and in the absence of any information. Part 2 of [Theorem 3](#) shows that the loss of surplus from this inefficiency can be large. Specifically, we show that, whenever quality control is necessary for experimentation, there is a unique equilibrium outcome in which the principal never experiments. This result shows that, whenever the principal needs to discipline the agent to make experimentation worthwhile, the agent’s need to establish reputation makes this impossible.

What makes our result stark is that breakdown can occur even with arbitrarily small amounts of uncertainty or even if “minimal” quality control can make experimentation profitable. We discuss each of these and their implications in turn. First, observe that there are parameter values such that quality control might be necessary even when the agent is almost surely known to be the good type. For instance, when failures are very costly to the principal, experimentation will not be profitable even with a high belief if the good type agent always acts on low-quality information. However, when the principal’s belief is high, it might be natural to expect that the agent’s reputational incentives are weaker and that he can be incentivized to act efficiently sufficiently often to make experimentation worthwhile. An implication of our main result is that this is not the case: the principal’s equilibrium payoff discontinuously drops from that of the first-best to zero as there is infinitesimal uncertainty about whether the agent is the good type. Returning to our first application, this implies that even a small amount of doubt in the veracity of a content provider (like a news source) can have a large impact for consumers. Specifically, the incentives that drive market breakdown in our model mirror the oft-cited criticism of the news media: the need to constantly generate new content drives down the quality of reporting in turn engendering public mistrust which contributes to the environment of dwindling web traffic.

When the principal’s payoff in the static benchmark is negative but small, this corresponds to the case where the principal would not experiment if the agent acts inefficiently at every opportunity but experimentation would be profitable if the agent chose to behave efficiently even a small fraction of the time. Our result implies that the principal cannot even provide the long run incentives to make efficient play occur at a small fraction of histories. In this sense, the loss of surplus due to reputational concerns can be large relative to the first best.

Strikingly, our main insight continues to hold even if the principal is allowed to make transfers ([Theorem 4](#)). Specifically, we show that inefficiency and market breakdown occur (under the exact same conditions) even if we allow one-sided transfers—the principal can make a payment to the agent after the outcome from the agent’s action is observed. This generalizes our main insight to a broader set of applications (such as our second example) where such transfers are a natural feature of the environment. This result suggests that performance-based bonuses cannot correct an expert’s strong reputational incentive to speculatively act as opposed to candidly admitting that he may not have any good information.

How then can the functioning of such markets be improved? We argue that a natural way in which market breakdown can be avoided is if the principal can commit to a longer duration of experimentation: say  $T > 1$  periods at a time. For our two highlighted applications, this would amount to getting a subscription for online content or providing the expert with a fixed contract length respectively and only reevaluating the decision to continue experimentation after this fixed period expires. In the language of our model, this would amount to altering the stage game so that after the principal decides to experiment, the agent sequentially receives information and makes action choices  $T$  times (with the cost of experimentation appropriately adjusted so we can make an apples-to-apples comparison). We argue that this commitment to longer experimentation can weaken the agent’s need to immediately establish reputation which in turn can make experimentation profitable in cases where it would not have been if  $T = 1$ . Thus, our model provides one rationale for why subscriptions might be a more efficient way to pay for online content and why longer contracts might reduce the amount of bad advice provided by experts in an attempt to prove their worth.

Finally, we explicitly highlight the role played by our refinement by showing that our main insights do not hold if it is dropped. Specifically, we construct Nash equilibria where the agent chooses efficient actions on path ([Theorem 5](#)). Moreover, we argue that the discontinuity of payoffs at belief one disappears: as the uncertainty about the agent’s type vanishes, there are Nash equilibria such that the payoffs to both players converge to the first-best. These equilibria (necessarily) have the unrealistic feature that there are on-path histories where the market breaks down even though incentives are perfectly aligned as the agent has revealed himself to be the good type with probability 1.

## 1.2. *Related Literature*

Our paper is most closely related to the literature on reputation in repeated games. One key difference between our paper and much of the remaining literature is that we consider a model with two-long lived players (one of whom has multiple strategic types) and that we can make robust equilibrium predictions without requiring either player to be arbitrarily patient.

Perhaps the closest work in the reputation literature is [Ely and Välimäki \(2003\)](#) (in particular, the refinement we impose is taken from their work). They also consider a two player repeated game and establish a “bad reputation” result where the reputational incentives of a long lived agent with a privately known type causes the loss of all surplus when faced with a sequence of short-lived principals. At a superficial level, our model differs because our game does not have the payoff structure of a bad reputation game (in the sense of [Ely, Fudenberg, and Levine \(2008\)](#)) and, additionally, our market breakdown result does not require the discount factor to approach one. More importantly, the forces that prevent the market from functioning are very different; because the principals are short lived in their setting, they do not internalize the future value that is gained from experimentation. Indeed, they show that their bad reputation result does not hold if both players are long lived. In [Ely, Fudenberg, and Levine \(2008\)](#), breakdown can occur because the short run temptation to separate may make it impossible to get good types to behave well. In our paper, breakdown occurs because eventually there will be a history where short run incentives

dominate, which reduces the incentives for good types to behave well all along the path to those histories.

[Bar-Isaac \(2003\)](#) studies a signaling framework where a seller of privately known quality can choose whether or not to sell a good in each period to a sequence of short-lived buyers. Like our model, the seller's quality determines the likelihood with which the good is publicly revealed to be a success or failure. Perhaps the most critical distinctions are that (i) we have additional private information that the agent receives periodically, and (ii) we aim to isolate robust properties of all equilibria. He conversely shows that there exists a Markov perfect equilibrium that delivers his positive result (the good-type seller never stops selling).

There is a comparatively smaller fraction of the reputation literature that analyzes repeated games with two long-lived equally patient players. An early example is [Cripps and Thomas \(1997\)](#), more recent papers are [Atakan and Ekmekci \(2012, 2013\)](#). These papers have fundamentally different goals in that they aim to establish when reputation can be achieved (in the sense that the player with the unknown type can attain the Stackelberg payoff) as players become arbitrarily patient. We, by contrast, are interested in analyzing robust equilibrium outcomes for impatient players. We can do so because we analyze a very specific game; the above mentioned papers work with more general classes of games (to the best of knowledge, the game we study does not lie in the class of any paper in this strand of the literature). Framed this way, our paper can also be thought of as an instance of a relational contracting problem; perhaps the closest related recent papers are [Li, Matouschek, and Powell \(2017\)](#) and [Mitchell \(2017\)](#) who analyze relational contracting models without transfers, but where the problem is one of moral hazard without adverse selection. The environment, applications and focus of these papers are quite distinct from ours.

While otherwise very different, the reputational incentives (and the fact that they can distort behavior) in our setting are similar in spirit to those in models where experts with private types choose actions in an attempt to demonstrate competence ([Prendergast and Stole \(1996\)](#), [Morris \(2001\)](#), [Ottaviani and Sørensen \(2006\)](#)). More recently, [Backus and Little \(2018\)](#) consider a single period, extensive-form game of expert advice where they derive conditions under which an expert can admit uncertainty. Our setting shares an essential modeling feature that good types may not be able to provide the principal with positive utility in all periods.

Finally, our paper is related to a few papers in the literature on dynamic mechanism design which analyze outcomes both with and without commitment. [Guo \(2016\)](#) studies a dynamic setting without transfers where an agent is privately informed about the quality of a risky arm and he prefers greater experimentation than the principal (both prefer the risky arm only when it is good). While our setting differs substantially in terms of the payoff structure (and, hence the application to which we speak), another critical difference is that the agent in our environment strategically acts in response to additional private information she receives over time. [Aghion and Jackson \(2016\)](#) consider a political economy setting where voters (principal) must incentivize a politician (agent). Formally, they consider a setting without transfers where a principal is trying to determine the type of a long-lived agent. While some features of our game are similar (the agent's payoff and the fact that she receives private information in each period), the main driving forces in their model are different. Specifically, signaling is not a source of inefficiency in their

setting; instead, the principal wants to the agent to take “risky” actions that are potentially damaging to the latter’s reputation. [Deb, Pai, and Said \(2018\)](#) study a pure adverse selection dynamic environment without transfers where a principal is trying to determine whether a strategic political forecaster has accurate information by comparing the dynamics of his predictions leading up to an election with the eventual winner. Once again, there are some superficial similarities in the payoff structure but the games are otherwise quite different.

## 2. THE MODEL

In this section, we set up the main model on which the bulk of the paper focuses. In the latter half of [Section 4](#), we discuss which of the assumptions can be altered without affecting our main results.

We study a discrete time, infinite horizon repeated game of imperfect public monitoring between a principal and an agent. We denote time by  $t \in \{1, \dots, \infty\}$  and both players have a common discount factor  $\delta \in (0, 1)$ .

*Agent’s Initial Type:* The agent starts the game with a privately known type  $\theta$  which can either be good ( $\theta_g$ ) or bad ( $\theta_b$ ). The agent is the good type  $\theta_g$  with commonly known prior probability  $0 < p_0 < 1$ . This initial type determines the rate at which the agent can generate positive payoffs for the principal.

We begin by describing the stage game (summarized in [Figure 1](#)) after which we define strategies and Nash equilibrium.

### 2.1. The Stage Game

At each period  $t$ , the principal and agent play the following extensive-form stage game where the order of our description matches the timing of moves.

*Principal’s Action:* The principal begins the stage game by choosing whether or not to experiment  $x_t \in \{0, 1\}$ , where  $x_t = 1$  corresponds to experimenting. The cost of action  $x_t$  is  $cx_t$  with  $c > 0$  so experimenting is costly.

If the principal chooses not to experiment, the stage game ends. When she experiments, the stage game proceeds as follows.

*Agent’s Information:* In each period, the agent either receives no information ( $i_t = i_\varphi$ ), low-quality information ( $i_t = i_l$ ) or high-quality information ( $i_t = i_h$ ).  $i_t$  is drawn independently in each period from a distribution that depends on the the agent’s type  $\theta$ .

If  $\theta = \theta_b$ , the agent never receives information: that is,  $i_t = i_\varphi$  with probability 1.

If  $\theta = \theta_g$ , the agent independently draws information  $i_t \in \{i_l, i_h\}$  with probability  $\lambda \in (0, 1)$  and, hence, receives no information  $i_t = i_\varphi$  with the complementary probability  $1 - \lambda$ . Conditional on receiving information, it is high, low quality with probability  $\frac{\lambda_h}{\lambda}, \frac{\lambda_l}{\lambda}$  respectively (where  $\lambda_h + \lambda_l = \lambda$ ).

*Agent’s Action:* After his information is realized, the agent decides whether or not to costlessly act based on his information. Formally, he picks a public action  $a_t \in \{0, 1\}$  where  $a_t = 1(0)$  denotes (not) acting.

*Public Outcomes:* If the principal experiments ( $x_t = 1$ ) and the agent acts ( $a_t = 1$ ), a public outcome  $o_t \in \{\bar{o}, \underline{o}\}$  is realized from a distribution  $\mu$  given by

$$\mu(\bar{o} | i_t) = \begin{cases} 1 & \text{if } i_t = i_h, \\ q_l & \text{if } i_t = i_l, \\ 0 & \text{if } i_t = i_\varphi, \end{cases}$$

where  $\underline{o}$  is realized with the complementary probability.

A success ( $\bar{o}$ ) can only be generated by the good type and the likelihood of a success is determined by the information quality: acting on high (low) quality information always (sometimes) generates a success. The tension in the model arises from the fact that the good type may want to act on low-quality information in an effort to signal his type. Since this can generate failures ( $\underline{o}$ ), the bad type may act in an attempt to pool even though he never receives any information.

If either the principal does not experiment ( $x_t = 0$ ) or the agent does not act ( $a_t = 0$ ), the stage game ends with the agent generating neither a success nor failure. We denote this outcome by  $o_t = o_\varphi$ . Note that the extensive form implies that the agent does not receive information or get to move if the principal chooses the safe arm; to simplify notation, we define  $a_t = 0$  and  $i_t = i_\varphi$  when  $x_t = 0$ .

We use the shorthand notation  $h_t = (x_t, a_t, o_t)$ ,  $h_t \in \{\not{h}, h_\varphi, \bar{h}, \underline{h}\}$  to describe the outcome of the stage game. Here,

$$\begin{aligned} \not{h} &:= (x = 0, a = \varphi, o = \varphi), & h_\varphi &:= (x = 1, a = 0, o = o_\varphi), \\ \bar{h} &:= (x = 1, a = 1, o = \bar{o}), & \underline{h} &:= (x = 1, a = 1, o = \underline{o}). \end{aligned}$$

In words,  $\not{h}$  denotes the case where principal chooses not to experiment, the remaining three correspond to the separate outcomes that can occur after the principal experiments.  $h_\varphi$  denotes the case where the agent does not act, and  $\bar{h}$ ,  $\underline{h}$  denote the cases where the agent acts and a success, failure respectively are observed.

We use time superscripts to denote vectors. Thus,  $h^t = (h_1, \dots, h_t)$  and  $i^t = (i_1, \dots, i_t)$ . Additionally,  $h^{t'} h^t$  (and analogously for other vectors) denotes the  $t' + t$  length vector where the first  $t'$  elements are given by  $h^{t'}$  and the  $t' + 1^{\text{st}}$  to  $t' + t^{\text{th}}$  elements are given by  $h^t$ .

*Stage Game Payoffs:* The agent receives a unit payoff whenever the principal experiments. Formally, his stage-game payoff  $u$  is given by

$$u(x_t) = x_t.$$

The principal wants to maximize (minimize) the number of successes (failures). Her payoff  $v$  is given by

$$v(h_t) = \begin{cases} 1 - c & \text{if } h_t = \bar{h}, \\ -\kappa - c & \text{if } h_t = \underline{h}, \\ -c & \text{if } h_t = h_\varphi, \\ 0 & \text{if } h_t = \not{h}. \end{cases}$$

In words, gross of cost, the principal realizes a normalized payoff of 1 for every success, a loss of  $\kappa > 0$  for every failure and 0 otherwise.

*Payoff Assumptions:* We assume that acting on low quality information yields a net loss for the principal:

$$q_l < (1 - q_l)\kappa.$$

This assumption creates one of the key tradeoffs in the model; absent its signaling value, the principal wants to prevent the agent from acting on low quality information.

Additionally, we assume that the cost of experimentation

$$c < \lambda^h$$

is sufficiently low to make the model nontrivial. If this assumption is not satisfied, the principal will never experiment even if the agent is known to be the good type.

## 2.2. The Repeated Game

*Histories:*  $h^{t-1}$  denotes the public history (henceforth, referred to simply as a history) at the beginning of period- $t$ . This contains all the previous actions of and outcomes observed by both players. The good type agent's private history additionally contains all previous information realizations  $i^{t-1}$  and the period- $t$  information  $i_t$  when he is deciding whether or not to act; recall, the bad type never receives any information. We use the convention that  $h^0 = \varphi$  denotes the start of the game. We use  $\mathcal{H}$  and  $\mathcal{H}^g$  to respectively denote the set of histories and the set of the good type agent's private histories.<sup>3</sup>

*Agent's Strategy:* We denote the agent's strategy by  $\tilde{a}_\theta$ . When the agent is the bad type,  $\tilde{a}_{\theta_b}(h^{t-1}) \in [0, 1]$  specifies the probability with which the agent acts at each period  $t$  as a function of the history  $h^{t-1} \in \mathcal{H}$ . When the agent is the good type,  $\tilde{a}_{\theta_g}(h^{t-1}, i^t) \in [0, 1]$  specifies the probability with which he acts at each period  $t$  as a function of his private history  $(h^{t-1}, i^t) \in \mathcal{H}^g$ . Since the agent can only act when  $x_t = 1$ , this is implicitly assumed in the notation and we do not add this as an explicit argument of  $\tilde{a}_\theta$  for brevity.

*Principal's Strategy:* The principal's strategy  $\tilde{x}(h^{t-1}) \in [0, 1]$  specifies the probability of experimenting in each period  $t$  as a function of the history.

*On- and off-path Histories:* Given strategies  $\tilde{x}$  and  $\tilde{a}_\theta$ , a history  $h^{t-1} \in \mathcal{H}$  is said to be on-path (off-path) if it can (cannot) be reached with positive probability for either (both) of the agent's possible types  $\theta \in \{\theta_g, \theta_b\}$ . Similarly, a private history  $(h^{t-1}, i^t) \in \mathcal{H}^g$  is said to be on-path (off-path) if it can (cannot) be reached with positive probability when  $\theta = \theta_g$ .

*Beliefs:* A belief  $\tilde{p}$  is associated with a pair of strategies  $\tilde{x}, \tilde{a}_\theta$  (we suppress explicit dependence on the strategies for notational convenience).  $\tilde{p}(h^{t-1})$  is principal's belief that the agent is the good type at history  $h^{t-1}$ . If this history is on-path,  $\tilde{p}(h^{t-1})$  is derived from the agent's strategy  $\tilde{a}_\theta$  by Bayes' rule. In the entirety of what follows, we impose no restriction on how off-path beliefs are formed.

<sup>3</sup>Note that  $\mathcal{H}^g$  only contains histories where the information received by the good type is compatible with the outcomes for the stage game; recall that the successes cannot arise when the good type has no information.

*Expected Payoffs:* We use  $U_{\theta_g}(h^{t-1}, i^{t-1}, \tilde{x}, \tilde{a}_{\theta_g})$ ,  $U_{\theta_b}(h^{t-1}, \tilde{x}, \tilde{a}_{\theta_b})$ ,  $V(h^{t-1}, \tilde{x}, \tilde{a}_\theta)$  to denote the expected payoff of the good, bad type agent and principal respectively at histories  $h^{t-1} \in \mathcal{H}$ ,  $(h^{t-1}, i^{t-1}) \in \mathcal{H}^g$  given strategies  $\tilde{x}, \tilde{a}_\theta$ . Note that the principal's payoff implicitly depends on her belief  $\tilde{p}(h^{t-1})$  at this history.

*Nash Equilibrium:* A Nash equilibrium consists of a pair of strategies  $\tilde{x}, \tilde{a}_\theta$  such that they are mutual best responses. Formally, this implies that  $U_{\theta_g}(h^0, i^0, \tilde{x}, \tilde{a}_{\theta_g}) \geq U_{\theta_g}(h^0, i^0, \tilde{x}, \tilde{a}'_{\theta_g})$ ,  $U_{\theta_b}(h^0, \tilde{x}, \tilde{a}_{\theta_b}) \geq U_{\theta_b}(h^0, \tilde{x}, \tilde{a}'_{\theta_b})$  and  $V(h^0, \tilde{x}, \tilde{a}_\theta) \geq V(h^0, \tilde{x}', \tilde{a}_\theta)$  for any other strategies  $\tilde{x}', \tilde{a}'_\theta$ .

A Nash equilibrium is *non-trivial* if there is an on-path history  $(h^t)$  where the principal experiments with positive probability ( $\tilde{x}(h^t) > 1$ ).

Figure 1 below summarizes the actions and the order of play in the repeated game.

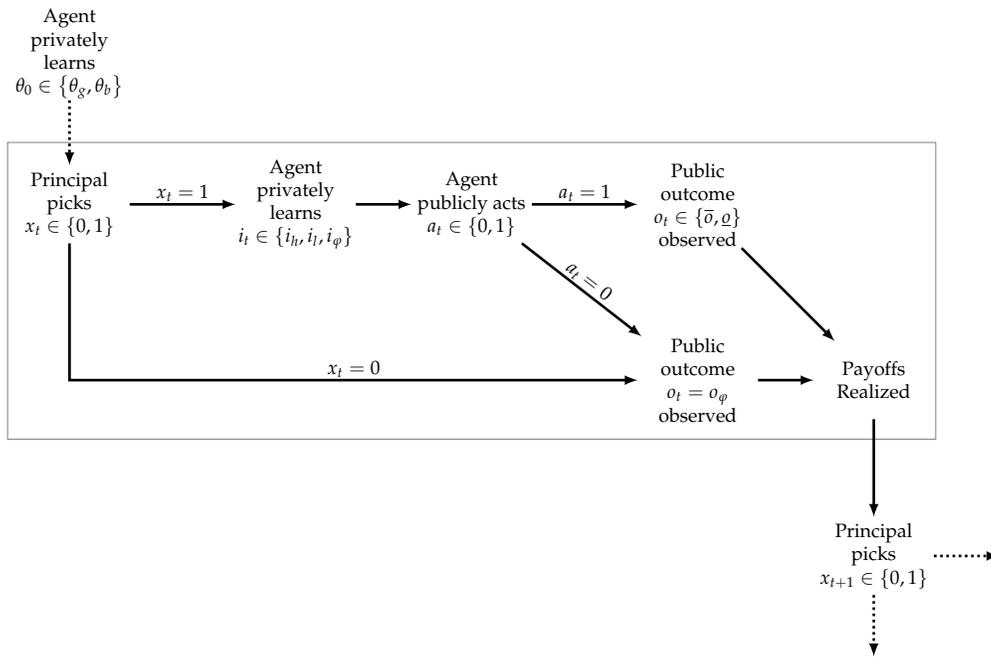


FIGURE 1. Flow chart describing the repeated game; the rectangle contains the stage game.

Before beginning the analysis, it is worth discussing the main assumption in our model: the monitoring structure mapping the agent's actions based on information to outcomes. As is common in bandit models, we consider a "good news" environment where only the good type can generate successes which are hence perfectly revealing. As we will discuss below, this assumption is deliberately stark. It is intended to highlight that market breakdown can arise even though the good type can separate perfectly at histories where he receives high-quality information and, hence, one might expect screening to be possible.

This assumption does have implications for the applications that correspond to our model. For our second leading application, the implication is clear; drug trials with no scientific basis will not succeed. For our first application, the monitoring structure implies that we are taking a specific stand on the meaning of "fake" content that is a result of acting without information.

Content providers can claim to have inside information, and their scoops could turn out, after public scrutiny, to be either correct or false. The key tension in our model is that the arrival of novel, accurate stories is stochastic. A genuine content provider, behaving efficiently, might go long stretches without a scoop. If so, they have strong incentives to publish content even though they have very little certainty about its veracity. In other words, fake content in our model is specifically an attempt by news sources to appear timely; it is *not* an attempt to confuse the public about actual facts (so should not be confused with the disinformation “fake news” campaigns designed to sway public opinion).

### 3. SOME BENCHMARKS AND THE EQUILIBRIUM REFINEMENT

We begin by considering some simple benchmarks which provide context for our main result. [Section 3.1](#) considers a setting where there is no incomplete information. This motivates the refinement that we define in [Section 3.2](#). Finally, we consider a “static” benchmark in [Section 3.3](#) that isolates the key tradeoffs in the model.

#### 3.1. The First-Best Complete Information Benchmark

The purpose of this section is to show that the key friction in the model is the principal’s need to screen the good from the bad type. To do so, we consider a benchmark in which the agent’s type is publicly known; formally, this corresponds to the case where the prior belief  $p_0 \in \{0, 1\}$ . The theorem below characterizes the “first-best” Nash equilibrium under complete information. As we often do in the paper, formal statements are presented in words without notation to make them easier to read. Where we feel it is helpful to the reader, we additionally describe the strategies using the notation in footnotes (and these mathematical statements also appear in the appendices).

**THEOREM 1.** *If the agent is known to be the bad type, there is a unique Nash equilibrium in which the principal never experiments.*<sup>4</sup>

*Conversely, suppose the agent is known to be the good type. Then, there is a unique Pareto-optimal Nash equilibrium outcome in which the principal always experiments and the agent only acts on high quality information.*<sup>5</sup>

This result is obvious so the following discussion serves as the proof. The principal never experiments when the agent is known to be the bad type since experimentation is costly and this type cannot generate any successes which are the only outcomes that yield a positive payoff to the principal. Conversely, if the agent is known to be the good type, there are no frictions. As long as the agent acts only on high-quality information, it is always profitable for the principal to experiment. Since actions are costless, the agent is indifferent between all strategies and so, in particular, always choosing the principal optimal action is a best response. In what follows, whenever we describe the behavior of either player as *efficient*, we are referring to the strategies described in

<sup>4</sup>Formally, the principal’s unique Nash equilibrium strategy when  $p_0 = 0$  is  $\bar{x}(h^{t-1}) = 0$  for all  $h^{t-1} \in \mathcal{H}$ . The agent’s strategies can be picked arbitrarily.

<sup>5</sup>Formally, the Pareto-optimal Nash equilibrium strategies when  $p_0 = 1$  are  $\bar{x}(h^{t-1}) = 1$ ,  $\bar{a}_{\theta_b}(h^{t-1}) = 0$  for all  $h^{t-1} \in \mathcal{H}$  and  $\bar{a}_{\theta_g}(h^{t-1}, i^{t-1}i_t) = 1(0)$  for all  $(h^{t-1}, i^{t-1}i_t) \in \mathcal{H}^8$  where  $i_t = (\neq)i_h$ .

Theorem 1 above. We use

$$\bar{\Pi} := \frac{\lambda^h - c}{1 - \delta} > 0$$

to denote the first-best payoff that the principal obtains from the good type which clearly is an upper bound for the payoff that the principal can achieve in any Nash equilibrium for  $p_0 \in [0, 1]$ .

### 3.2. The Equilibrium Refinement

Despite the lack of frictions when the agent is known to be the good type, there exist other, inefficient Nash equilibria where the principal does not always experiment because the agent decides to act inefficiently at certain histories. Indeed, complete market breakdown is also an equilibrium: the principal never experiments and, in response, the agent never acts. We view such Nash equilibria as “implausible” as there is nothing that prevents both players from moving to the Pareto-superior equilibrium described in Theorem 1. Similar unrealistic Nash equilibria also arise in the model Ely and Välimäki (2003) and we borrow the following refinement from them to eliminate these, and only these, Nash equilibria.

*Equilibrium Refinement:* An equilibrium (with no additional qualifier) is a Nash equilibrium in which, at on-path histories where the principal believes that the agent is the good type with probability one, both players choose efficient strategies (formally described in Theorem 1) in the continuation game.<sup>6</sup>

While Ely and Välimäki (2003) refer to this refinement as renegotiation-proofness, we choose not to do so as we feel this terminology overstates the strength of the refinement. This is because it only restricts equilibrium behavior at a single belief whereas renegotiation-proofness refinements typically apply at all histories. To the best of our knowledge, there is no single, accepted renegotiation-proofness refinement in repeated games with uncertainty but the above refinement (imposed after the agent’s type uncertainty is resolved) is implied by many renegotiation-proofness refinements in the literature for complete information games. For instance, the consistency requirements in Farrell and Maskin (1989), Abreu and Stacchetti (1993) and, most recently, the sustainability requirement (without renegotiation frictions) of Safronov and Strulovici (2017) would yield the efficient outcome when the agent’s type is initially known to be good.<sup>7</sup>

Finally, note that the above refinement is, in a sense, weaker than standard refinements such as Markov perfection. This is, once again, because the refinement only has bite at one belief whereas Markov perfection restricts equilibrium strategies at all beliefs. That said, note that the strategies required by our refinement are not implied by Markov perfection as the latter does not prevent both players from behaving inefficiently when the belief is one. Note that, with our refinement in place, inefficiency is not a consequence of the players choosing inefficient strategies under complete information; instead, inefficiency arises exclusively from the information frictions.

<sup>6</sup>Formally, at all on-path histories  $h^{t-1} \in \mathcal{H}$  such that  $\tilde{p}(h^{t-1}) = 1$ , we have  $\tilde{x}(h^{t-1}) = 1$ ,  $\tilde{a}_{\theta_b}(h^{t-1}) = 0$  and, for all  $(i^{t-1}, h^{t-1}) \in \mathcal{H}^S$ ,  $\tilde{a}_{\theta_g}(h^{t-1}, i^{t-1}i_t) = 1(0)$  when  $i_t = (\neq)i_h$ .

<sup>7</sup>All three are defined for complete information repeated games with normal form stage games but can be applied to our extensive-form stage game as well.

### 3.3. The “Static” Benchmark and the Need for Quality Control

The goal of this section is to highlight the tradeoff between the agent’s short term need to establish reputation and the principal’s value from experimentation. To do so, we examine a “static” benchmark where experimentation only occurs once, but the full dynamic benefits of that single experiment are calculated. When the payoff in this benchmark is negative, then experimentation is only potentially worthwhile if the principal can use the dynamics of the relationship to dampen the agent’s reputational incentives. Specifically, consider the following strategies.

**Principal:** Experiments in the first period. If no success is generated, she stops experimenting at all subsequent histories. If a success is generated, she always experiments at all future histories.

**Good type agent:** Acts if and only if he receives information of either high or low quality in the first period. If no success is generated, he never acts whereas if a success is generated, he only acts on high quality information at all future private histories.

**Bad type agent:** Acts with positive probability in the first period and then stops acting.

Formally, these “static” strategies are given by

$$\begin{aligned}
 \tilde{x}(h^0) = 1 \quad \text{and} \quad \tilde{x}(h_1 h^{t-1}) &= \begin{cases} 1 & \text{if } h_1 = \bar{h}, \\ 0 & \text{if } h_1 \neq \bar{h}, \end{cases} \quad \text{for all } h_1 h^{t-1} \in \mathcal{H}, t \geq 1, \\
 \tilde{a}_{\theta_g}(h^0, i_1) &= \begin{cases} 1 & \text{if } i_1 \in \{i_h, i_l\}, \\ 0 & \text{if } i_1 = i_\varphi, \end{cases} \\
 \tilde{a}_{\theta_g}(h_1 h^{t-1}, i^t i_{t+1}) &= \begin{cases} 1 & \text{if } h_1 = \bar{h} \text{ and } i_{t+1} = i_h, \\ 0 & \text{otherwise,} \end{cases} \quad \text{for all } (h_1 h^{t-1}, i^t) \in \mathcal{H}^g, t \geq 1. \\
 \tilde{a}_{\theta_b}(h^0) > 0 \quad \text{and} \quad \tilde{a}_{\theta_b}(h^t) = 0, & \quad \text{for all } h^t \in \mathcal{H}, t \geq 1.
 \end{aligned} \tag{1}$$

To summarize, these strategies correspond to the principal experimenting for a single period in the hope that the agent produces a success in which case she receives the first-best continuation payoff. If no success is generated, the principal stops experimenting and the game effectively ends. Faced with this strategy, it is a strict best response for the good type to act both on high and low quality information because only successes generate positive continuation payoffs. The bad type is indifferent between acting or not (as is the good type when he does not receive information), so acting with positive probability is, in particular, a best response. We term these strategies as static (despite the game continuing after a success) because both players’ incentives in this game are identical to a static game in which they receive positive (first-best) payoffs only after a success.

An upper bound for the principal’s payoff from these strategies is given by

$$\underline{\Pi}(p_0) := p_0(\lambda_h + \lambda_l q_l)(1 + \delta \bar{\Pi}) - p_0 \lambda_l (1 - q_l) \kappa - c.$$

This is the expected payoff that the principal receives from the actions of the good type; it is an upper bound because it does not account for the losses from failures generated by the bad type. The first term corresponds to the payoff after a success (the principal gets a payoff of 1 in period one and the first-best continuation payoff  $\bar{\Pi}$ ) and the second term is the loss from a failure.

It is straightforward to observe that a necessary and sufficient condition for these strategies to constitute an equilibrium is that this bound is positive.

**THEOREM 2.** *There exists an equilibrium in static strategies (given by (1)) iff  $\underline{\Pi}(p_0) > 0$ .*

When  $\underline{\Pi}(p_0) > 0$ , there is always a low enough positive probability  $\varepsilon > 0$  of action by the bad type ( $\tilde{a}_{\theta_b}(h^0) = \varepsilon$ ) such that the principal's payoff from experimentation is positive. Note that, as long as the bad type sometimes, but not always, acts ( $\tilde{a}_{\theta_b}(h^0) \in (0, 1)$ ), the principal's belief following both non-success outcomes is interior (since  $\tilde{p}(\underline{h}), \tilde{p}(h_\varphi) < 1$ ) and thus the continuation play as prescribed by (1) is allowed by the refinement since it has no bite at these histories.

Conversely, when  $\underline{\Pi}(p_0) \leq 0$ , it is a strict best response for the principal to not experiment in period one. This is because her payoff is strictly lower than this upper bound since the bad type acts with positive probability. Note that there cannot be an equilibrium where the bad type does not act ( $\tilde{a}_{\theta_b}(h^0) = 0$ ). Suppose to the contrary that there was such an equilibrium. In this case, only the good type acts and so, irrespective of outcome, the principal's posterior belief must go to one after an action ( $\tilde{p}(h_1) = 1$  for  $h_1 \in \{\bar{h}, \underline{h}\}$ ). Our refinement implies that this will result in positive continuation value for the agent and so acting ( $\tilde{a}_{\theta_b}(h^0) = 1$ ) will be a strict best response for the bad type which is a contradiction.

When  $\underline{\Pi}(p_0) \leq 0$ , we say *quality control is necessary* for experimentation to generate a positive expected payoff. In this case, the principal must use the dynamics of her relationship with the agent in order to provide the good type with the necessary incentives to act sufficiently infrequently on low-quality information. Note that the monitoring structure we have chosen in our model is intentionally stark. By assuming successes are perfectly revealing, it should be easier for the good type to establish reputation compared to a setting where acting on no information could also generate a success with positive probability. In other words, one might think that screening would be relatively easier for the principal compared to the case where bad types too could sometimes produce successes. Our main insight is that inefficiency and market breakdown arise *despite* this stark good news monitoring structure.

## 4. MAIN RESULTS

In this section we present the main results of the paper and show that our key insights are robust to allowing for one-sided transfers and a few other alterations of the model.

### 4.1. Equilibrium Properties: Inefficiency and Market Breakdown

We now present the main result of the paper in two parts. We first show that it is impossible for the principal to provide incentives that uniformly prevent inefficient actions on path and then argue that this inefficiency can have extreme payoff consequences.

**THEOREM 3.** *Suppose  $p_0 \in (0, 1)$ . Then:*

- (1) *In every nontrivial equilibrium, the good type agent acts on low-quality information and the bad type agent acts with positive probability (on no information) on path.*<sup>8</sup>
- (2) *If quality control is necessary ( $\underline{\Pi}(p_0) \leq 0$ ), the unique equilibrium outcome is that the principal never experiments.*

Taken together, these results have stark economic implications. The first statement of [Theorem 3](#) implies that a principal who is willing to experiment will necessarily face inefficient actions. Phrased in terms of our first application, this implies that a consumer of online content must resign herself to the reality that, in addition to tolerating fake content from bad providers, even genuine providers will willingly publish low-quality content that has not been fully vetted. The inevitability of fake content follows from a simple argument that we have already articulated in the context of the static benchmark. At any history where the good type acts, the bad type must also act; if he did not, acting alone will take the principal’s posterior belief to one which guarantees the highest possible continuation value due to our refinement and hence, not acting cannot be a best response. The reason low quality content from the good type always arises on path, which is the critical feature of the equilibrium and the potential cause for breakdown, is more subtle and we postpone the intuition for this until later in the section.

The second statement of [Theorem 3](#) argues that the payoff implications of the above mentioned inefficiency can be large. Specifically, the agent’s reputational incentives prevent any quality control whenever it is necessary for experimentation and hence, the principal can never internalize the long run benefits of experimentation. This result highlights perhaps the most significant economic difference between our results and those of [Ely and Välimäki \(2003\)](#) in whose setting market breakdown occurs precisely because, and only when, the principal is not long-lived and thus she does not account for the future value of experimentation.

A consequence of this result is that there are two separate discontinuities that arise in the equilibrium payoff set. First observe that there are parameter values such that  $\underline{\Pi}(1) \leq 0$ ; for instance, this can arise when experimentation is costly (high  $c$ ), failures yield large losses (high  $\kappa$ ) or when low-quality information is very unlikely to generate successes (low  $q_l$ ). In this case,  $\underline{\Pi}(p_0) \leq 0$  for all  $p_0 \in (0, 1)$  and so the principal will never experiment. Thus, both players’ payoffs discontinuously fall from the first-best when  $p_0 = 1$  to zero when  $p_0 < 1$ .

Similarly, if we fix the prior  $p_0$  but instead alter the parameter values so that the payoff upper bound from the static benchmark  $\underline{\Pi}(p_0) \downarrow 0$ , the agent’s payoff set once again discontinuously shrinks to zero. As long as  $\underline{\Pi}(p_0) > 0$ , there always exists at least one nontrivial equilibrium (described in [Section 3.3](#)) where the principal experiments in period one and the good type can attain the first best continuation payoff with probability  $\lambda_h + \lambda_l q_l$ .

Returning to our application, the implication of these results is that small additional frictions can have large impacts on the functioning of online content markets. For instance, American politicians have taken to decrying unfavorable coverage as “fake news” and now pounce on minor reporting mistakes as evidence of orchestrated bias. In our model, we can think of these as

<sup>8</sup>Formally, in every nontrivial equilibrium, there exist on-path histories  $h^t \in \mathcal{H}$  and  $(h^t, i^t i_l) \in \mathcal{H}^g$  such that  $\tilde{a}_{\theta_b}(h^t) > 0$  and  $\tilde{a}_{\theta_g}(h^t, i^t i_l) = 1$ .

lowering the belief in the content provider  $p_0$  or increasing the perceived disutility from failures—changes that can switch the sign of  $\underline{\Pi}$ —which can in turn result in a large drop of web traffic.

We end the section with the main intuition underlying [Theorem 3](#). We first observe that there is a positive belief below which the principal does not experiment in any Nash equilibrium. Since experimentation is costly, there is a belief below which experimentation is not worth it even if uncertainty about the agent’s type could somehow surely resolve itself with one period of experimentation (and be followed by the first-best continuation equilibrium).

Now suppose, to the contrary, that there is a nontrivial equilibrium despite quality control being necessary. Then, after each time the principal experiments, her posterior belief following one of the two non-success outcomes—failure or inaction—must not only fall (because beliefs follow a martingale) but must fall at a sufficiently fast rate (because successes must arrive frequently enough to make experimentation profitable). It is then possible to show that every nontrivial equilibrium must eventually have a “last” on-path history at which the principal’s belief  $p$  is less than  $p_0$  and at which the agent must generate a success or otherwise the principal stops experimenting. But note that the agent’s incentives at such a last history are identical to his incentives in the static benchmark. Here, it will be a strict best response for the good type to act on low quality information and, since  $\underline{\Pi}(p_0) \leq 0$  implies  $\underline{\Pi}(p) \leq 0$  for all  $p \leq p_0$ , it cannot be a best response for the principal to experiment. Thus, such a last history cannot exist contradicting the existence of a nontrivial equilibrium.

Very loosely speaking, the intuition underlying our result is similar to the reason that there is a unique equilibrium outcome in a finitely repeated Prisoner’s dilemma. The difference is that the backward induction “unraveling” in our model occurs in the space of beliefs (importantly, without having to appeal to Markov perfection) as opposed to calendar time.

#### 4.2. Allowing Transfers

For some applications, it is more appropriate to consider a setting where the principal can make transfers to the agent. For instance, in addition to a fixed wage, firms can choose to pay contingent bonuses to experts they employ. We introduce transfers by altering the stage game to allow the principal to make payments to the agent after the public outcome is observed. We assume that the principal can only make transfers in periods where she experiments. This assumption is not required for our result ([Theorem 4](#)); instead, we impose it because we feel it is realistic for our applications—it would be highly unusual for a firm to be making bonus payments to an expert not in their employ—and it shortens the proof. Formally, we denote the transfer in period  $t$  by  $\tau_t \geq 0$ . These transfers are observed by both players and so become part of the public history. For easy reference, [Figure 2](#) below describes how this alters the stage game. The formal description of histories and strategies can be found in the appendix.

We do not alter our equilibrium refinement: at any on-path history where the principal’s belief is one, she experiments and the agent only acts on high quality information. Importantly, the refinement applied to this version of the game does not additionally restrict transfers in any way and so the principal can transfer an arbitrary share of her first-best profits to the agent after he reveals himself to be the good type.

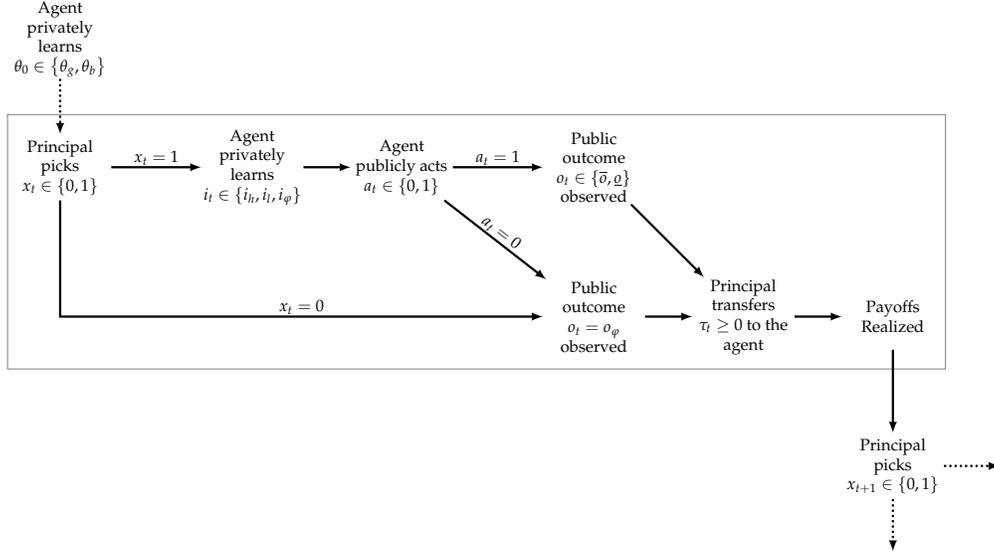


FIGURE 2. Flow chart describing the repeated game with transfers (the rectangle contains the stage game)

**THEOREM 4.** *Suppose  $p_0 \in (0, 1)$ . Then:*

- (1) *In every nontrivial equilibrium of the game with transfers, the good type agent acts on low-quality information and the bad type agent acts (on no information) on path.*
- (2) *If quality control is necessary ( $\underline{\Pi}(p_0) \leq 0$ ), the unique equilibrium outcome of the game with transfers is that the principal never experiments.*

Transfers do not help the principal avoid inefficient actions or the resulting market breakdown arising from the agent’s reputational incentives. In a nutshell, inefficiency is the result of the agent’s desire to generate successes and one-sided transfers from the principal can only further reward successes instead of punishing failures. To see this, first note that, once again, every non-trivial equilibrium must have a last on-path history at which the principal experiments for similar reasons to the case without transfers. Transfers do not help prevent the agent from acting on low-quality information at this last history: any payment after a success only exacerbates the problem and the principal can never credibly promise payments after a failure since she stops experimenting and thus her continuation payoff is 0 which in turn implies that she has no reason to honor the promise. Thus,  $\underline{\Pi}(p_0)$  remains an upper bound for the principal’s payoff at such last period histories since one-sided transfers can only lower her continuation utility.

We end this subsection by observing that inefficiency can trivially be eliminated by allowing two-sided transfers. This essentially allows the agent to buy the experimentation technology from the principal. As is the case in many contracting problems, this completely removes any frictions because only the good type will be willing to pay a sufficiently high amount.

#### 4.3. Other Modeling Assumptions

Here, we briefly discuss a few other assumptions that can be relaxed. First, we can allow for the principal and agent to have different discount factors. As should be clear from the preceding

analysis, the agent’s discount factor plays no role. The statement of [Theorem 3](#) remains as is and the only difference is that the key payoffs  $\bar{\Pi}$  and  $\underline{\Pi}$  from the two benchmarks need to be computed using the principal’s discount factor.

A few other assumptions can be relaxed. Our main result is unaffected if we allow the agent to discard or destroy information. In our applications, this amounts to an agent deliberately lowering the likelihood of generating a success, for instance, by discarding well reported content or good research ideas and acting anyway. Formally, we model this by allowing the good-type agent to pick, in response to information  $i_t$ , both an action  $a_t$  and a quality  $q_t$  of the action.  $q_t$  determines the likelihood of a success and is bounded above by the information quality

$$q_t \leq \begin{cases} 1 & \text{if } i_t = i_h, \\ q_l & \text{if } i_t = i_l, \\ 0 & \text{if } i_t = i_\varphi. \end{cases}$$

It is apparent that this generality plays no role as the agent does not have an incentive to increase the likelihood of a failure in equilibrium. This is because failures lead to losses for the principal and do not help in screening (since the bad type can always surely generate failures).

Finally, the timing of moves in our extensive form stage game ([Figure 1](#)) implies that the agent only receives information when the principal experiments. This assumption clearly does not matter because, information being perishable, implies that past information is payoff irrelevant and the agent cannot act on any information unless the principal experiments.

## 5. IMPROVING MARKET FUNCTIONING WITH PARTIAL COMMITMENT

As we showed in the previous section, transfers do not improve market functioning when quality control is necessary. In this section, we argue instead that commitment from the principal to a longer duration of experimentation can ease the agent’s immediate need to develop reputation.

Specifically, we model this as partial commitment: whenever the principal decides to experiment, she commits to experimenting for  $T > 1$  periods. Formally, the principal’s strategy

$$\hat{x}(h_1 \dots h_t) \in \begin{cases} [0, 1] & \text{if } h_{t-T+2} = \dots = h_t = \not{h}, \\ \{1\} & \text{otherwise,} \end{cases}$$

specifies the probability of experimentation at public histories  $h_1 \dots h_t \in \mathcal{H}$ . This probability must be one if she has experimented in any of the past  $T - 1$  periods. Here we use the convention that  $h_{t'} = \not{h}$  if  $t' \leq 0$ . A different way of interpreting this particular form of partial commitment is that when a principal decides to experiment, she sinks the  $T$  period discounted cost of experimentation  $\frac{1-\delta^T}{1-\delta}c$  up front. Note that our refinement applies verbatim since the only restriction on the principal’s strategy is to force her to experiment at histories where she might otherwise not. In addition to being relevant for our applications (which we discuss further at the end of this section), this is one of the reasons we chose to focus on this particular type of partial commitment.

Partial commitment allows the principal to realize benefits from experimentation even when quality control is necessary. By allowing her to sink the cost of experimentation up front, the principal’s decision to experiment is based on her *total* profit over  $T$  periods. To see this consider

the following strategies. The principal commits to experimenting for the first  $T$  periods

$$\hat{x}(h^{t-1}) = 1 \quad \text{for all } t \leq T, h^{t-1} \in \mathcal{H}.$$

If a failure or no success arrives over these  $T$  periods, the principal stops experimenting. If the agent generates at least one success (and no failures) over the first  $T$  periods, the principal experiments forever. Formally, for  $t \geq T + 1$  and  $h_1 \dots h_{t-1} \in \mathcal{H}$ ,

$$\hat{x}(h_1 \dots h_{t-1}) = \begin{cases} 1 & \text{if } h_{t'} \neq \underline{h} \text{ for all } 1 \leq t' \leq T \text{ and } h_{t''} = \bar{h} \text{ for some } 1 \leq t'' \leq T, \\ 0 & \text{otherwise.} \end{cases}$$

In response, the good type agent only acts on high-quality information for the first  $t \leq T - 1$  periods

$$\tilde{a}_{\theta_g}(h^{t-1}, i^{t-1} i_t) = \begin{cases} 1 & \text{if } i_t = i_h, \\ 0 & \text{otherwise,} \end{cases} \quad \text{for all } (h^{t-1}, i^{t-1} i_t) \in \mathcal{H}^g$$

and, additionally, on low-quality information at period  $T$  if no success has arrived,

$$\tilde{a}_{\theta_g}(h^{T-1}, i_1 \dots i_T) = \begin{cases} 1 & \text{if } i_T = i_h, \\ 1 & \text{if } i_T = i_l \text{ and } i_{t'} \neq i_h \text{ for all } 1 \leq t' \leq T - 1, \\ 0 & \text{otherwise,} \end{cases}$$

for all  $(h^{T-1}, i^{T-1} i_T) \in \mathcal{H}^g$ . If a success arrives in the first  $T$  periods, he acts efficiently thereafter. Formally, for  $t \geq T + 1$  and  $(h_1 \dots h_{t-1}, i^{t-1} i_t) \in \mathcal{H}^g$ ,

$$\tilde{a}_{\theta_g}(h_1 \dots h_{t-1}, i^{t-1} i_t) = \begin{cases} 1 & \text{if } i_t = i_h, h_{t'} \neq \underline{h} \text{ for all } 1 \leq t' \leq T \text{ and } h_{t''} = \bar{h} \text{ for some } 1 \leq t'' \leq T, \\ 0 & \text{otherwise.} \end{cases}$$

Finally, the bad type only acts with positive probability  $\varepsilon \in (0, 1)$  in period  $T$  so

$$\tilde{a}_{\theta_b}(h^{T-1}) = \varepsilon \text{ for all } h^{T-1} \in \mathcal{H} \text{ and } \tilde{a}_{\theta_b}(h^{t-1}) = 0 \text{ for all } t \neq T, h^{t-1} \in \mathcal{H}.$$

We now argue that there are parameter values such that the above strategies constitute an equilibrium even when quality control is necessary. First note that experimentation can be profitable. For instance, if  $p_0 \lambda^h > c$ , then the principal's payoff taking  $T \rightarrow \infty$  converges to  $\frac{p_0 \lambda^h - c}{1 - \delta} > 0$  so experimentation is also profitable for some finite  $T$  irrespective of the value of  $\underline{\Pi}(p_0)$  (the optimal such  $T$  is finite).

Clearly, the bad type is best responding since he is indifferent between all strategies.<sup>9</sup> Finally, the good type's strategy is a best response if he does not have an incentive to act on low-quality information at any period prior to  $T$ . A sufficient condition for this is  $\lambda_h \geq q_l$  since the likelihood of getting high quality information (and thereby a sure success) in the next period makes the option value of waiting more attractive than the lottery of generating a success today with probability  $q_l$ .

To summarize, partial commitment to  $T$  periods of experimentation implies that the good type does not need to always act on low-quality information and this in turn can generate a positive average profit for the principal. Indeed, when  $\underline{\Pi}(p_0)$  is negative but small, there are parameter values such that the principal receives a positive expected payoff from experimentation in *every*

<sup>9</sup>Note that the bad type has no incentive to act prior to  $T$  (even though the good type acts with positive probability) because failures between periods 1 and  $T - 1$  are off path and are punished by the principal.

*period* (not just a positive total payoff) gross of the sunk cost. This latter observation is important for our applications because otherwise the principal could choose to “ignore” the agent’s action in period  $T$ . This corresponds to a reader refusing to read any content or a firm refusing to run a drug trial at period  $T$  because the expected payoff (even gross of sunk cost  $c$ ) is negative due to the cost of failure  $\kappa$ .

We specifically chose to highlight this form of partial commitment because it has a natural counterpart in our leading applications. In the case of online content, this is akin to a subscription based payment model. Our analysis suggests that, having been paid a subscription fee, a content provider can be more judicious about his choice of content as he does not need to incentivize each additional click. The same is true for experts in organizations as (even short term) job security allows them to admit that they might not always have actionable information which in turn prevents unnecessary losses to the firm. Interestingly, the above equilibrium strategies in this section mirror employment contracts that are commonly used in academia which consist a single fixed period of employment followed by either dismissal or tenure.

## 6. DISCUSSION

### 6.1. *Storable Information*

In the above analysis, we have assumed that information is perishable; if an agent does not act on information in a given period, he cannot use it in the future. This assumption is suitable for time sensitive information such as breaking news or research in competitive industries but might not be appropriate for other applications. A natural version of the game is one in which the agent can choose to act on any information he has received till date but not acted on in the past. In other words, the agent can store information and, at any point, can act on any piece of information he has previously stored. After the agent acts on a piece of information, it is then spent. Our refinement can be applied verbatim noting that efficient actions for the good type means that he acts only on (potentially previously stored) high quality information.

The ability to store information can improve market functioning. The easiest way to see this is to consider the following two period example. Here, the principal experiments for two periods at which point, she stops experimenting from period three onwards unless the agent generates a success in period two. The good type, in response, does not act in period one but acts on both high and low quality information in period two. The bad type only acts in period two with small probability.

The fact that the agent is required to not act in period one, gives him one additional period to receive high-quality information before acting eventually in period two. At that point, it is thus relatively more likely that the agent has at least one piece of (potentially stored) high-quality information (in which case she will act on it) as opposed to only low-quality information. It is easy to find parameter values (for instance, when the cost of experimentation is small relative to the loss from failure) where  $\underline{\Pi}(p_0) \leq 0$  but the principal is willing to pay for an additional period of experimentation to reduce the relative likelihood of a failure.

## 6.2. Nash Equilibrium

While we feel that our refinement is both natural and weak, it is worth pointing out that the main insights of [Theorem 3](#) do not hold in the full class of Nash equilibria. Recall that, when our refinement is not imposed, there are Nash equilibria in which principal stops experimenting despite it being public knowledge that the agent is the good type. This (seemingly unrealistic) behavior can destroy the benefits from establishing reputation which, as we have argued, is the key driver that leads to both inefficient actions and market breakdown.

**THEOREM 5.** *Suppose  $p_0 \in (0, 1)$ . Then:*

- (1) *There exist nontrivial Nash equilibria in which the good-type agent only acts on high-quality information and the bad-type agent never acts.*<sup>10</sup>
- (2) *As  $p_0 \rightarrow 1$ , there is a sequence of Nash equilibria such that the payoffs of the principal and the agent converge to the Pareto optimal.*

This result is easy to show; for brevity, we do not include proofs as the discussion below makes the formal argument self-evident. Both parts follow from essentially the same intuition. Define  $\hat{T}$  as largest time period  $t$  that satisfies

$$\frac{p_0}{p_0 + \frac{(1-p_0)}{(1-\lambda_h)^{t-1}}} \geq \frac{c}{\lambda_h}. \quad (2)$$

The term on the left is the posterior belief in period  $t$  when the agent has not acted in any prior period and the agent's strategy is to act efficiently. The term on the right is the minimal belief that would make the principal willing to experiment if the stage game were only played for one period and the agent was acting efficiently.

With  $\hat{T}$  thus defined, consider the following strategies. The principal experiments for the first  $\hat{T}$  periods and the agent acts efficiently. At all subsequent periods, the principal never experiments and the agent never acts.<sup>11</sup>

It is easy to see that these strategies constitute a Nash equilibrium whenever  $\hat{T} \geq 1$ . Because the outcomes in the first  $\hat{T}$  periods have no effect on continuation play from period  $\hat{T} + 1$  onwards, both types of the agent are indifferent between all strategies (irrespective of information received) and so acting efficiently is, in particular, a best response. The principal on the other hand has an incentive to experiment for the first  $\hat{T}$  periods because her belief is never below the right side of (2) which in turn implies that her expected payoff from experimenting in each stage game before  $\hat{T}$  is nonnegative (as the agent is acting efficiently). Because the agent switches to never acting after  $\hat{T}$ , it is no longer optimal to experiment even if the principal's belief has jumped to one (which

<sup>10</sup>Formally, there exist nontrivial Nash equilibria such at all on path histories  $h^t \in \mathcal{H}$ ,  $(h^t, i^t i_{t+1}) \in \mathcal{H}^S$  where  $\tilde{x}(h^t) > 0$ , we have  $a_{\theta_g}(h^t, i^t i_{t+1}) = 1(0)$  if  $i_{t+1} = (\neq) i_h$  and  $a_{\theta_b}(h^t) = 0$ .

<sup>11</sup>Formally,  $\tilde{x}, \tilde{a}$  are defined as follows. For all  $0 \leq t \leq \hat{T} - 1$ ,  $h^t \in \mathcal{H}$ ,  $(h^t, i^{t+1}) \in \mathcal{H}^S$ , strategies are

$$\tilde{x}(h^t) = 1, \quad \tilde{a}_{\theta_g}(h^t, i^t i_{t+1}) = \begin{cases} 1 & \text{if } i_{t+1} = i_h, \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad \tilde{a}_{\theta_b}(h^t) = 0.$$

When  $t \geq \hat{T}$ , strategies are

$$\tilde{x}(h^t) = \tilde{a}_{\theta_g}(h^t, i^{t+1}) = \tilde{a}_{\theta_b}(h^t) = 0 \quad \text{for all } h^t \in \mathcal{H}, (h^t, i^{t+1}) \in \mathcal{H}^S.$$

happens if the agent acts at or before  $\hat{T}$ ). Note that it is precisely this particular feature of the equilibrium strategies that is ruled out by our refinement.

The first part of [Theorem 5](#) states that such a nontrivial equilibrium exists. This is true whenever  $p_0\lambda_h > c$  as this implies that  $\hat{T} \geq 1$ . For the second part of the theorem, note that  $\hat{T} \rightarrow \infty$  as  $p_0 \rightarrow 1$  and therefore the principal's payoff from the constructed equilibrium  $\frac{1-\delta^{\hat{T}}}{1-\delta}(p_0\lambda_h - c)$  converges to the first best  $\bar{\Pi}$ .

### 6.3. Full Commitment

In [Section 5](#), we showed that partial commitment (that does not conflict with our refinement) to  $T$  periods of experimentation at a time can lead to Pareto improvements over the outcome of [Theorem 3](#). However, [Theorem 5](#) shows that inefficient actions can be eliminated without any commitment as long as the players are not forced into efficient continuation play once the agent is revealed to be the good type. An implication is that, as uncertainty disappears, there are Nash equilibria in which the principal's payoff approaches the first best. In light of this, it is natural to ask whether having full commitment can lead to even greater profits? There are recent instances of dynamic mechanism design problems (for instance, [Guo \(2016\)](#) and [Deb, Pai, and Said \(2018\)](#)) where the principal does not require commitment to maximize profits. More generally, [Ben-Porath, Dekel, and Lipman \(2018\)](#) have shown that, for a large class of static mechanism design problems with "evidence," the commitment optimal can be achieved even without commitment. They have ongoing work which shows that these insights also extend to certain dynamic environments.

We now allow the principal to announce her strategy  $\tilde{x}$  in advance and commit to it. In response, the agent's strategy  $\tilde{a}_\theta$  is the best response to  $\tilde{x}$  that maximizes the principal's payoff (an assumption that is always made in mechanism design).

**THEOREM 6.** *Full commitment to  $\tilde{x}$  allows the principal to attain a strictly higher payoff than the principal-optimal Nash equilibrium.*

In the appendix, we state the result more formally by providing sufficient conditions under which commitment makes the principal strictly better off. Full commitment allows the principal to experiment at histories where her belief is so low that she would not do so in any Nash equilibrium. This in turn gives the principal one additional instrument: it allows her to provide the agent continuation value at all histories irrespective of her belief. [Theorem 6](#) shows that this is a more profitable way of screening the agent as opposed to foregoing surplus despite knowing that the agent is the good type.

We end this section by noting that commitment to a strategy  $\tilde{x}$  is *not* the same as commitment to a direct revelation mechanism. This is because our model is a setting with both adverse selection and strategic actions. A direct mechanism should solicit the agent's type at the beginning of the game and should provide (possibly mixed) action recommendations after the agent reports his period- $t$  information. These are both features that are absent from the principal's strategy  $\tilde{x}$  in the repeated game. A full characterization of the optimal mechanism is hard and we have only been

able to derive it in special cases. This difficulty is to be expected. In order to solve for the optimal mechanism, we have to first derive the most profitable way to implement a given incentive compatible action strategy for the agent and then optimize over all such possible action strategies. Even the first of these two steps is hard as mechanism design problems without transfers can be complex even in static pure adverse selection environments without strategic actions.

## 7. CONCLUDING REMARKS

With the aim of studying markets for expertise (broadly defined), we develop and analyze a novel repeated game. A principal screens an agent with a privately known type who periodically receives private information (the arrival rate of which depends on his type) and strategically decides whether or not to act on it. Actions generate public outcomes which depend on the quality of the information received (if any). A key feature of our model is that, while the good-type agent can provide the principal with a positive average payoff, he cannot generate positive payoffs for the principal in every period. We show that reputational incentives can be exceedingly strong in this environment: a consequence is inefficiency in every equilibrium (subject to a mild refinement) that can lead to market breakdown leaving potentially large surplus gains on the table. We argue that this breakdown can occur discontinuously as we change model parameters and thus small changes in the environment can have large effects on market functioning. We show that this central insight is robust to several generalizations of the model; the most significant of which is allowing transfers from the principal to the agent. We suggest partial commitment as a way to improve market functioning. Our results speak, in particular, to the effect of advertising based revenue (as opposed to traditional subscription based payment) on the quality of online content and to the improvement of contract design, with the aim of inducing better decision making by career concerned experts, in organizations.

We end by discussing some limitations of our analysis and the natural avenues for future research they suggest. The key assumptions driving our main insights are (i) that both information acquisition and actions are costless for the agent and (ii) the monitoring structure that determines how information quality affects the distribution of outcomes via actions. We address each in turn. While costless actions are a natural assumption for the applications we have focused on (there is no physical cost of publishing a story on a website, drug trials are paid for by the pharmaceutical company), there are other environments where this assumption is less than ideal. Costly actions significantly complicate the analysis because the agent can then signal by bearing the cost of acting. Similarly, if information acquisition is costly, the efficacy with which the principal can screen is determined by her ability to provide the agent appropriate incentives to acquire information.

There are two separate aspects of the monitoring structure that are worth discussing. First, in our model, bad types can only generate failures and thus good types can perfectly out themselves by producing a success. As we have argued, this assumption makes our results more stark because one might expect that this facilitates screening by allowing the good type to perfectly separate. That said, a natural alternative to this assumption is to allow bad types to also generate successes with positive probability by acting on no information. Conceptually, as long as we could ensure that the good type always acts on high-quality information, our main results should continue to hold. At a high level, this is because, the good type is more likely to generate successes and beliefs

will then drift down when successes do not arrive leading to the problematic last histories. The technical issue of moving away from a pure good news setting is that the principal's belief would never go to one on path and thus our refinement would have no bite whatsoever. As we have argued, unrealistic equilibria can exist absent a refinement. A stronger alternative refinement to ours would be to impose some version of renegotiation-proofness at all histories. This would introduce technical difficulties and, moreover, to the best of our knowledge, there is no natural renegotiation-proofness refinement that we can borrow from the literature.

Finally, there are applications where our assumption of imperfect public monitoring may be inappropriate. Instead, in such environments, the principal could privately observe the outcome which is an imperfect signal (à la [Compte \(1998\)](#)) of the agent's information. Here, the principal can communicate with the agent by sending him cheap talk messages. An example of such a setting is a firm hiring a consultant to provide advice whose outcome is privately observed by the firm which then provides feedback. These and numerous other variants of our model are not mere extensions as they involve different technical challenges and speak to different real-world problems. We hope to address at least some of these questions in future work.

APPENDIX A. PROOFS OF THEOREM 3 AND THEOREM 4

We will prove [Theorem 4](#) for the game with transfers and then argue that it implies [Theorem 3](#). Before we proceed with the proofs, we want to formally define histories and strategies for the game with transfers which we chose to omit from [Section 4.2](#).

*Transfers:* After the public outcome is realized in period  $t$ , the principal makes a one-sided transfer  $\tau_t \in \mathbb{R}_+$  to the agent.

*Histories:* A (public) history  $(h_1 \dots h_{t-1}, \tau_1 \dots \tau_{t-1})$  at the beginning of period  $t \geq 2$  satisfies  $h_1 \dots h_{t-1} \in \mathcal{H}$  and  $\tau_{t'} \geq (=)0$  when  $h_{t'} \neq (=)\not{h}$  for all  $1 \leq t' \leq t$ . In addition to the previous actions and outcomes, it also contains the previous transfers (that are zero whenever the principal does not experiment).  $(h^0, \tau^0)$  denotes the beginning of the game. We denote the set of histories using  $\mathcal{T}$ .

As before, the good type agent's period- $t$  private history  $(h^{t-1}, \tau^{t-1}, i^t)$  additionally contains the current and previous information realizations  $i^t$ . We use  $\mathcal{T}^g$  to denote the set of the good type of agent's private histories.

We use  $\tilde{\mathcal{T}}^t$  where  $t \geq 1$  to denote the set of on-path histories at the beginning of period  $t + 1$ . Here, the notation reflects the fact that this set depends on the equilibrium strategies  $(\tilde{x}, \tilde{\tau}, \tilde{a}_\theta)$ .

*Agent's Strategy:* The bad type's strategy  $\tilde{a}_{\theta_b}(h^{t-1}, \tau^{t-1}) \in [0, 1]$  specifies the probability of acting at each period  $t$  public history  $(h^{t-1}, \tau^{t-1}) \in \mathcal{T}$ . Similarly,  $\tilde{a}_{\theta_g}(h^{t-1}, \tau^{t-1}, i^t) \in [0, 1]$  for each private history  $(h^{t-1}, \tau^{t-1}, i^t) \in \mathcal{T}^g$ .

*Principal's Strategy:* The principal's strategy consists of two functions: an experimentation decision  $\tilde{x}(h^{t-1}, \tau^{t-1}) \in [0, 1]$  and a transfer strategy  $\tilde{\tau}(h^{t-1}h_t, \tau^{t-1}) \in \Delta(\mathbb{R}_+)$  which specifies the distribution over transfers for each  $(h^{t-1}, \tau^{t-1}) \in \mathcal{T}$ ,  $h_t \in \{\bar{h}, \underline{h}, h_\varphi\}$ . Note that, by definition,  $\tilde{\tau}(h^{t-1}h_t, \tau^{t-1})$  is the Dirac measure at 0 when  $h_t = \not{h}$ .

*Beliefs:* Finally, the principal's beliefs  $\tilde{p}(h^{t-1}, \tau^{t-1})$ ,  $\tilde{p}(h^{t-1}h_t, \tau^{t-1})$  at the moment she makes her experimentation and transfer decision respectively also depend on the past history of transfers. Once again, these beliefs are derived by Bayes' rule on path and are not restricted off path.

*Expected Payoffs:* Expected payoffs now additionally also have as an argument the history of transfers and are denoted by  $U_{\theta_g}(h^{t-1}, \tau^{t-1}, i^t)$ ,  $U_{\theta_b}(h^{t-1}, \tau^{t-1})$ ,  $V(h^{t-1}, \tau^{t-1})$  to denote the expected payoff of the good, bad type agent and principal respectively. We sometimes also refer to the agent's expected payoff  $U_{\theta_g}(h^t, \tau^{t-1}, i^t)$ ,  $U_{\theta_b}(h^t, \tau^{t-1})$  after the outcome but before the principal's transfer in period- $t$  is realized. Note that, for brevity, we are suppressing dependence on strategies.

We now define "last histories." The necessity of such histories arising on-path will be crucial for our arguments.

*Last History:* A last history is an on-path history at which the principal experiments such that, if no success is generated, the principal stops experimenting (almost surely) at all future histories. Formally, a last history is an on-path  $(h^t, \tau^t) \in \tilde{\mathcal{T}}^t$  such that  $\tilde{x}(h^t, \tau^t) > 0$ ,  $\tilde{p}(h^t, \tau^t) < 1$  and  $U_{\theta_b}(h^t h_{t+1}, \tau^t, \tilde{x}, \tilde{a}_{\theta_b}) = 0$  for all on-path  $h_{t+1} \in \{\underline{h}, h_\varphi\}$ . Note that this also implies  $U_{\theta_g}(h^t h_{t+1}, \tau^t, i^{t+1}) = 0$  for  $h_{t+1} \in \{\underline{h}, h_\varphi\}$ .

Note that we define a last history in terms of the agent's payoff (as opposed to the principal's experimentation decision) so that we can avoid qualifiers about the principal's actions on continuation on-path histories that are reached with probability 0 (after the principal's transfer is realized via her mixed strategy). Also observe that  $U_{\theta_b}(h^t h_{t+1}, \tau^t \tau_{t+1}) = 0$  implies that  $\tau_{t+1} = 0$  since it cannot be a best response for the principal to make a positive transfer to the agent in period  $t + 1$  when her continuation value is 0.

We now argue that every nontrivial Nash equilibrium must have such a last history.

**LEMMA 1.** *Suppose  $p_0 \in (0, 1)$ . Every nontrivial Nash equilibrium  $(\tilde{x}, \tilde{\tau}, \tilde{a}_\theta)$  of the game with transfers must have a last history.*

**PROOF.** We first define,

$$\underline{p} = \inf \left\{ \tilde{p}(\hat{h}^t, \hat{\tau}^t) \mid (\hat{h}^t, \hat{\tau}^t) \in \mathcal{F}^t, t' \geq 0 \right\} < 1,$$

to denote the infimum of the beliefs at on-path histories. We will now assume to the converse that there is no last history. With this assumption in place, the following steps yield the requisite contradiction.

*Step 1: Experimentation at low beliefs.* For all  $\varepsilon > 0$ , there exists an on-path history  $(h^t, \tau^t) \in \mathcal{F}^t$  such that the principal experiments  $\tilde{x}(h^t, \tau^t) > 0$  and the belief is close to the infimum  $\tilde{p}(h^t, \tau^t) < 1$ ,  $\tilde{p}(h^t, \tau^t) - \underline{p} < \varepsilon$ .

By definition, there exists a history  $(\hat{h}^t, \hat{\tau}^t) \in \mathcal{F}^t$  such that the belief satisfies  $\tilde{p}(\hat{h}^t, \hat{\tau}^t) < 1$  and  $\tilde{p}(\hat{h}^t, \hat{\tau}^t) - \underline{p} < \varepsilon$ . It remains to be shown that there is a history at which the principal experiments. We consider the case where  $\tilde{x}(\hat{h}^t, \hat{\tau}^t) = 0$  as otherwise  $(h^t, \tau^t) = (\hat{h}^t, \hat{\tau}^t)$  would be the required history.

If  $U_{\theta_b}(\hat{h}^t, \hat{\tau}^t) > 0$ , then this implies that there is an on-path continuation history  $(\hat{h}^t \hat{h}^s, \hat{\tau}^t \hat{\tau}^s) \in \mathcal{F}$ ,  $\hat{h}^s = (\mathcal{H}, \dots, \mathcal{H})$ ,  $\hat{\tau}^s = (0, \dots, 0)$ ,  $s \geq 1$  where the principal eventually experiments  $\tilde{x}(\hat{h}^t \hat{h}^s, \hat{\tau}^t \hat{\tau}^s) > 0$ ; this is because the principal is not allowed to make transfers unless she experiments first. But since the belief does not change without experimentation, this implies that  $\tilde{p}(\hat{h}^t \hat{h}^s, \hat{\tau}^t \hat{\tau}^s) = \tilde{p}(\hat{h}^t, \hat{\tau}^t)$  and thus  $(h^t, \tau^t) = (\hat{h}^t \hat{h}^s, \hat{\tau}^t \hat{\tau}^s)$  is the requisite history.

So suppose instead that  $U_{\theta_b}(\hat{h}^t, \hat{\tau}^t, \tilde{x}, \tilde{a}_{\theta_b}) = 0$ . Let  $0 \leq s < t$  be the last period in the history  $(\hat{h}^t, \hat{\tau}^t)$  at which the principal experiments. Formally,  $\tilde{x}(\hat{h}^s, \hat{\tau}^s) > 0$  and  $(\hat{h}^t, \hat{\tau}^t) = (\hat{h}^s h_{s+1} \hat{h}^{t-s-1}, \hat{\tau}^s \hat{\tau}^{t-s})$  where  $\hat{h}^{t-s-1} = (\mathcal{H}, \dots, \mathcal{H})$  when  $s < t - 1$  and  $\hat{\tau}^{t-s} = (0, \dots, 0)$ . Such a history  $(\hat{h}^s, \hat{\tau}^s)$  must exist because the equilibrium is nontrivial. If principal's realized choice was not to experiment at period  $s + 1$ , then we have  $h_{s+1} = \mathcal{H}$  which in turn implies  $\tilde{p}(\hat{h}^s, \hat{\tau}^s) = \tilde{p}(\hat{h}^t, \hat{\tau}^t)$  and  $(h^t, \tau^t) = (\hat{h}^s, \hat{\tau}^s)$  is the requisite history.

So finally suppose that the realized action of the principal was to experiment at  $s + 1$ . Then we must have  $\hat{h}_{s+1} \in \{\underline{h}, h_\varphi\}$  (note that if  $\hat{h}_{s+1} = \bar{h}$ , this would imply  $\tilde{p}(\hat{h}^t, \hat{\tau}^t) = 1$  which is a contradiction). If  $U_{\theta_b}(\hat{h}^s \hat{h}_{s+1}, \hat{\tau}^s) > 0$ , this would imply the existence of a first continuation history  $(\hat{h}^s \hat{h}_{s+1} \check{h}^{s'}, \hat{\tau}^s \check{\tau}^{s'+1}) \in \mathcal{F}^{s+s'+1}$  such that  $\tilde{x}(\hat{h}^s \hat{h}_{s+1} \check{h}^{s'}, \hat{\tau}^s \check{\tau}^{s'+1}) > 0$  and  $\tilde{p}(\hat{h}^s \hat{h}_{s+1} \check{h}^{s'}, \hat{\tau}^s \check{\tau}^{s'+1}) = \tilde{p}(\hat{h}^s \hat{h}_{s+1}, \hat{\tau}^s) = \tilde{p}(\hat{h}^t, \hat{\tau}^t)$  and so  $(h^t, \tau^t) = (\hat{h}^s \hat{h}_{s+1} \check{h}^{s'}, \hat{\tau}^s \check{\tau}^{s'+1})$  would be the requisite history.

Therefore the only remaining case to analyze is when  $U_{\theta_b}(\hat{h}^s \hat{h}_{s+1}, \hat{\tau}^s) = 0$ . Since we have assumed that there is no last history, it must be the case that the principal experiments after the other non-success outcome or that  $U_{\theta_b}(\hat{h}^s \check{h}_{s+1}, \hat{\tau}^s) > 0$  for  $\check{h}_{s+1} \in \{\underline{h}, \hat{h}_\varphi\}$ ,  $\check{h}_{s+1} \neq \hat{h}_{s+1}$ . But then it cannot be a best response for type  $\theta_b$  to reach history  $(\hat{h}^s \hat{h}_{s+1}, \hat{\tau}^s)$  since he can always reach history  $(\hat{h}^s \check{h}_{s+1}, \hat{\tau}^s)$  costlessly with probability 1. This implies that  $\tilde{p}(\hat{h}^s \hat{h}_{s+1}, \hat{\tau}^s) = 1$  (since this history is on path) which once again yields the contradiction  $\tilde{p}(\hat{h}^t, \hat{\tau}^t) = 1$ . Thus  $\hat{h}_{s+1} \in \{\underline{h}, \hat{h}_\varphi\}$  is not possible and the proof of this step is complete.

*Step 2: Upper bound for the principal's payoff.* Fix an  $\varepsilon > 0$  and a history  $(h^t, \tau^t) \in \mathcal{F}^t$  such that the principal experiments  $\tilde{x}(h^t, \tau^t) > 0$  and the belief satisfies  $\tilde{p}(h^t, \tau^t) < 1$ ,  $\tilde{p}(h^t, \tau^t) - \underline{p} < \varepsilon$ . For any integer  $s \geq 1$ ,

$$\frac{\varepsilon}{1 - \underline{p}} s - c + \delta^s \bar{\Pi} \geq V(h^t, \tau^t) \quad (3)$$

is an upper bound for the principal's payoff at history  $(h^t, \tau^t)$ .

For the remainder of this step, we assume that the principal's realized action choice is to experiment at  $(h^t, \tau^t)$ . Since  $\tilde{x}(h^t, \tau^t) > 0$ , experimenting at  $(h^t, \tau^t)$  must be a (weak) best response and so it suffices to compute an upper bound for the principal's payoff assuming she experiments for sure at  $(h^t, \tau^t)$ .

We use  $q$  to denote the probability that at least one success arrives in the periods between  $t + 1$  and  $t + s$ . Now consider the principal's beliefs at on-path histories  $(h^t h^s, \tau^t \tau^s) \in \mathcal{F}^{t+s}$ . Since the beliefs follow a martingale, they must average to the belief  $\tilde{p}(h^t, \tau^t)$ . This immediately yields an upper bound for  $q$  since

$$q + (1 - q)\underline{p} \leq \tilde{p}(h^t, \tau^t) \implies q \leq \frac{\tilde{p}(h^t, \tau^t) - \underline{p}}{1 - \underline{p}} \leq \frac{\varepsilon}{1 - \underline{p}}.$$

Since the belief jumps to 1 after a success, the maximal probability of observing a success (subject to Bayes' consistency) can be obtained by assuming that the belief is the lowest possible ( $\underline{p}$ ) when a success does not arrive (which happens with probability  $1 - q$ ).

We can write the principal's payoff  $V(h^t, \tau^t)$  by summing three separate terms: (i) the expected payoff from the outcomes in the  $s$  periods following  $t$ , (ii) the expected cost of experimentation and transfers in the  $s$  periods and (iii) the expected continuation value at  $t + s + 1$ .

To derive the upper bound for the principal's payoff from (3), we label each of the terms of

$$\underbrace{\frac{\varepsilon}{1 - \underline{p}} s}_{(i)} - \underbrace{c}_{(ii)} + \underbrace{\delta^s \bar{\Pi}}_{(iii)},$$

so that they individually correspond to a bound for each component (i)-(iii) of the payoffs.

- (i)  $qs$  is an upper bound for the expected payoff that the principal can receive from outcomes in the  $s$  periods following history  $(h^t, \tau^t)$ . This corresponds to getting a success in each period of the  $s$  periods (with no loss from discounting) whenever at least one success arrives (which occurs with probability  $q \leq \varepsilon / (1 - \underline{p})$ ) and no losses from failures.

- (ii) Since the principal experiments for sure at  $(h^t, \tau^t)$  her expected cost of experimentation must be greater than  $c$ . Additionally, her cost from transfers must be at weakly greater than 0.
- (iii) The principal's expected continuation value must be less than  $\bar{\Pi}$  since this is the first-best payoff corresponding to the case where the agent is known to be the good type for sure.

*Step 3: Final contradiction.* By simultaneously taking  $s$  large and  $\varepsilon s$  small, we can find a history  $(h^t, \tau^t) \in \tilde{\mathcal{T}}^t$  where the principal experiments  $\tilde{x}(h^t, \tau^t) > 0$  but the maximal payoff she can get (given by the bound (3)) is negative.

Since this cannot be true in any Nash equilibrium, this contradicts the assumption that there is no last history and completes the proof of the lemma. ■

**PROOF OF THEOREM 4, PART (1).** Suppose  $(\tilde{x}, \tilde{\tau}, \tilde{a}_\theta)$  is a nontrivial equilibrium. Then Lemma 1 shows that there must be a last history. At this history, it is a strict best response  $\tilde{a}_{\theta_g}(h^t, \tau^t, i^t_{i_l}) = 1$  for type  $\theta_g$  to act on low-quality information since our refinement implies that

$$q_l U_{\theta_g}(h^t \bar{h}, \tau^t) + (1 - q_l) U_{\theta_g}(h^t \underline{h}, \tau^t) \geq \frac{\delta}{1 - \delta} > 0 = U_{\theta_g}(h^t h_\varphi, \tau^t).$$

This additionally implies that failure at this history  $(h^t \underline{h}, \tau^t)$  must be on path. A final consequence is that we must have  $\tilde{a}_{\theta_b}(h^t, \tau^t) > 0$  as otherwise  $\tilde{p}(h^t \underline{h}, \tau^t) = 1$  which, due to the refinement, would contradict the fact that  $(h^t, \tau^t)$  is a last history. This proves the first part of the theorem. ■

We now prove the second part of Theorem 4.

**PROOF OF THEOREM 4, PART (2).** Suppose  $(\tilde{x}, \tilde{\tau}, \tilde{a}_\theta)$  is a nontrivial equilibrium. Then Lemma 1 implies that there must be a last history. Since the principal's expected payoff at any on-path history must be nonnegative, this implies that her beliefs at *all* last histories must be strictly greater than  $p_0$  (since  $\underline{\Pi}(p_0) \leq 0$ ). We will show this is not possible via the following sequence of steps.

We first define  $\tilde{\pi}(\mathcal{T} | h^t, \tau^t)$  which denotes the probability of reaching the set of histories  $\mathcal{T} \subset \mathcal{T}$  starting at history  $(h^t, \tau^t) \in \mathcal{T}$  via equilibrium strategies  $(\tilde{x}, \tilde{\tau}, \tilde{a}_\theta)$ .

*Step 1: Partitioning the histories.* We can partition the set of on-path period- $t + 1$  histories  $\tilde{\mathcal{T}}^t$  into three mutually disjoint sets:

- (1) The set  $\tilde{\mathcal{L}}^t \subset \tilde{\mathcal{T}}^t$  of all histories that follow from a last history: this consists of all histories  $(h^s h^{t-s}, \tau^s \tau^{t-s}) \in \tilde{\mathcal{T}}^t$  such that  $(h^s, \tau^s)$  is a last history for some  $1 \leq s \leq t$ .
- (2) The set  $\tilde{\mathcal{S}}^t \subset \tilde{\mathcal{T}}^t$  of all histories that follow from a success being generated in a non-last history: this consists of all histories  $(h^t, \tau^t) \in \tilde{\mathcal{T}}^t$  such that  $h_s = \bar{h}$  for some  $1 \leq s \leq t$ ,  $h_{s'} \neq \bar{h}$  for all  $1 \leq s' < s$  and  $(h^{s-1}, \tau^{s-1}) \notin \tilde{\mathcal{L}}^{s-1}$  is not a last history.
- (3) All other remaining histories  $\tilde{\mathcal{H}}^t = \tilde{\mathcal{T}}^t \setminus (\tilde{\mathcal{S}}^t \cup \tilde{\mathcal{L}}^t)$ . Note that for all  $(h^t, \tau^t) \in \tilde{\mathcal{H}}^t$ , we must have  $h_s \neq \bar{h}$  for all  $1 \leq s \leq t$ .

*Step 2: The lowest belief amongst histories in the set  $\tilde{\mathcal{H}}^t$  is less than  $p_0$ .* Formally, for all  $t$ ,  $\tilde{\mathcal{H}}^t$  is nonempty and there is a history  $(h^t, \tau^t) \in \tilde{\mathcal{H}}^t$  such that  $\tilde{p}(h^t, \tau^t) \leq p_0$ .

Since beliefs follow a martingale, the expected value of the beliefs at the beginning of period  $t + 1$  must equal the prior at the beginning of the game or

$$p_0 = \mathbb{E}[\tilde{p}(h^t, \tau^t)].$$

Note that the expectation above and those that follow in the proof of this step are taken with respect to the distribution  $\tilde{\pi}(\cdot | h_0, \tau_0)$  over period  $t + 1$  histories induced by the equilibrium strategies. We can rewrite this expression as

$$\begin{aligned} p_0 &= \tilde{\pi}(\mathcal{S}^t | h^0, \tau^0) \mathbb{E}[\tilde{p}(h^t, \tau^t) | (h^t, \tau^t) \in \mathcal{S}^t] + \tilde{\pi}(\mathcal{L}^t | h^0, \tau^0) \mathbb{E}[\tilde{p}(h^t, \tau^t) | (h^t, \tau^t) \in \mathcal{L}^t] \\ &\quad + \tilde{\pi}(\tilde{\mathcal{H}}^t | h^0, \tau^0) \mathbb{E}[\tilde{p}(h^t, \tau^t) | (h^t, \tau^t) \in \tilde{\mathcal{H}}^t]. \end{aligned}$$

Since  $\tilde{p}(h^t, \tau^t) = 1$  for all  $(h^t, \tau^t) \in \mathcal{S}^t$ , the above equation becomes

$$\begin{aligned} p_0 &= \tilde{\pi}(\mathcal{S}^t | h^0, \tau^0) + \tilde{\pi}(\mathcal{L}^t | h^0, \tau^0) \mathbb{E}[\tilde{p}(h^t, \tau^t) | (h^t, \tau^t) \in \mathcal{L}^t] \\ &\quad + \tilde{\pi}(\tilde{\mathcal{H}}^t | h^0, \tau^0) \mathbb{E}[\tilde{p}(h^t, \tau^t) | (h^t, \tau^t) \in \tilde{\mathcal{H}}^t]. \end{aligned} \tag{4}$$

Now recall that the beliefs at all last histories must be strictly greater than  $p_0$ . This implies that  $\mathbb{E}[\tilde{p}(h^t, \tau^t) | (h^t, \tau^t) \in \mathcal{L}^t] > p_0$  whenever  $\tilde{\pi}(\mathcal{L}^t | h^0, \tau^0) > 0$  since these are the histories that follow last histories and beliefs follow a martingale. Of course, this in turn implies that  $\tilde{\pi}(\tilde{\mathcal{H}}^t | h^0, \tau^0) > 0$  and that  $\mathbb{E}[\tilde{p}(h^t, \tau^t) | (h^t, \tau^t) \in \tilde{\mathcal{H}}^t] \leq p_0$ , which shows that  $\tilde{\mathcal{H}}^t$  is nonempty and that there must be a history  $(h^t, \tau^t) \in \tilde{\mathcal{H}}^t$  such that  $\tilde{p}(h^t, \tau^t) \leq p_0$ .

*Step 3: The principal experiments following a history in  $\tilde{\mathcal{H}}^t$  where the belief is low.* Formally, for all  $t$  and  $\varepsilon > 0$ , there is a history  $(h^t, \tau^t) \in \tilde{\mathcal{H}}^t$  such that

$$\tilde{p}(h^t, \tau^t) \leq \inf_{(\hat{h}^t, \hat{\tau}^t) \in \tilde{\mathcal{H}}^t} \tilde{p}(\hat{h}^t, \hat{\tau}^t) + \varepsilon, \quad \tilde{p}(h^t, \tau^t) \leq p_0 \quad \text{and} \quad \tilde{x}(h^t h^{t'}, \tau^t \tau^{t'}) > 0,$$

for some  $(h^t h^{t'}, \tau^t \tau^{t'}) \in \mathcal{S}^{t+t'}$ ,  $t' \geq 0$ .

First suppose  $\inf_{(\hat{h}^t, \hat{\tau}^t) \in \tilde{\mathcal{H}}^t} \tilde{p}(\hat{h}^t, \hat{\tau}^t) = p_0$ . Then, for all histories  $(\hat{h}^s, \hat{\tau}^s) \in \mathcal{S}^s$ ,  $0 \leq s \leq t$ , we must have  $\tilde{p}(\hat{h}^s, \hat{\tau}^s) = p_0$  which implies  $(\hat{h}^s, \hat{\tau}^s) \in \mathcal{R}^s$ . Suppose this were not true. Take an earliest history  $(\hat{h}^s \hat{h}_{s+1}^s, \hat{\tau}^s \hat{\tau}_{s+1}^s) \in \mathcal{S}^{s+1}$  (with the smallest  $s < t$ ) such that  $\tilde{p}(\hat{h}^s \hat{h}_{s+1}^s, \hat{\tau}^s \hat{\tau}_{s+1}^s) \neq p_0$ . Then, by definition, we must have  $\tilde{p}(\hat{h}^s, \hat{\tau}^s) = p_0$  and  $(\hat{h}^s, \hat{\tau}^s) \in \mathcal{R}^s$  since  $(\hat{h}^s, \hat{\tau}^s) \notin \mathcal{L}^s$  because the beliefs at all last histories are strictly greater than  $p_0$ . Then, by Bayes' consistency, there must be a continuation history  $(\hat{h}^s \check{h}_{s+1}^s, \hat{\tau}^s \check{\tau}_{s+1}^s) \in \mathcal{R}^{s+1}$  such that  $\tilde{p}(\hat{h}^s \check{h}_{s+1}^s, \hat{\tau}^s \check{\tau}_{s+1}^s) < p_0$ .<sup>12</sup> But then we can use the argument in Step 2 starting at  $(\hat{h}^s \check{h}_{s+1}^s, \hat{\tau}^s \check{\tau}_{s+1}^s)$  and belief  $\tilde{p}(\hat{h}^s \check{h}_{s+1}^s, \hat{\tau}^s \check{\tau}_{s+1}^s)$  (instead of the beginning of the game  $(h^0, \tau^0)$  with prior  $p_0$ ) to show that there exists a history  $(\hat{h}^s \check{h}^{t-s}, \hat{\tau}^s \check{\tau}^{t-s}) \in \mathcal{R}^t$  and  $\tilde{p}(\hat{h}^s \check{h}^{t-s}, \hat{\tau}^s \check{\tau}^{t-s}) \leq \tilde{p}(\hat{h}^s \check{h}_{s+1}^s, \hat{\tau}^s \check{\tau}_{s+1}^s) < p_0$  which would be a contradiction.

Now note that, if the principal does not experiment after all period  $t$  histories  $(\hat{h}^t, \hat{\tau}^t) \in \tilde{\mathcal{H}}^t = \mathcal{S}^t$ , it cannot be a best response for her to experiment at any history before period  $t + 1$  either because no successes are ever generated on path. This contradicts the fact that the equilibrium is nontrivial.

<sup>12</sup>Clearly, if  $\tilde{p}(\hat{h}^s \hat{h}_{s+1}^s, \hat{\tau}^s \hat{\tau}_{s+1}^s) < p_0$ , then  $(\hat{h}^s \check{h}_{s+1}^s, \hat{\tau}^s \check{\tau}_{s+1}^s) = (\hat{h}^s \hat{h}_{s+1}^s, \hat{\tau}^s \hat{\tau}_{s+1}^s)$  is such a history.

Next suppose  $\inf_{(\hat{h}^t, \hat{\tau}^t) \in \mathcal{R}^t} \tilde{p}(\hat{h}^t, \hat{\tau}^t) < p_0$ . We define

$$\bar{s} = \max \left\{ s \leq t \mid \begin{array}{l} (\check{h}^{s-1} \check{h}_s \check{h}^{t-s}, \check{\tau}^t) \in \mathcal{R}^t, \check{h}_s \in \{\underline{h}, h_\varphi\}, \tilde{p}(\check{h}^{s-1} \check{h}_s \check{h}^{t-s}, \check{\tau}^t) \leq p_0 \text{ and} \\ \tilde{p}(\check{h}^{s-1} \check{h}_s \check{h}^{t-s}, \check{\tau}^t) \leq \inf_{(\hat{h}^t, \hat{\tau}^t) \in \mathcal{R}^t} \tilde{p}(\hat{h}^t, \hat{\tau}^t) + \varepsilon \end{array} \right\},$$

to be the last period amongst the low belief histories where the principal observes a non-success outcome after experimenting. Let the set of period- $t+1$  histories that yield the above maximum be  $\mathcal{R}_{\bar{s}}^t$ . Note that, by definition,  $\check{h}^{t-\bar{s}} = (\check{h}, \dots, \check{h})$  for all  $(\check{h}^{\bar{s}} \check{h}^{t-\bar{s}}, \check{\tau}^t) \in \mathcal{R}_{\bar{s}}^t$ . Observe that the maximum is well defined because  $\mathcal{R}^t$  is nonempty (from Step 2) and the principal experiments before  $t$  (otherwise the infimum of the beliefs cannot be strictly lower than  $p_0$ ).

We now show that, for any history  $(\check{h}^t, \check{\tau}^t) \in \mathcal{R}_{\bar{s}}^t$ , we must have  $\tilde{x}(\check{h}^{\bar{s}} \hat{h}^{s-\bar{s}}, \check{\tau}^{\bar{s}-1} \hat{\tau}^{s-\bar{s}+1}) = 0$  for all  $\bar{s} \leq s < t$ ,  $(\check{h}^{\bar{s}} \hat{h}^{s-\bar{s}}, \check{\tau}^{\bar{s}-1} \hat{\tau}^{s-\bar{s}+1}) \in \mathcal{T}^s$ . In words, this says that the principal does not experiment between periods  $\bar{s}+1$  and  $t$  at *any* on-path continuation history following  $(\check{h}^{\bar{s}}, \check{\tau}^{\bar{s}-1})$ . A consequence of this statement is that the principal's belief does not change until period  $t+1$  and, since the beliefs at all last histories are strictly greater than  $p_0$ ,  $(\check{h}^{\bar{s}} \hat{h}^{t-\bar{s}}, \check{\tau}^{\bar{s}-1} \hat{\tau}^{t-\bar{s}+1}) \in \mathcal{T}^t$  implies  $(\check{h}^{\bar{s}} \hat{h}^{t-\bar{s}}, \check{\tau}^{\bar{s}-1} \hat{\tau}^{t-\bar{s}+1}) \in \mathcal{R}_{\bar{s}}^t$ .

Suppose this were not the case, and consider the smallest  $\bar{s} \leq s < t$  such that there is an on-path history  $(\check{h}^{\bar{s}} \hat{h}^{s-\bar{s}}, \check{\tau}^{\bar{s}-1} \hat{\tau}^{s-\bar{s}+1}) \in \mathcal{T}^s$  where the principal experiments  $\tilde{x}(\check{h}^{\bar{s}} \hat{h}^{s-\bar{s}}, \check{\tau}^{\bar{s}-1} \hat{\tau}^{s-\bar{s}+1}) > 0$ . Then, there must be an on-path continuation history  $(\check{h}^{\bar{s}} \hat{h}^{s-\bar{s}+1}, \check{\tau}^{\bar{s}-1} \hat{\tau}^{s-\bar{s}+2}) \in \mathcal{T}^{s+1}$ ,  $\hat{h}_{s-\bar{s}+1} \in \{\underline{h}, h_\varphi\}$  such that

$$\tilde{p}(\check{h}^{\bar{s}} \hat{h}^{s-\bar{s}+1}, \check{\tau}^{\bar{s}-1} \hat{\tau}^{s-\bar{s}+2}) \leq \tilde{p}(\check{h}^{\bar{s}} \hat{h}^{s-\bar{s}}, \check{\tau}^{\bar{s}-1} \hat{\tau}^{s-\bar{s}+1}) = \tilde{p}(\check{h}^{\bar{s}}, \check{\tau}^{\bar{s}-1} \hat{\tau}_s) = \tilde{p}(\check{h}^{\bar{s}}, \check{\tau}^{\bar{s}}) \leq p_0.$$

The first inequality follows from Bayes' consistency and the equalities follow from the facts that there is no experimentation from periods  $\bar{s}+1$  to  $s$  and the transfer in period  $\bar{s}$  does not change the belief at period  $\bar{s}+1$ . This combined with the fact that  $(\check{h}^{\bar{s}-1}, \check{\tau}^{\bar{s}-1}) \in \mathcal{R}_{\bar{s}-1}^{\bar{s}-1}$ , implies that  $(\check{h}^{\bar{s}} \hat{h}^{s-\bar{s}+1}, \check{\tau}^{\bar{s}-1} \hat{\tau}^{s-\bar{s}+2}) \in \mathcal{R}^{s+1}$  since the beliefs at all last histories are strictly greater than  $p_0$ . But then, we can once again use the argument in Step 2 starting at the history  $(\check{h}^{\bar{s}} \hat{h}^{s-\bar{s}+1}, \check{\tau}^{\bar{s}-1} \hat{\tau}^{s-\bar{s}+2})$  with associated belief  $\tilde{p}(\check{h}^{\bar{s}} \hat{h}^{s-\bar{s}+1}, \check{\tau}^{\bar{s}-1} \hat{\tau}^{s-\bar{s}+2})$  (instead of the beginning of the game  $(h^0, \tau^0)$  and belief  $p_0$ ) to conclude that there must be a period- $t+1$  history  $(\check{h}^{\bar{s}} \hat{h}^{t-\bar{s}}, \check{\tau}^{\bar{s}-1} \hat{\tau}^{t-\bar{s}+1}) \in \mathcal{R}^t$  satisfying

$$\tilde{p}(\check{h}^{\bar{s}} \hat{h}^{t-\bar{s}}, \check{\tau}^{\bar{s}-1} \hat{\tau}^{t-\bar{s}+1}) \leq \tilde{p}(\check{h}^{\bar{s}} \hat{h}^{s-\bar{s}+1}, \check{\tau}^{\bar{s}-1} \hat{\tau}^{s-\bar{s}+2})$$

which would contradict the maximality of  $\bar{s}$ .

So finally suppose, to the converse, that the principal stopped experimenting after all histories  $(\check{h}^t, \check{\tau}^t) \in \mathcal{R}_{\bar{s}}^t$ . An implication is that

$$U_{\theta_b}(\check{h}^{\bar{s}}, \check{\tau}^{\bar{s}-1}) = 0.$$

The equality follows from the above argument which shows that the principal does not experiment between periods  $\bar{s}+1$  and  $t$  after history  $(\check{h}^{\bar{s}}, \check{\tau}^{\bar{s}-1})$  and that the transfer in period  $\bar{s}$  must be 0 since the principal's continuation value after this history is 0. Then it must be the case that the history corresponding to the other period  $\bar{s}$  non-success outcome  $(\check{h}^{\bar{s}-1} \hat{h}_{\bar{s}}, \check{\tau}^{\bar{s}})$ ,  $\hat{h}_{\bar{s}} \in \{\underline{h}, h_\varphi\}$ ,  $\hat{h}_{\bar{s}} \neq \check{h}_{\bar{s}}$  must either be off-path or we must have  $U_{\theta_b}(\check{h}^{\bar{s}-1} \hat{h}_{\bar{s}}, \check{\tau}^{\bar{s}-1}) = 0$ . To see this, note that if this was not the case, then it is a strict best response for the bad type to costlessly reach history  $(\check{h}^{\bar{s}-1} \hat{h}_{\bar{s}}, \check{\tau}^{\bar{s}-1})$  which is not possible since this would imply  $\tilde{p}(\check{h}^{\bar{s}}, \check{\tau}^{\bar{s}}) = 1$ .

But then  $\tilde{x}(\check{h}^{\bar{s}-1}, \check{\tau}^{\bar{s}-1}) > 0$  is not possible as otherwise we would get the contradiction that  $(\check{h}^{\bar{s}-1}, \check{\tau}^{\bar{s}-1}) \in \mathcal{L}^{\bar{s}-1}$  is a last history. Thus, the principal must experiment after at least one history  $(\check{h}^t, \check{\tau}^t) \in \mathcal{R}_s^t$  and the proof of this step is complete.

*Step 4: Final contradiction.* We define

$$\underline{p}^{\mathcal{R}} = \inf \{ \tilde{p}(h^t, \tau^t) \mid (h^t, \tau^t) \in \mathcal{R}^t, t \geq 0 \} \leq p_0,$$

to be the lowest belief that can arise at a history that does not follow a last history. For  $\varepsilon > 0$ , we pick an  $(h^t, \tau^t) \in \mathcal{R}^t$  such that

$$\tilde{p}(h^t, \tau^t) \leq \underline{p}^{\mathcal{R}} + \varepsilon \text{ and } \tilde{x}(h^t, \tau^t) > 0,$$

and we will argue that the principal's payoff must be negative at such a history if  $\varepsilon$  is small enough.

First observe that a consequence of Step 3 is that such a history must exist. We now define a few additional terms

$$q^{\mathcal{S}} := \tilde{\pi}(\mathcal{S}^{t+s} \mid h^t, \tau^t), \quad q^{\mathcal{L}} := \tilde{\pi}(\mathcal{L}^{t+s} \mid h^t, \tau^t) \text{ and } p^{\mathcal{L}} := \mathbb{E}[\tilde{p}(h^t \hat{h}^s, \tau^t \hat{\tau}^s) \mid (h^t \hat{h}^s, \tau^t \hat{\tau}^s) \in \mathcal{L}^{t+s}].$$

The first two terms are the probabilities that, after  $s$  periods, the players reach a history that follows from a success being generated in a non-last history or a history that follows a last history respectively. The third term is the expected belief at the latter set of histories where the expectation is taken with respect to the distribution  $\tilde{\pi}(\cdot \mid h^t, \tau^t)$ .

The martingale property of beliefs implies

$$\begin{aligned} \tilde{p}(h^t, \tau^t) &= q^{\mathcal{S}} + q^{\mathcal{L}} p^{\mathcal{L}} + (1 - q^{\mathcal{S}} - q^{\mathcal{L}}) \mathbb{E}[\tilde{p}(h^t \hat{h}^s, \tau^t \hat{\tau}^s) \mid (h^t \hat{h}^s, \tau^t \hat{\tau}^s) \in \mathcal{R}^{t+s}] \\ &\geq q^{\mathcal{S}} + q^{\mathcal{L}} p^{\mathcal{L}} + (1 - q^{\mathcal{S}} - q^{\mathcal{L}}) \underline{p}^{\mathcal{R}}, \end{aligned}$$

where, once again, the expectation is taken with respect to the distribution  $\tilde{\pi}(\cdot \mid h^t, \tau^t)$ . This implies that

$$q^{\mathcal{S}} \leq \frac{\varepsilon}{1 - \underline{p}^{\mathcal{R}}} \text{ and } q^{\mathcal{L}} p^{\mathcal{L}} \leq \tilde{p}(h^t, \tau^t), \quad (5)$$

where the first inequality follows from the fact that either  $q^{\mathcal{L}} p^{\mathcal{L}} = 0$  or  $p^{\mathcal{L}} > p_0 \geq \underline{p}^{\mathcal{R}}$  (because beliefs follow a martingale and the beliefs at all last histories are strictly greater than  $p_0$ ) and  $\tilde{p}(h^t, \tau^t) - \underline{p}^{\mathcal{R}} < \varepsilon$ .

Now note that at any last history  $(h^t, \tau^t)$ , we must have  $\underline{\Pi}(\tilde{p}(h^t, \tau^t)) \geq 0$  (because otherwise  $\tilde{x}(h^t, \tau^t) > 0$  cannot be an equilibrium action by the principal) and

$$\begin{aligned} \underline{\Pi}(\tilde{p}(h^t, \tau^t)) &= \underline{\Pi} \left( \mathbb{E} \left[ \tilde{p}(h^t \hat{h}^{s-t'}, \tau^t \hat{\tau}^{s-t'}) \mid (h^t \hat{h}^{s-t'}, \tau^t \hat{\tau}^{s-t'}) \in \mathcal{L}^{s'} \right] \right) \\ &= \mathbb{E} \left[ \underline{\Pi} \left( \tilde{p}(h^t \hat{h}^{s-t'}, \tau^t \hat{\tau}^{s-t'}) \right) \mid (h^t \hat{h}^{s-t'}, \tau^t \hat{\tau}^{s-t'}) \in \mathcal{L}^{s'} \right], \end{aligned} \quad (6)$$

where the expectation is taken with respect to  $\tilde{\pi}(\cdot \mid h^t, \tau^t)$ . The first equality is a consequence of the martingale property of beliefs and the second follows from the fact that  $\underline{\Pi}(\cdot)$  is a linear function.

We can write the principal's payoff at  $(h^t, \tau^t)$  when her realized action choice is to experiment by summing four separate terms: (i) the expected continuation value after a success arrives from

a non-last history, (ii) the expected continuation value at last histories, (iii) the cost of experimentation, transfers and failures at histories  $(h^t h^{s'}, \tau^t \tau^{s'}) \in \mathcal{R}^{t+s'}$ ,  $0 \leq s' \leq s-1$  and (iv) the expected continuation value at all remaining period- $t+s+1$  histories  $(h^t h^s, \tau^t \tau^s) \in \mathcal{R}^{t+s}$ .

We can derive a simple upper bound

$$\underbrace{q^{\mathcal{L}}(1 + \delta\bar{\Pi})}_{(i)} - \underbrace{\delta q^{\mathcal{L}} \mathbb{E} \left[ \underline{\Pi} \left( \tilde{p}(h^t \hat{h}^s, \tau^t \hat{\tau}^s) \right) \mid (h^t \hat{h}^s, \tau^t \hat{\tau}^s) \in \mathcal{L}^{t+s} \right]}_{(ii)} - \underbrace{c}_{(iii)} + \underbrace{\delta^s \bar{\Pi}}_{(iv)},$$

for the principal's payoff when she experiments at  $(h^t, \tau^t)$  and each term is labeled to individually correspond to a bound for each above mentioned component (i)-(iv) of the principal's payoff. The expectation is taken with respect to the distribution  $\tilde{\pi}(\cdot | h^t, \tau^t)$ .

- (i) This term is an upper bound because it assumes that the successes from non-last histories that arrive between periods  $t+1$  and  $t+s$  arrive immediately.
- (ii) This term upper bounds the the sum of payoffs at last histories by assuming that these payoffs are not discounted beyond period  $t+1$ . Consider a last history  $(h^t h^{s'}, \tau^t \tau^{s'})$ ,  $1 \leq s' \leq s$ . From the perspective of period  $t+1$ , the payoff from this history is bounded above by

$$\begin{aligned} 0 &\leq \delta^{s'} \tilde{\pi}(h^t h^{s'}, \tau^t \tau^{s'} | h^t, \tau^t) \underline{\Pi} \left( \tilde{p}(h^t h^{s'}, \tau^t \tau^{s'}) \right) \\ &= \delta^{s'} \tilde{\pi}(h^t h^{s'}, \tau^t \tau^{s'} | h^t, \tau^t) \underline{\Pi} \left( \mathbb{E} \left[ \tilde{p}(h^t h^{s'} h^{s-s'}, \tau^t \tau^{s'} \tau^{s-s'}) \mid (h^t h^{s'} h^{s-s'}, \tau^t \tau^{s'} \tau^{s-s'}) \in \mathcal{L}^{t+s} \right] \right) \\ &\leq \delta \tilde{\pi}(h^t h^{s'}, \tau^t \tau^{s'} | h^t, \tau^t) \mathbb{E} \left[ \underline{\Pi} \left( \tilde{p}(h^t h^{s'} h^{s-s'}, \tau^t \tau^{s'} \tau^{s-s'}) \right) \mid (h^t h^{s'} h^{s-s'}, \tau^t \tau^{s'} \tau^{s-s'}) \in \mathcal{L}^{t+s} \right], \end{aligned}$$

where the expectations are taken with respect to  $\tilde{\pi}(\cdot | h^t h^{s'}, \tau^t \tau^{s'})$ . Summing the last term over all last histories and using the law of iterated expectations gives us the required bound.

- (iii) This term only accounts for the cost of experimentation at  $(h^t, \tau^t)$  but no subsequent costs of experimentation, transfers or losses due to failures.
- (iv) This term is trivially an upper bound for the continuation payoff as it assumes that the principal gets the first best payoff at period  $t+s+1$  for sure.

Now observe that

$$\begin{aligned} &q^{\mathcal{L}}(1 + \delta\bar{\Pi}) + \delta q^{\mathcal{L}} \mathbb{E} \left[ \underline{\Pi} \left( \tilde{p}(h^t \hat{h}^s, \tau^t \hat{\tau}^s) \right) \mid (h^t \hat{h}^s, \tau^t \hat{\tau}^s) \in \mathcal{L}^{t+s} \right] - c + \delta^s \bar{\Pi} \\ &= q^{\mathcal{L}}(1 + \delta\bar{\Pi}) + \delta q^{\mathcal{L}} \underline{\Pi}(p^{\mathcal{L}}) - c + \delta^s \bar{\Pi} \\ &= q^{\mathcal{L}}(1 + \delta\bar{\Pi}) + \delta \underline{\Pi}(q^{\mathcal{L}} p^{\mathcal{L}}) + \delta(1 - q^{\mathcal{L}})c - c + \delta^s \bar{\Pi} \\ &\leq \frac{\varepsilon}{1 - p^{\mathcal{L}}} (1 + \delta\bar{\Pi}) - (1 - \delta(1 - q^{\mathcal{L}}))c + \delta^s \bar{\Pi}, \end{aligned}$$

where the inequality follows from (5) and the fact that  $q^{\mathcal{L}} p^{\mathcal{L}} \leq \tilde{p}(h^t, \tau^t) \leq p_0$ ,  $\underline{\Pi}(p_0) \leq 0$ . This last term is negative is we take  $\varepsilon$  sufficiently small and  $s$  sufficiently large. This shows that there must exist an on-path history  $(\hat{h}^t, \hat{\tau}^t) \in \mathcal{R}^t$  where  $\tilde{x}(\hat{h}^t, \hat{\tau}^t) > 0$  and the principal's payoff is less than 0 which contradicts the existence of a nontrivial equilibrium and completes the proof.  $\blacksquare$

We now show that [Theorem 4](#) implies [Theorem 3](#).

**PROOF OF THEOREM 3.** Take a Nash equilibrium  $(\tilde{x}, \tilde{a}_\theta)$  of the original game and consider the following strategies  $(\tilde{x}', \tilde{\tau}', \tilde{a}'_\theta)$  in the game with transfers:

$$\begin{aligned}\tilde{x}'(h^t, \tau^t) &= \begin{cases} \tilde{x}(h^t, \tau^t) & \text{if } h^t \in \tilde{\mathcal{H}}^t \text{ and } \tau^t = (0, \dots, 0), \\ 0 & \text{otherwise,} \end{cases} \\ \tilde{\tau}'(h^{t+1}, \tau^t) &= 0 \quad (\text{the Dirac measure at } 0), \\ \tilde{a}'_{\theta_b}(h^t, \tau^t) &= \begin{cases} \tilde{a}_{\theta_b}(h^t, \tau^t) & \text{if } h^t \in \tilde{\mathcal{H}}^t \text{ and } \tau^t = (0, \dots, 0), \\ 0 & \text{otherwise,} \end{cases} \\ \tilde{a}'_{\theta_s}(h^t, \tau^t, i^t) &= \begin{cases} \tilde{a}_{\theta_s}(h^t, \tau^t, i^t) & \text{if } h^t \in \tilde{\mathcal{H}}^t \text{ and } \tau^t = (0, \dots, 0), \\ 0 & \text{otherwise,} \end{cases}\end{aligned}$$

where  $\tilde{\mathcal{H}}^t$  is the set of period- $t + 1$  histories that are on path when players use strategies  $(\tilde{x}, \tilde{a}_\theta)$  in the game without transfers. In words, these strategies are identical to  $(\tilde{x}, \tilde{a}_\theta)$  at histories where the principal has not made a transfer in the past. Hence, the only on-path histories  $\tilde{\mathcal{H}}^t$  when players use strategies  $(\tilde{x}', \tilde{\tau}', \tilde{a}'_\theta)$  in the game with transfers will be those where the principal never makes a transfer. Note that  $(\tilde{x}', \tilde{\tau}', \tilde{a}'_\theta)$  will be a Nash equilibrium because, if either player deviates off path, they receive a payoff of 0 since the principal stops experimenting and the agent stops acting.

Finally, note that every equilibrium outcome of the original game is also an equilibrium of the game with transfers; indeed we can use the identical construction above. This is because the refinement does not restrict transfers after a success is generated. Thus, [Theorem 3](#) is an immediate consequence of [Theorem 4](#). ■

## APPENDIX B. PROOF OF [THEOREM 6](#)

We begin by defining an *optimal mechanism*  $\tilde{x}^*$  as

$$\tilde{x}^* \in \underset{\tilde{x}}{\operatorname{argmax}} V(h^0, \tilde{x}, \tilde{a}_\theta),$$

such that  $\tilde{a}_\theta$  is a best response to  $\tilde{x}$ .

Since the principle has commitment, note that the above implies that  $\tilde{x}^*$  does not have to be sequentially rational at all on-path histories.

We now define  $\underline{\kappa}$  as the solution to the equation

$$\lambda_h - c + \max \{ \lambda_l(q_l - (1 - q_l)\underline{\kappa}), -(1 - \lambda_h - \lambda_l)\underline{\kappa} \} + \delta\bar{\Pi} = 0, \quad (7)$$

This value of  $\underline{\kappa}$  is high enough that a single inefficient action by an agent (who otherwise acts efficiently) makes the principal's payoff negative even when the belief assigns probability one to the good type. Note that clearly this also implies that the principal's payoff is not positive at any history (where the belief is less than one) where the good type acts either on low-quality or no information with probability one.

Recall from the statement of [Lemma 1](#), that we had defined a last on-path history  $h^t \in \mathcal{H}$  to be a history where the principal experiments ( $\tilde{x}(h^t) > 0, \tilde{p}(h^t) < 1$ ) such that she stops experimenting

if no success is generated ( $\tilde{x}(h^t h_{t+1} h^t) = 0$  for all  $h_{t+1} \in \{\underline{h}, h_\varphi\}$  and  $h^t h_{t+1} h^t \in \mathcal{H}$ ). While the lemma was for the game with transfers (in Section 4.2), note that it also applies to the game defined in Section 2. This is because the outcome of every Nash equilibrium of the game without transfers can also be achieved in Nash equilibrium when transfers are allowed (since when transfers are off-path, they can be punished).

We now restate Theorem 6 more precisely.

**THEOREM** (Formal Restatement of Theorem 6). *Suppose the prior belief  $p_0$  satisfies  $p_0 \lambda_h - c > 0$ . Then, for all  $\kappa > \underline{\kappa}$  and every optimal mechanism  $\tilde{x}^*$  with associated agent best response  $\tilde{a}_\theta$ , there exists an on-path history  $h^t \in \mathcal{H}$  such that  $V(h^t, \tilde{x}^*, \tilde{a}_\theta) < 0$ .*

This theorem implies that whenever  $p_0 \lambda_h - c > 0$  and  $\kappa > \underline{\kappa}$ , the principal optimal Nash equilibrium generates a strictly lower payoff for the principal than an optimal mechanism since  $V(h^t, \tilde{x}^*, \tilde{a}_\theta) < 0$  is not possible at any on-path history in any Nash equilibrium. The condition on the belief  $p_0$  is sufficient to ensure that all optimal mechanisms are nontrivial.

**PROOF.** We begin by arguing that it is without loss to restrict attention to pure strategy best responses. Formally, consider the pure strategy  $\tilde{a}_\theta^p$  constructed from  $\tilde{a}_\theta$  as follows: at any on-path history where the agent mixes,  $\tilde{a}_\theta^p$  chooses the action that benefits the principal. The agent never acts in  $\tilde{a}_\theta^p$  at any off-path history.

Note that, under  $\tilde{a}_\theta^p$ , the set of on-path histories is a subset of those under  $\tilde{a}_\theta$ . Moreover, for any on-path history  $h^t$  under  $\tilde{a}_\theta^p$ , we must have  $V(h^t, \tilde{x}^*, \tilde{a}_\theta^p) = V(h^t, \tilde{x}^*, \tilde{a}_\theta)$ ; we cannot have  $V(h^t, \tilde{x}^*, \tilde{a}_\theta^p) > V(h^t, \tilde{x}^*, \tilde{a}_\theta)$  as this would imply  $V(h^0, \tilde{x}^*, \tilde{a}_\theta^p) > V(h^0, \tilde{x}^*, \tilde{a}_\theta)$  contradicting the optimality of the latter.

We now assume to the contrary that there is an optimal mechanism  $\tilde{x}^*$  with associated pure strategy best response  $\tilde{a}_\theta^p$  such that  $V(h^t, \tilde{x}^*, \tilde{a}_\theta^p) \geq 0$  for all on-path histories  $h^t \in \tilde{\mathcal{H}}^t$ . We will now show that this is not possible which, in turn, proves the theorem because of the above argument.

*Step 1: The good type never acts on low-quality information or in the absence of information.* Formally,  $\tilde{a}_{\theta_g}^p(h^t, i^t i_{t+1}) = 0$  for all on-path  $(h^t, i^t i_{t+1}) \in \mathcal{H}^s$  with  $i_{t+1} \in \{i_l, i_\varphi\}$ .

Suppose he did. Then, we would immediately have the contradiction  $V(h^t, \tilde{x}^*, \tilde{a}_\theta) < 0$  since we have assumed that  $\kappa > \underline{\kappa}$  which implies that the principal's continuation value is negative whenever the good type acts without high-quality information.

*Step 2: The principal does not experiment at any on-path history where a failure has been generated in the past.* Formally,  $\tilde{x}^*(h^t) = 0$  at every on-path  $h^t = (h_1, \dots, h_t) \in \tilde{\mathcal{H}}^t$ ,  $t \geq 1$  such that  $h_{t'} = \underline{h}$  for some  $1 \leq t' \leq t$ . Consequently, it is without loss to consider  $\tilde{a}_{\theta_b}(h^t) = 0$  for all  $h^t \in \tilde{\mathcal{H}}^t$ .

From Step 1, at any such on-path history  $h^t$ , the principal's belief must assign probability one to the bad type  $\tilde{p}(h^t) = 0$  which would imply that  $V(h^t, \tilde{x}^*, \tilde{a}_\theta^p) < 0$  which is a contradiction. This immediately implies that  $\tilde{a}_{\theta_b}^p(h^t) = 0$  for all  $h^t \in \tilde{\mathcal{H}}^t$  is a best response to the optimal mechanism  $\tilde{x}^*$  since experimentation stops after failure.

Following Steps 1 and 2, it is without loss to consider the following best responses for the agent

$$\tilde{a}_{\theta_b}^p(h^t) = 0 \quad \text{and} \quad \tilde{a}_{\theta_g}^p(h^t, i^t, i_{t+1}) = \begin{cases} 1 & \text{if } i_{t+1} = i_h, \\ 0 & \text{otherwise,} \end{cases}$$

for all (and not just on-path) histories  $h^t \in \mathcal{H}$ .

*Step 3: There must be a last history.*

Suppose not. Then, from the proof of [Lemma 1](#) (which allows for the case where the principal never makes transfers), there must exist an on-path history  $h^t \in \tilde{\mathcal{H}}^t$  such that  $\tilde{x}^*(h^t) > 0$  and  $V(h^t, \tilde{x}^*, \tilde{a}_{\theta}^p) < 0$  which is a contradiction.

*Step 4: At every last history, the principal stops experimenting even after a success.* Formally, following every last history  $h^t \in \tilde{\mathcal{H}}^t$ , it must be the case that  $\tilde{x}^*(h^t \bar{h} h^t) = 0$  for all  $h^t \bar{h} h^t \in \tilde{\mathcal{H}}^{t+t'+1}$ ,  $t' \geq 0$ .

First observe that  $\tilde{p}(h^t) > 0$  as otherwise  $\tilde{x}^*(h^t) > 0$  implies the contradiction that  $V(h^t, \tilde{x}^*, \tilde{a}_{\theta}^p) < 0$ . Now suppose to the converse that the principal experimented following a success at some last history  $h^t$ . Then,  $\tilde{a}_{\theta_g}^p(h_t, i_t, i_t) = 1$  is a strict best response which contradicts Step 1.

*Step 5: Final contradiction.* The principal's payoff can be strictly raised by altering the mechanism at any last history  $h^t$ .

We now define an alternate mechanism  $\tilde{x}$  and show that it can strictly raise the principal's payoff. We begin by taking  $\tilde{x} = \tilde{x}^*$  and then alter  $\tilde{x}$  in what follows; at any history where  $\tilde{x}$  has not been explicitly altered, it is the same as  $\tilde{x}^*$ .

We first change the optimal mechanism  $\tilde{x}^*$  by reducing the amount of experimentation at  $h^t$  but increasing the experimentation in the continuation mechanism (so  $h^t$  is no longer a last history).

We define

$$\tilde{x}(h^t) = \tilde{x}^*(h^t) - \varepsilon, \quad \tilde{x}(h^t \bar{h}) = \gamma \quad \text{and} \quad \tilde{x}(h^t h_\varphi) = q_l \gamma,$$

where  $\varepsilon, \gamma > 0$  are chosen to satisfy

$$(\tilde{x}^*(h^t) - \varepsilon)(1 + (\lambda_h + (1 - \lambda_h)q_l)\delta\gamma) = \tilde{x}^*(h^t). \quad (8)$$

As should be clear from the equation above, the good type's payoff when the principal decides to experiment at  $h^t$  is the same under  $\tilde{x}$  (the left side) as it is under  $\tilde{x}^*$  (the right side).

We will now alter the continuation mechanism when the principal's realized choice at  $h^t$  is not to experiment. The aim is to ensure that the payoff of the good type at  $h^t$  is the same in  $\tilde{x}$  and  $\tilde{x}^*$ . Note that this is necessary because under mechanism  $\tilde{x}$  there is an  $\varepsilon$  greater likelihood that the continuation history  $h^t \bar{h}$  is reached.<sup>13</sup>

We first use  $\bar{h}^{t'} = (\bar{h}, \dots, \bar{h})$  to denote length  $t' \geq 1$  vector where each element corresponds to the principal not experimenting;  $\bar{h}^{t'} = \varphi$  by convention when  $t' = 0$ . We now recursively define  $\tilde{x}$

<sup>13</sup>If we had assumed that there was a public randomization device, this could have been done in a very straightforward way: with probability  $\varepsilon$  at history  $h^t$ , the principal could stop experimenting forever and with probability  $1 - \varepsilon$  the principal's continuation strategy at histories  $h^t \bar{h} h^t \in \mathcal{H}$  is given by  $\tilde{x}(h^t \bar{h} h^t) = \tilde{x}^*(h^t \bar{h} h^t)$ . We instead replicate the above described mechanism without resorting to a public randomization device as follows.

for all  $t' \geq 1$ ,  $h^t \mathcal{H}^{t'} \in \mathcal{H}$  as

$$\tilde{x}(h^t \mathcal{H}^{t'}) \prod_{s=0}^{t'-1} (1 - \tilde{x}(h^t \mathcal{H}^s)) = \tilde{x}^*(h^t \mathcal{H}^{t'}) \prod_{s=0}^{t'-1} (1 - \tilde{x}^*(h^t \mathcal{H}^s)). \quad (9)$$

These terms capture the probability that the principal experiments at period  $t + t' + 1$  but does not experiment between periods  $t + 1$  and  $t + t'$  in mechanisms  $\tilde{x}$  and  $\tilde{x}^*$  respectively. Note that, conditional on experimenting at any period  $t + 2$  onwards, mechanisms  $\tilde{x}$  and  $\tilde{x}^*$  are identical.

We now argue that  $\tilde{a}_{\theta}^p$  is also a best response to  $\tilde{x}$ . First observe, that  $\tilde{x}$  has not been altered at any history where a failure has occurred in the past. Thus, like  $\tilde{x}^*$ , the principal stops experimenting in  $\tilde{x}$  whenever a failure is observed (Step 2) and thus  $\tilde{a}_{\theta_b}^p$  remains a best response.

We finally argue that  $\tilde{a}_{\theta_g}^p$  is a best response to  $\tilde{x}$ . First observe that the agent is indifferent between acting or not on low-quality information at  $h^t$ . Then observe that the good type's payoff at  $h^t$  (before the principal's experimentation decision is realized) is given by

$$\begin{aligned} & U_{\theta_g}(h^t, i^t, \tilde{x}, \tilde{a}_{\theta_g}^p) \\ &= \tilde{x}(h^t) [1 + (\lambda_h + (1 - \lambda_h)q_l)\delta\gamma] \\ & \quad + \sum_{t'=1}^{\infty} \left( \prod_{s=0}^{t'-1} (1 - \tilde{x}(h^t \mathcal{H}^s)) \right) \tilde{x}(h^t \mathcal{H}^{t'}) \left[ 1 + \lambda_h \delta U(h^t \mathcal{H}^{t'} \bar{h}, i^{t+t'} i_h, \tilde{x}, \tilde{a}_{\theta_g}^p) + \lambda_l \delta U(h^t \mathcal{H}^{t'} h_\varphi, i^{t+t'} i_l, \tilde{x}, \tilde{a}_{\theta_g}^p) \right. \\ & \quad \left. + (1 - \lambda_h - \lambda_l) \delta U(h^t \mathcal{H}^{t'} h_\varphi, i^{t+t'} i_\varphi, \tilde{x}, \tilde{a}_{\theta_g}^p) \right] \\ &= \tilde{x}^*(h^t) + \sum_{t'=1}^{\infty} \left( \prod_{s=0}^{t'-1} (1 - \tilde{x}^*(h^t \mathcal{H}^s)) \right) \tilde{x}^*(h^t \mathcal{H}^{t'}) \left[ 1 + \lambda_h \delta U(h^t \mathcal{H}^{t'} \bar{h}, i^{t+t'} i_h, \tilde{x}^*, \tilde{a}_{\theta_g}^p) + \lambda_l \delta U(h^t \mathcal{H}^{t'} h_\varphi, i^{t+t'} i_l, \tilde{x}^*, \tilde{a}_{\theta_g}^p) \right. \\ & \quad \left. + (1 - \lambda_h - \lambda_l) \delta U(h^t \mathcal{H}^{t'} h_\varphi, i^{t+t'} i_\varphi, \tilde{x}^*, \tilde{a}_{\theta_g}^p) \right] \\ &= U_{\theta_g}(h^t, i^t, \tilde{x}^*, \tilde{a}_{\theta_g}^p), \end{aligned}$$

where the equalities follow from (8), (9) and the fact that  $\tilde{x}$  and  $\tilde{x}^*$  are identical at histories  $h^t \mathcal{H}^{t'} h_{t+t'+1}$  for  $h_{t+t'+1} \in \{\bar{h}, h_\varphi\}$ . Hence, the incentive at any history prior to period  $t$  remains unaffected. Also, note that at any history after period  $t + 1$  where the good type can act, his incentives are identical under  $\tilde{x}$  and  $\tilde{x}^*$  and so  $\tilde{a}_{\theta_g}^p$  is also a best response to  $\tilde{x}$ .

We end the proof by showing that the principal's payoff from  $\tilde{x}$  is strictly higher than  $\tilde{x}^*$ . In the algebra that follows we use the shorthand notation  $\eta = \lambda_h - c$  and  $p = \tilde{p}(h^t)$ . Now the principal's payoff from  $\tilde{x}^*$  at history  $h_t$  is:

$$V(h^t, \tilde{x}^*, \tilde{a}_{\theta}^p) = \tilde{x}^*(h^t) (p\lambda_h - c) + (1 - \tilde{x}^*(h^t)) V(h^t \mathcal{H}, \tilde{x}^*, \tilde{a}_{\theta}^p). \quad (10)$$

The principal's payoff from the altered mechanism  $\tilde{x}$  is:

$$\begin{aligned} V(h^t, \tilde{x}, \tilde{a}_{\theta}^p) &= (\tilde{x}^*(h_t) - \varepsilon) [p\lambda_h(1 + \delta\gamma(\lambda_h - c)) + p(1 - \lambda_h)(\delta q_l \gamma(\lambda_h - c)) - (1 - p)\delta q_l \gamma c - c] \\ & \quad + (1 - \tilde{x}(h^t)) V(h^t \mathcal{H}, \tilde{x}, \tilde{a}_{\theta}^p) \\ &= (\tilde{x}^*(h_t) - \varepsilon) [p\lambda_h(1 + \delta\gamma(\lambda_h - c)) + p(1 - \lambda_h)(\delta q_l \gamma(\lambda_h - c)) - (1 - p)\delta q_l \gamma c - c] \\ & \quad + (1 - \tilde{x}^*(h^t)) V(h^t \mathcal{H}, \tilde{x}^*, \tilde{a}_{\theta}^p) \end{aligned} \quad (11)$$

We finally show that the perturbed mechanism gives the principal a strictly higher continuation utility. To see this, take the difference between (11) and (10) to get

$$\begin{aligned} V(h^t, \tilde{x}, \tilde{a}_\theta^p) - V(h^t, \tilde{x}^*, \tilde{a}_\theta^p) &= (\tilde{x}^*(h_t) - \varepsilon)(p\eta\delta\gamma(\lambda_h + (1 - \lambda_h)q_l) - c(1 - p)\delta q_l\gamma) - \varepsilon(p\lambda_h - c) \\ &= (\tilde{x}^*(h_t) - \varepsilon)(p\eta\delta\gamma(\lambda_h + (1 - \lambda_h)q_l) - c(1 - p)\delta q_l\gamma) - \varepsilon p\eta + \varepsilon(1 - p)c. \end{aligned}$$

Substituting in (8), we get

$$\begin{aligned} V(h^t, \tilde{x}, \tilde{a}_\theta^p) - V(h^t, \tilde{x}^*, \tilde{a}_\theta^p) &= -(\tilde{x}^*(h_t) - \varepsilon)c(1 - p)\delta q_l\gamma + \varepsilon(1 - p)c \\ &= (1 - p)c(\varepsilon - (\tilde{x}^*(h_t) - \varepsilon)\delta q_l\gamma) \\ &> 0. \end{aligned}$$

To see the last inequality observe that we can rewrite (8) as

$$\begin{aligned} (\tilde{x}^*(h_t) - \varepsilon)(\delta(\lambda_h + (1 - \lambda_h)q_l)\gamma) &= \varepsilon \\ \implies (\tilde{x}^*(h_t) - \varepsilon)(\delta(\lambda_h(1 - q_l) + q_l)\gamma) &= \varepsilon \\ \implies (\tilde{x}^*(h_t) - \varepsilon)\delta q_l\gamma &< \varepsilon. \end{aligned}$$

■

## REFERENCES

- ABREU, DILIP, P. D., AND E. STACCHETTI (1993): "Renegotiation and Symmetry in Repeated Games," *Journal of Economic Theory*, 60(2), 217–240.
- AGHION, P., AND M. O. JACKSON (2016): "Inducing Leaders to Take Risky Decisions: Dismissal, Tenure, and Term Limits," *American Economic Journal: Microeconomics*, 8(3), 1–38.
- ATAKAN, A. E., AND M. EKMEKCI (2012): "Reputation in Long-Run Relationships," *Review of Economic Studies*, 79(2), 451–480.
- (2013): "A Two-Sided Reputation Result with Long-Run Players," *Journal of Economic Theory*, 148(1), 376–392.
- BACKUS, M., AND A. T. LITTLE (2018): "I Don't Know," Unpublished manuscript, Columbia University.
- BAR-ISAAC, H. (2003): "Reputation and Survival: Learning in a Dynamic Signalling Model," *Review of Economic Studies*, 70(2), 231–251.
- BEN-PORATH, E., E. DEKEL, AND B. L. LIPMAN (2018): "Mechanisms with Evidence: Commitment and Robustness," *Econometrica*, forthcoming.
- COMPTE, O. (1998): "Communication in Repeated Games with Imperfect Private Monitoring," *Econometrica*, 66(3), 597–626.
- CRIPPS, M. W., AND J. P. THOMAS (1997): "Reputation and Perfection in Repeated Common Interest Games," *Games and Economic Behavior*, 18(2), 141–158.
- DEB, R., M. M. PAI, AND M. SAID (2018): "Evaluating Strategic Forecasters," *American Economic Review*, 108(10), 3057–3103.
- ELY, J., D. FUDENBERG, AND D. K. LEVINE (2008): "When is Reputation Bad?," *Games and Economic Behavior*, 63(2), 498–526.
- ELY, J. C., AND J. VÄLIMÄKI (2003): "Bad Reputation," *Quarterly Journal of Economics*, 118(3), 785–814.
- FARRELL, J., AND E. MASKIN (1989): "Renegotiation in Repeated Games," *Games and Economic Behavior*, 1(4), 327–360.
- GUO, Y. (2016): "Dynamic Delegation of Experimentation," *American Economic Review*, 106(8), 1969–2008.
- LI, J., N. MATOUSCHEK, AND M. POWELL (2017): "Power Dynamics in Organizations," *American Economic Journal: Microeconomics*, 9(1), 217–41.
- MITCHELL, M. (2017): "Free (Ad)vice," Unpublished manuscript, University of Toronto.
- MORRIS, S. (2001): "Political Correctness," *Journal of Political Economy*, 109(2), 231–265.
- OTTAVIANI, M., AND P. N. SØRENSEN (2006): "Reputational Cheap Talk," *RAND Journal of Economics*, 37(1), 155–175.
- PRENDERGAST, C., AND L. STOLE (1996): "Impetuous Youngsters and Jaded Old-Timers: Acquiring a Reputation for Learning," *Journal of Political Economy*, pp. 1105–1134.
- SAFRONOV, M., AND B. STRULOVICI (2017): "From Self-Enforcing Agreements to Self-Sustaining Norms," Unpublished manuscript, Northwestern University.