Klein, Nicolas

Article

The importance of being honest

Theoretical Economics

Provided in Cooperation with:
The Econometric Society

Suggested Citation: Klein, Nicolas (2016) : The importance of being honest, Theoretical Economics, ISSN 1555-7561, The Econometric Society, New Haven, CT, Vol. 11, Iss. 3, pp. 773-811, https://doi.org/10.3982/TE1913

This Version is available at:
https://hdl.handle.net/10419/150294

Theoretical Economics 11 (2016), 773–811 1555-7561/20160773

The importance of being honest

Nicolas Klein
Département de Sciences Économiques, Université de Montréal and CIREQ

This paper analyzes the case of a principal who wants to provide an agent with proper incentives to explore a hypothesis that can be either true or false. The agent can shirk, thus never proving the hypothesis, or he can avail himself of a known technology to produce fake successes. This latter option either makes the provision of incentives for honesty impossible or does not distort its costs at all. In the latter case, the principal will optimally commit to rewarding later successes even though he only cares about the first one. Indeed, after an honest success, the agent is more optimistic about his ability to generate further successes. This, in turn, provides incentives for the agent to be honest before a first success.

Keywords. Dynamic moral hazard, continuous-time principal–agent models, optimal incentive scheme, experimentation, bandit models, Poisson process, Bayesian learning.

JEL classification. C79, D82, D83, O32.

1. Introduction

In this paper, I analyze the problem of a principal interested in learning an initially unknown state of the world. To this end, he provides incentives to an agent to engage in experimentation. In particular, it is assumed that, at any point in time, the agent can choose between two projects. One project yields apparent "successes," which are not informative about the state of the world and hence not valuable to the principal, according to a state-independent commonly known distribution. The other project, which is socially valuable, involves the investigation of a hypothesis that is uncertain. This project

Nicolas Klein: kleinnic@yahoo.com

I am most grateful to the editor and to two anonymous referees for their many detailed and helpful comments and suggestions. I thank Johannes Hörner and Sven Rady, for their advice, patience, and encouragement, as well as Guy Arie, Dirk Bergemann, Tri-Vi Dang, Federico Echenique, Philippe Jehiel, Reinoud Joosten, Daniel Krähmer, Lucas Maestri, Thomas Mariotti, Benny Moldovanu, Pauli Murto, Tymofiy Mylovanov, Frank Rosar, Dinah Rosenberg, Francisco Ruiz-Aliseda, Sergei Severinov, Andy Skrzypacz, Eilon Solan, Bruno Strulovici, Juuso Välimäki, Tom Wiseman, Jianjun Wu, and seminar attendees at Arizona State University, Berlin, Bonn, UC Davis, Exeter, Maastricht, Montréal, Paris, Queen's University, Rochester, Stanford, UBC Vancouver, University of Iowa, University of Western Ontario, and various conferences, for helpful comments and discussions. I am especially grateful to the Cowles Foundation for Research in Economics at Yale University for an extended stay during which the idea for this paper took shape. I thank *American Journal Experts* (*AJE*) for editing assistance. Financial support from the National Research Fund of Luxembourg, the German Research Fund through SFB TR-15, and the Fonds de Recherche du Québec Société et Culture is gratefully acknowledged. These funding sources had no involvement in the actual writing or submission of this article.

DOI: 10.3982/TE1913

only yields a success if the hypothesis is true. It is furthermore assumed that the principal cares about the validity of the uncertain hypothesis but can neither tell nor contract upon whether an observed success reflects true success or cheating. Additionally, the agent could shirk, which gives him some private flow benefit. In this case, he will never achieve an observable success. The agent's effort choice is also unobservable to the principal. This paper demonstrates how to implement an honest investigation of the uncertain hypothesis subject to the aforementioned informational restrictions. Specifically, the principal's objective is to minimize the wage costs of implementing honesty with probability 1 on the equilibrium path up to the first success while only observing the occurrence and timing of successes; he does not observe whether a given success was achieved by cheating or by honest means.

Because his actions are unobservable, the agent's pay must depend on his performance to incentivize him to exert effort. Thus, the agent will be paid a substantial bonus if and only if he proves the validity of the hypothesis. This may provide him with the necessary incentives to work, but, unfortunately, it might also tempt him to cheat to try to achieve a fake success. That the mere provision of incentives to exert effort is not sufficient to induce agents to engage in the pursuit of innovation is shown empirically by Francis et al. (2011). Using data from ExecuComp firms for the 1992–2002 period, they demonstrate that the performance sensitivity of chief executive officer (CEO) pay has no impact on a firm's innovation performance, as measured by the number of patents filed or by the number of patent citations.

If the investigation of a correct hypothesis yields breakthroughs at a lower frequency than cheating, then honesty is never implementable. It is thus impossible to incentivize the agent to pursue a low-yield, high-risk project. In the more interesting case of a high-risk project that is also high yield, I describe the schemes that the principal can use to ensure that the agent is always honest, at least until the first breakthrough.

While investigating the hypothesis, the agent grows increasingly pessimistic about its being true as long as no breakthrough arrives. As an honest investigation can never reveal a false hypothesis to be true, all uncertainty is resolved at the first breakthrough, and the agent will be certain of the state of the world. If the agent did not have the option to cheat, the principal could simply offer the agent a reward for the first success, with the reward being just high enough that the agent is willing to exert the effort. Yet, if the agent becomes so pessimistic about the prospects of honesty that the expected arrival rate of a first success is higher when cheating, this scheme could no longer implement honesty.

Thus, to keep the agent honest, the principal will devise a scheme that makes the production of information valuable for the agent as well. While there may be many means of achieving this goal, in one optimal scheme I identify, the principal will reward the agent only for the $(m+1)$st breakthrough, with the chosen m being sufficiently large to deter him from cheating. Whereas the principal has no learning motive because he is only interested in the *first* breakthrough that the agent achieves by honest means, he makes information valuable to the agent to provide incentives. Indeed, the first breakthrough makes the agent more optimistic about his prospects of achieving m future successes but *only if this first breakthrough is achieved by honest means.*

All optimal schemes share the property that cheating is made so unattractive that it is dominated even by shirking.[1] Hence, the agent only needs to be compensated for forgoing the benefits of being lazy. In other words, the presence of a cheating action creates no distortions in players' values, i.e., the payoffs to the principal and the agent are identical regardless of whether the agent has access to the cheating technology.[2]

Still, when the principal can also choose the end date of the interaction, conditional on no breakthrough having been obtained, he will stop the project inefficiently early. The reason for this is that, as in Hörner and Samuelson (2013), future rewards adversely impact today's incentives: If the agent will receive a large payment for achieving his first success tomorrow, he is loath to "risk" having his first success today and thereby forgo the possibility of collecting tomorrow's reward. To overcome this reticence, the principal needs to pay the agent an extra *procrastination rent*, which is increasing in the amount of time remaining. This, in turn, makes longer deadlines less attractive to the principal.

The threshold number of successes m will be chosen to be high enough that, even for an off-path agent, who has achieved his first breakthrough via cheating, m further breakthroughs are so unlikely to be achieved by cheating that he prefers to be honest after his first breakthrough. This puts an off-path agent at a distinct disadvantage, as in contrast to an on-path agent, he has not learned that the hypothesis is true. Thus, only an honest agent has a high level of confidence about his ability to produce many additional successes. Therefore, the agent will want to ensure that he only enters the continuation regime after an honest success. Indeed, an off-path agent, fully aware of his dishonesty, will be comparatively pessimistic about his ability to produce a large number of future successes in the continuation game following the first success. The importance of being honest thus arises endogenously as a tool the principal can use to provide incentives in the cheapest possible way: the principal, who enjoys full commitment power, leverages this difference in beliefs between on-path and off-path agents. Thus, even though later breakthroughs are of no intrinsic value to the principal, it is still optimal for him to tie rewards to consistently outstanding performance as evidenced by a large number of later breakthroughs produced in quick succession.

The continuation phase after a first success can be thought of as a statistical test constructed by the principal. The number of successes required, m, plays the role of a review period in repeated games.[3] In a sense, the construction is also somewhat reminiscent of a Cremer and McLean (1988) mechanism in that the agent is forced to make

[1]Note that the (opportunity) costs of cheating and of being honest are the same, namely the forgone benefits of shirking. If cheating were (much) cheaper than honesty, this conclusion would, of course, no longer hold.

[2]This is because both parties' payoffs are 0 after the game stops. Therefore, from an ex ante perspective, the costs to the parties are the same whether a given sum is transferred via an immediate lump-sum payment and the game ends right away or whether there is a continuation game with the same expected payments. Thus, requiring additional breakthroughs does not entail any waste of resources.

[3]I am indebted to an anonymous referee for this observation. See, e.g., Radner (1985) for the use of *review strategies* in a dynamic moral hazard setting. There, an agent's performance is reviewed every R periods, which allows the principal to achieve the targeted precision of the test through a judicious choice of R. In my setting, the principal does not choose the length of a review period but rather a target number of successes m that makes it unattractive for the agent to enter the continuation phase after a spurious success.

a side bet after a first success is observed, which he evaluates differently depending on whether he is an on-path or off-path agent. The principal, in turn, avails himself of this difference in evaluation to construct the continuation phase in such a way that the agent will not venture off path.

While paying only for the $(m+1)$st breakthrough ensures that off-path agents do not persist in cheating in the continuation game after a first "success," they will nevertheless continue to update their beliefs. Thus, they might be tempted to switch to shirking once they have grown too pessimistic about the hypothesis, a possibility that gives them a positive option value. Because in my model, the agent never makes an "honest mistake," and later breakthroughs are thus of no intrinsic value to the principal, one way for the principal to address this challenge is for him to end the game soon after the first breakthrough, thereby reducing the option value associated with the safe arm by curtailing the amount of time that the agent has access to it. Then, given this end date, the reward for the $(m+1)$st breakthrough is chosen appropriately to yield the intended continuation value to the on-path agent.

The rest of this paper proceeds as follows: Section 2 reviews some of the relevant literature; Section 3 introduces the model; Section 4 addresses the provision of a certain continuation value; Section 5 analyzes the optimal mechanisms before a first breakthrough; Section 6 considers the point at which the principal will optimally elect to stop the project, conditional on no success having occurred; and Section 7 concludes. The technical details of the construction of the continuation scheme are addressed in Appendix A; Appendix B addresses the agent's problem before a first breakthrough, and the proofs not provided within the text are given in Appendix C.

2. Related literature

Holmstrom and Milgrom (1991) analyze a case in which the agent performs several tasks, some of which may be undesirable from the principal's point of view, not unlike my model. The principal may be able to monitor certain activities more accurately than others. While their model could be extended into a dynamic model with the agent controlling the drift rate of a Brownian motion signal,[4] the learning motive I introduce fundamentally changes the basic trade-offs involved. Indeed, in my model, the optimal mechanisms extensively leverage the fact that only an honest agent will have experienced a discontinuous jump in his beliefs.

By contrast, a learning motive is present in Bergemann and Hege (1998, 2005) as well as in Hörner and Samuelson (2013). Those papers examine a venture capitalist's provision of funds for an investment project of initially uncertain quality that is managed by an entrepreneur. The investor cannot observe the entrepreneur's allocation of funds, so off path, the entrepreneur's belief about the quality of the project will differ from the public belief. If the project is good, it yields a success with a probability that is increasing in the amount of funds invested in it; if it is bad, it never yields a success. These papers differ from my model chiefly in that there is no way for the entrepreneur to

[4] See Holmstrom and Milgrom (1987).

"fake" a success; any success that is publicly observed will have been achieved by honest means alone.

Fong (2007) explicitly considers the possibility of cheating in a model without monetary transfers. In my model, the agent is initially no better informed than the principal, while in her model, the agent knows his type from the beginning and adapts his behavior accordingly.[5]

One paper that is close in spirit to mine is Manso (2011), which analyzes a two-period model wherein an agent can shirk, try to produce in some established manner with a known success probability, or experiment with a risky alternative. He demonstrates that to induce experimentation, the principal will optimally not pay for a success in the first period and might even pay for early failure. This distortion is an artifact of the discrete structure of the model and the limited signal space, as early failure can be a very informative signal that the agent has not exploited the known technology but has chosen the risky, unknown alternative. By contrast, while confirming Manso (2011) central intuition that it is better to provide incentives through later rewards, I observe that the presence of the alternative production method does not distort the players' payoffs in continuous time. Now indeed, arbitrary precision of the signal can be achieved by choosing a critical number of successes that is high enough, as will become clear *infra*. Moreover, the dynamic structure allows me to analyze the principal's optimal stopping time.

Considering very general learning processes, Bhaskar (2012) demonstrates that in dynamic moral hazard settings with learning, an agent is always in a position to exploit the misalignment of beliefs following a deviation. This, in turn, makes deviations more attractive and incentive provision more expensive than in a static setting. By virtue of an effect that is somewhat reminiscent of our procrastination rent, he further demonstrates that in a dynamic setting, high-powered future incentives aggravate the incentive problem today by increasing the agent's temptation to exploit a misalignment in beliefs.

A common challenge in moral hazard settings with persistent private information is to demonstrate that local incentive compatibility implies global incentive compatibility. In Arie (2014), the agent's current (private) effort costs are increasing in past effort. Thus, after histories in which the agent has previously deviated by shirking, effort is less costly than the principal would expect. By demonstrating that, at an optimum, the agent is asked to work more following a success than following a failure, Arie (2014) finds that the strongest incentive to shirk prevails on path. In Kwon (2015), the temptation to deviate is also strongest on the equilibrium path. This is intuitively implied by the observation that an agent is more optimistic about the state of the world, and hence more willing to work, after a deviation than when on path. A similar conclusion applies in the present paper (see Proposition 3): an off-path agent will have stronger incentives to be honest than an on-path agent will have.

[5] In Halac et al. (forthcoming), the agent privately knows his type as well. If the agent is of the good type, (unobservable) effort is more productive in that it yields a higher probability of success, provided the project is good. If the project is bad, it never yields a success regardless of the agent's type.

Like Bonatti and Hörner (2011), I embed an exponential bandit framework à la Keller et al. (2005) in a moral hazard problem, availing myself of the tractability of the framework. Garfagnini (2011) and Guo (forthcoming) analyze delegated experimentation without monetary transfers in a related framework.

3. The model

There is one principal and one agent, who are both risk neutral. The agent operates a bandit machine with three arms. One arm is safe in that it yields him a private benefit flow $s > 0$. One arm is known to yield breakthroughs according to a Poisson process with intensity $\lambda_0 > 0$ (arm 0). Finally, arm 1 either yields breakthroughs according to a Poisson process with intensity $\lambda_1 > 0$ (if the time-invariant state of the world $\theta = 1$) or never yields a breakthrough (if the state is $\theta = 0$). The initial probability that $\theta = 1$ is $p_0 \in (0, 1)$. The principal observes all breakthroughs and the time at which they occur; however, he does not observe the arms on which the breakthroughs have been achieved. In addition to observing what the principal can observe, the agent also sees the arms on which the breakthroughs have occurred. The principal and the agent share a common discount rate r. The decision problem (in particular, all the parameter values) is common knowledge.

If the first breakthrough achieved on arm 1 occurs at time t, the principal receives a payoff of $e^{-rt}\Pi > 0$. Later breakthroughs, as well as breakthroughs on arm 0, give the principal no payoff. The principal chooses an end date $\check{T}(t) \in [t, \overline{T})$ (where $\overline{T} < \infty$ is an exogenous finite horizon[6]) in case the first breakthrough occurs at time t. Conditional on there having been no breakthrough, the game ends at time $T < \overline{T}$. Once the game ends, utilities are realized. In the first part of the paper, the end date T is exogenous, and the principal's objective is to ensure, at a minimal cost, that the agent's best response is to use arm 1 until the first breakthrough with probability 1. In Section 6, I let the principal choose the end date T, assuming that he is restricted to implementing arm 1 a.s. before time T.

Formally, the number of breakthroughs achieved on arm i up to and including time t defines the point processes $\{N_t^i\}_{0\le t\le \overline{T}}$ (for $i \in \{0, 1\}$). In addition, let the point process $\{N_t\}_{0\le t\le \overline{T}}$ be defined by $N_t := N_t^0 + N_t^1$ for all t. Moreover, let $\mathfrak{F} := \{\mathfrak{F}_t\}_{0\le t\le \overline{T}}$ and $\mathfrak{F}^N := \{\mathfrak{F}_t^N\}_{0\le t\le \overline{T}}$ denote the filtrations generated by the processes $\{(N_t^0, N_t^1)\}_{0\le t\le \overline{T}}$ and $\{N_t\}_{0\le t\le \overline{T}}$, respectively. The former encodes the evolution of the agent's information over time and captures the idea that at any point in time, the agent knows the timing of all previous breakthroughs and can identify the arm on which each of them has occurred. The latter filtration, capturing the evolution of the principal's information over time, differs from the former in that it groups together previous breakthroughs from the two arms; it captures the idea that at any point in time, the principal can condition his

[6]As we shall see in Section 4, the assumption of a finite horizon will be crucial for the construction of our continuation scheme after a first breakthrough has been observed. This assumption ensures that the agent's belief will always be bounded away from zero after any history. This, in turn, allows us to choose a finite target number of successes m such that off-path agents will not want to continue to pull arm 0 after a first breakthrough on that arm.

actions on the timing of all previous breakthroughs without being able to identify the arm on which they occurred.

By choosing which arm to pull, the agent affects the probabilities of breakthroughs on the different arms. Specifically, if he commits a constant fraction k_0 of his unit endowment flow to arm 0 over a time interval of length $\Delta > 0$, the probability that he achieves at least one breakthrough on arm 0 in that interval is given by $1 - e^{-\lambda_0 k_0 \Delta}$. If he commits a constant fraction of k_1 of his endowment to arm 1 over a time interval of length $\Delta > 0$, the probability of achieving at least one breakthrough on arm 1 in that interval is given by $\theta(1 - e^{-\lambda_1 k_1 \Delta})$.

Formally, a strategy for the agent is a process $\mathbf{k} := \{(k_{0,t}, k_{1,t})\}_t$, which satisfies $(k_{0,t}, k_{1,t}) \in \{(a,b) \in \mathbb{R}^2_+ : a + b \leq 1\}$ for all t and is $\mathfrak{F}$-predictable, where $k_{i,t}$ $(i \in \{0,1\})$ denotes the fraction of the agent's resources that he devotes to arm i at instant t. The $\mathfrak{F}$-predictability captures the idea that the agent chooses his action at instant t before knowing the current outcome of this action. The agent's strategy space, which I denote $\mathcal{U}$, is given by all the processes $\mathbf{k}$ satisfying these requirements. I denote the set of abridged strategies $\mathbf{k}_T$ that prescribe the agent's actions *before the first breakthrough* as $\mathcal{U}_T$.

A *wage scheme* offered by the principal is a process $\{\mathcal{W}_t\}_{0 \leq t \leq \overline{T}}$, which is $\mathfrak{F}^N$-adapted, where $\mathcal{W}_t$ denotes the cumulated discounted time-0 values of the payments that the principal has consciously made to the agent up to and including time t. The assumption that the wage scheme $\{\mathcal{W}_t\}_{0 \leq t \leq \overline{T}}$ is $\mathfrak{F}^N$-adapted captures the idea that payments at instant t can condition on the current outcome. I assume that the agent is protected by limited liability, so that for any history, payments to the agent must be nonnegative at each point in time. Thus, every realization of the process of cumulated payments $\{\mathcal{W}_t\}_{0 \leq t \leq \overline{T}}$ will be nonnegative and nondecreasing.[7] I furthermore assume that the principal has full commitment power, i.e., he commits to a wage scheme $\{\mathcal{W}_t\}_{0 \leq t \leq \overline{T}}$ as well as to a schedule of end dates $\{\check{T}(t)\}_{t \in [0,T]}$, which he announces to the agent at the outset of the game. To ensure that the agent has a best response, I restrict the principal to choosing a piecewise continuous function $t \mapsto \check{T}(t)$.

Over and above the payments he receives as a function of breakthroughs, the agent can secure a safe payoff flow s from the principal by pulling the safe arm, which is unobservable to the principal. The idea is that the principal cannot observe the agent shirking in real time, and such information will surface only after the project is shut down, when the principal will find out ex post that he has been robbed of the payoff flow s during the project. Thus, even though there is no explicit cost to the principal's provision of the bandit in my model, this assumption ensures that the implied flow costs from doing so are at least s.

The principal's objective is to minimize his costs subject to the incentive compatibility constraint, ensuring that it is a best response for the agent to use arm 1 with probability 1 up to the first breakthrough. Thus, I denote the set of *full-experimentation strategies* as $\mathcal{K} := \{\mathbf{k} \in \mathcal{U} : N_t = 0 \Rightarrow k_{1,t} = 1 \text{ for a.a. } t \in [0,T]\}$ and the corresponding set of abridged strategies as $\mathcal{K}_T$.

[7] If the game ends at time $\check{T}$, we set $\mathcal{W}_{\check{T}+\Delta} = \mathcal{W}_{\check{T}}$ for all $\Delta > 0$.

While the state of the world is uncertain, the agent gets new information about the quality of arm 1 whenever he uses it. This *learning* is captured in the evolution of his (private) belief $\hat{p}_t$ that arm 1 is good. Formally, $\hat{p}_t := E[\theta \mid \mathfrak{F}_t, \{(k_{0,\tau}, k_{1,\tau})\}_{0\leq\tau<t}]$. The evolution of beliefs is easy to describe as only a good arm 1 can ever yield a breakthrough. By Bayes' rule,

$$\hat{p}_t = \frac{p_0 e^{-\lambda_1 \int_0^t k_{1,\tau}\, d\tau}}{p_0 e^{-\lambda_1 \int_0^t k_{1,\tau}\, d\tau} + 1 - p_0}$$

and

$$\dot{\hat{p}}_t = -\lambda_1 k_{1,t} \hat{p}_t (1 - \hat{p}_t)$$

prior to the first breakthrough. After the agent has achieved at least one breakthrough on arm 1, his belief will be $\hat{p}_t = 1$ thereafter.

On the equilibrium path, the principal will correctly anticipate $\hat{p}_t$. Thus, in equilibrium, the principal's belief about the agent's belief will be given by $p_t := E[\hat{p}_t \mid \mathfrak{F}_t^N, \mathbf{k} \in \mathcal{K}]$. As in equilibrium, the agent will always operate arm 1 until the first breakthrough, it is clear that if $N_t \geq 1$, then $p_{t+\Delta} = 1$ for all $\Delta > 0$. If $N_t = 0$, Bayes' rule implies that

$$p_t = \frac{p_0 e^{-\lambda_1 t}}{p_0 e^{-\lambda_1 t} + 1 - p_0}.$$

Clearly, as the principal wants to minimize wage payments subject to implementing a full-experimentation strategy, it is never a good idea for him to pay the agent in the absence of a breakthrough. Moreover, as the principal is only interested in the first breakthrough, the notation can be simplified. Let $\{\mathcal{W}_t\}_{0\leq t\leq \overline{T}}$ be the principal's wage scheme and let t be the time of the first breakthrough. I write ϕ_t for the instantaneous lump sum the principal pays the agent as a reward for his first breakthrough; i.e., if $N_t = 1$ and $\lim_{\tau\uparrow t} N_\tau = 0$, $\phi_t := e^{rt}(\mathcal{W}_t - \lim_{\tau\uparrow t} \mathcal{W}_\tau)$. By w_t I denote the expected continuation value of an agent who has achieved his first breakthrough on arm 1 at time t, given he behaves optimally in the future; formally,

$$w_t := \sup_{\{(k_{0,\tau},k_{1,\tau})\}_{t<\tau\leq\check{T}(t)}} E\Bigg[e^{rt}(\mathcal{W}_{\check{T}(t)} - \mathcal{W}_t) + s\int_t^{\check{T}(t)} e^{-r(\tau-t)}(1 - k_{0,\tau} - k_{1,\tau})\, d\tau \;\Big|\; \mathcal{A}_t, \{(k_{0,\tau}, k_{1,\tau})\}_{t<\tau\leq\check{T}(t)} \Bigg],$$

where $\mathcal{A}_t$ denotes the event that the first breakthrough has been achieved on arm 1 at time t, and $\{(k_{0,\tau}, k_{1,\tau})\}_{t<\tau\leq\check{T}(t)}$ is the agent's continuation strategy. Thus, the expectation conditions on the agent's knowledge that the first breakthrough has been achieved on arm 1 at time t. While our previous assumption of limited liability guarantees that $\phi_t \geq 0$ and $w_t \geq 0$ for all $t \in [0, T]$, I additionally impose piecewise continuity of the mappings $t \mapsto \phi_t$ and $t \mapsto w_t$ to ensure that the agent has a best response (see Lemma 2). The corresponding expected continuation payoff of an off-path agent who achieves his first

breakthrough on arm 0 at time t is denoted ω_t, an event I designate $\mathfrak{B}_t$. Formally,

$$\omega_t := \sup_{\{(k_{0,\tau},k_{1,\tau})\}_{t<\tau\leq\check{T}(t)}} E\Big[e^{rt}(\mathcal{W}_{\check{T}(t)} - \mathcal{W}_t) + s\int_t^{\check{T}(t)} e^{-r(\tau-t)}(1-k_{0,\tau}-k_{1,\tau})\,d\tau \;\Big|\; \mathfrak{B}_t, \{(k_{0,\tau},k_{1,\tau})\}_{0\leq\tau\leq\check{T}(t)}\Big],$$

where $\{(k_{0,\tau},k_{1,\tau})\}_{0\leq\tau\leq\check{T}(t)}$ collects the agent's past actions and his continuation strategy. The agent's past actions influence ω_t only via his private belief at time t, $\hat{p}_t$. I therefore write ω_t as a function of $\hat{p}_t$. In the paragraph immediately preceding Lemma 2, I impose assumptions guaranteeing the piecewise continuity of the mapping $t \mapsto \omega_t(\hat{p})$ for any given $\hat{p}$.

Before a first breakthrough, the agent's objective, given a wage scheme $\{\mathcal{W}_t\}_{0\leq t\leq\overline{T}}$ and a schedule of end dates $\{\check{T}(t)\}_{t\in[0,T]}$, depends only on the *first-stage incentives* $\mathbf{g} := (\phi_t, w_t, \omega_t(\hat{p}))_{0\leq t\leq T}$. Given that his belief at time t is given by $\hat{p}_t$, the instantaneous probability of a breakthrough occurring on arm 1 at instant t depends on the agent's action choice at instant t and is given by $k_{1,t}\hat{p}_t\lambda_1$ from the agent's perspective. The instantaneous probability of a breakthrough occurring on arm 0 at instant t is given by $k_{0,t}\lambda_0$. The probability that no breakthrough has occurred before time t depends on the agent's action choices before time t and is given by $e^{-\lambda_1\int_0^t \hat{p}_\tau k_{1,\tau}\,d\tau-\lambda_0\int_0^t k_{0,\tau}\,d\tau}$. Thus, the agent seeks to choose $\mathbf{k}_T \in \mathcal{U}_T$ to maximize

$$\int_0^T \Big\{e^{-rt-\lambda_1\int_0^t \hat{p}_\tau k_{1,\tau}\,d\tau-\lambda_0\int_0^t k_{0,\tau}\,d\tau} \times\big[(1-k_{0,t}-k_{1,t})s + k_{0,t}\lambda_0(\phi_t+\omega_t(\hat{p}_t)) + k_{1,t}\lambda_1\hat{p}_t(\phi_t+w_t)\big]\Big\}\,dt$$

subject to $\dot{\hat{p}}_t = -\lambda_1 k_{1,t}\hat{p}_t(1-\hat{p}_t)$.

The following impossibility result is immediate.

PROPOSITION 1. *If $\lambda_0 \geq \lambda_1$, there does not exist a wage scheme $\{\mathcal{W}_t\}_{0\leq t\leq\overline{T}}$ implementing any strategy in $\mathcal{K}$.*

PROOF. Suppose that $\lambda_0 \geq \lambda_1$ and that there exists a wage scheme $\{\mathcal{W}_t\}_{0\leq t\leq\overline{T}}$ implementing some strategy $\mathbf{k} \in \mathcal{K}$. Now, consider the alternative strategy $\tilde{\mathbf{k}} \notin \mathcal{K}$, which is defined as follows. The agent sets $\tilde{k}_{1,t} = 0$ after all histories and $\tilde{k}_{0,t} = [p_0 e^{-\lambda_1 t}/(p_0 e^{-\lambda_1 t} + 1 - p_0)](\lambda_1/\lambda_0)$ before the first breakthrough. After the first breakthrough, he sets $\tilde{k}_{0,t} = k_{0,t} + (\lambda_1/\lambda_0)k_{1,t} \leq k_{0,t} + k_{1,t}$, history by history. By construction, $\tilde{\mathbf{k}}$ leads to the same distribution over $\{N_t\}_{0\leq t\leq\overline{T}}$ and, hence, over $\{\mathcal{W}_t\}_{0\leq t\leq\overline{T}}$, as $\mathbf{k}$. However, the agent strictly prefers $\tilde{\mathbf{k}}$, as it yields a strictly higher payoff from the safe arm, a contradiction to $\{\mathcal{W}_t\}_{0\leq t\leq\overline{T}}$ implementing $\mathbf{k}$. □

In the rest of this paper, I assume $\lambda_1 > \lambda_0$. Denoting the set of solutions to the agent's problem that are implemented by first-stage incentives $\mathbf{g}$ as $\mathbf{K}^*(\mathbf{g})$, the principal's problem is to choose $\mathbf{g} = (\phi_t, w_t, \omega_t(\hat{p}))_{0\le t\le T}$ to minimize his wage bill

$$\int_0^T e^{-rt-\lambda_1\int_0^t p_\tau\,d\tau}\, p_t\lambda_1(\phi_t + w_t)\,dt$$

subject to $p_t = p_0e^{-\lambda_1 t}/(p_0e^{-\lambda_1 t}+1-p_0)$ and $\mathbf{K}^*(\mathbf{g})\cap\mathcal{K}_T \neq \varnothing$. In fact, the solution to this problem coincides with the solution to the problem in which $\mathbf{K}^*(\mathbf{g}) \subseteq \mathcal{K}_T$ is additionally imposed. That is, it is no costlier to the principal to implement full experimentation in *any* Nash equilibrium than it is to ensure that there exists a Nash equilibrium in which the agent employs a full-experimentation strategy (see Section 5).

4. Incentives after a first breakthrough

4.1 *Introduction*

The purpose of this section is to analyze how the principal will deliver a promised continuation value $w_t > 0$ given that a first breakthrough has occurred at some given time $t \in [0, T]$. His goal will be to find a scheme that maximally discriminates between an agent who has achieved his breakthrough on arm 1 and an agent who has been "cheating," i.e., who has achieved the breakthrough on arm 0. Put differently, for any w_t promised to the on-path agent, the principal strives to reduce the off-path agent's continuation value ω_t, as this will yield a "bigger bang for his buck" in terms of incentives. Because an off-path agent will always have experimented less than an on-path agent, $\hat{p}_t$, his private (off-path) belief at time t will satisfy $\hat{p}_t \in [p_t, p_0]$. As an off-path agent always has the option of imitating the on-path agent's strategy, we know that $\omega_t \ge \hat{p}_t w_t$. The following proposition summarizes the main result of this section: It shows that ω_t can be pushed arbitrarily close to this lower bound.

Proposition 2. *Fix the time of the first breakthrough $t \in [0, T]$. For every $\epsilon > 0$, $w_t > 0$, there exists a continuation scheme such that $\omega_t(\hat{p}_t) \le \hat{p}_t w_t + (s/r)(1 - e^{-r\epsilon})$ for all $\hat{p}_t \in [p_t, p_0]$.*

The proof of this proposition is constructive and is shown in Section 4.2. The construction of the wage scheme relies on the assumption that $\lambda_1 > \lambda_0$, implying that the variance in the number of successes with a good risky arm 1 is higher than with arm 0. Therefore, the principal will structure his wage scheme to reward a number of later breakthroughs that is "extreme enough" that they are very unlikely to have been achieved on arm 0 as opposed to arm 1. Thus, even the most pessimistic of off-path agents would prefer to bet on his arm 1 being good rather than to pull arm 0. In contrast to the off-path agents, an on-path agent knows for sure that his arm 1 is good and therefore he has a distinct advantage in expectation when facing the principal's payment scheme after a first breakthrough. The agent's anticipation of this advantage in turn gives him the proper incentives to use arm 1 rather than arm 0 before the first breakthrough occurs.

4.2 *Construction of an optimal continuation scheme*

The idea of the construction is to approximate to a situation in which an agent who had his first breakthrough on arm 1 continues to use arm 1 until the end of the game, and off-path agents, who had their first breakthrough on arm 0, have no better option than to imitate this behavior. Because $\lambda_1 > \lambda_0$, on-path agents, who know that their arm 1 is good, will never use arm 0. The purpose of the first step of my construction is to make sure that the same holds true for all off-path agents. To this effect, the principal will only pay the agent for the mth breakthrough after time t, where m is chosen to be large enough that even the most pessimistic off-path agents will deem m breakthroughs more likely to occur on arm 1 than on arm 0. Then, in a second step, the end date $\check{T}(t) > t$ is chosen so that $\check{T}(t) - t \leq \epsilon$. This ensures that the agent's option value from being able to switch to the safe arm is bounded from above by $(s/r)(1 - e^{-r\epsilon})$. Then, given the end date $\check{T}(t)$, the reward is chosen appropriately so that the on-path agent receives exactly his promised continuation value of w_t in expectation.

Specifically, the agent is paid only a constant lump sum of $\overline{V}_0$ after his mth breakthrough after time t, where m is sufficiently high that, even for the most pessimistic of all possible off-path agents, arm 1 dominates arm 0. Because $\lambda_1 > \lambda_0$, such an m exists, as the following lemma shows:

LEMMA 1. *There exists an integer m such that if the agent is paid only a lump-sum reward $\overline{V}_0 > 0$ for the mth breakthrough, arm 1 dominates arm 0 for any type of off-path agent whenever m breakthroughs remain to be achieved to collect the lump-sum reward.*

(Recall that all proofs not given in the text are provided in Appendix C.)

Intuitively, the likelihood ratio of m breakthroughs being achieved on arm 1 versus arm 0 in the time interval $(t, \check{T}(t)]$, $\hat{p}_t(\lambda_1/\lambda_0)^m e^{-(\lambda_1-\lambda_0)(\check{T}(t)-t)}$, is unbounded in m. Using the assumption that $\overline{T} < \infty$, which implies that the agent's belief after all histories is bounded from below by some $p_{\overline{T}} > 0$, the proof shows, by virtue of a first-order stochastic dominance argument, that when m exceeds certain thresholds, which can be chosen independently of $\check{T}(t)$, it never pays for the agent to use arm 0.

Thus, Lemma 1 shows that we can ensure that off-path agents will never continue to use arm 0 after time t. Ending the game soon after a first breakthrough, namely, at some time $\check{T}(t) \in (t, t + \epsilon]$, bounds off-path agents' option values from access to the safe arm by $(s/r)(1 - e^{-r\epsilon})$. Hence, an off-path agent of type $\hat{p}_t$ can, at most, obtain $\hat{p}_t w_t + (s/r)(1 - e^{-r\epsilon})$. What remains to be shown is that, given $\check{T}(t)$ and m, $\overline{V}_0$ can be chosen in a manner that ensures that the on-path agent receives the precise payment he is supposed to receive, namely, w_t. While this is essentially a continuity argument, its details are somewhat intricate and technical and are hence relegated to Appendix A.

The principal can apply a similar construction for all times $t \in [0, T]$ at which a first breakthrough is observed. Thus, in summary, the mechanism I have constructed delivers a certain given continuation value of w_t to the on-path agent; it must take care of two distinct concerns to harness maximal incentive power at a given cost. On the one hand, it must ensure that off-path agents never continue to play arm 0. This is achieved by rewarding only the mth breakthrough after time t, where m is sufficiently high. On the other hand, the mechanism must preclude more pessimistic off-path agents from

collecting an excessive option value from their ability to switch between the safe arm and arm 1. This is achieved by ending the game appropriately soon after a first breakthrough. Note that, given the continuation value w_t to be delivered, the principal does not need to know the agent's exact prior belief p_0 for the implementation of this continuation scheme; he only needs to be able to bound the agent's pessimism away from 0.[8] However, to optimally fine tune this w_t, exact knowledge of p_0 becomes necessary, as we shall see in the following section.

5. Incentive provision before a breakthrough

Whereas the previous section addressed the optimal provision of a given *continuation* value w_t, in this section, we analyze optimal incentive provision *before* a first breakthrough. I shall show that, thanks to the continuation scheme we have constructed in the previous section (see Proposition 2), arm 0 can be made so unattractive that in any optimal scheme, it is dominated by the safe arm. Thus, to be induced to use arm 1, the agent only needs to be compensated for his outside option of playing safe, which pins down the principal's wage costs (Proposition 4).

In a first step, we analyze the agent's best responses to given first-stage incentives $(\phi_t, w_t, \omega_t(\hat{p}))_{0\le t\le T}$ so as to derive conditions for the agent to best respond by always using arm 1 until the first breakthrough. In a second step, we will then use these conditions as constraints in the principal's problem as he seeks to minimize his wage bill. While the literature on experimentation with bandits would typically use dynamic programming techniques, this would not be expedient here, as an agent's optimal strategy will depend not only on his current belief and incentives but also on the entire path of future incentives. To the extent that it would be inappropriate to impose any ex ante monotonicity constraints on the incentive scheme, today's scheme need not be a perfect predictor for the future path of incentives; therefore, even a three-dimensional state variable $(\hat{p}_t, \phi_t, w_t)$ would be inadequate. Thus, I shall be using Pontryagin's optimal control approach.

The agent's problem

Given first-stage incentives $(\phi_t, w_t, \omega_t(\hat{p}))_{0\le t\le T}$, the agent chooses $(k_{0,t}, k_{1,t})_{0\le t\le T}$ to maximize

$$\int_0^T \Big\{ e^{-rt-\lambda_1\int_0^t \hat{p}_\tau k_{1,\tau}\,d\tau-\lambda_0\int_0^t k_{0,\tau}\,d\tau} \times\big[(1-k_{0,t}-k_{1,t})s+k_{0,t}\lambda_0(\phi_t+\omega_t(\hat{p}_t))+k_{1,t}\lambda_1\hat{p}_t(\phi_t+w_t)\big]\Big\}\,dt,$$

subject to $\dot{\hat{p}}_t=-\lambda_1 k_{1,t}\hat{p}_t(1-\hat{p}_t)$.

[8] Note that to choose m, exact knowledge of λ_1 is not required either. Indeed, provided that λ_1 is known to be in $[\underline{\lambda}_1,\overline{\lambda}_1]$, where $\underline{\lambda}_1>\lambda_0$, m can be chosen to be high enough given the bounds $\underline{\lambda}_1$ and $\overline{\lambda}_1$. To fine tune the lump sum $\overline{V}_0$ so that the agent obtains precisely a given w_t in expectation, however, the principal does need to know λ_1 precisely. Moreover, the strategic problem if the principal does not know λ_1 precisely is far more complicated, as he would now continue to learn even after a first breakthrough, and the agent could strategically manipulate this learning process. A thorough investigation of these aspects is beyond the scope of this paper.

It will be useful to work with the log-likelihood ratio $x_t := \ln((1-\hat{p}_t)/\hat{p}_t)$ and the probability of no success on arm 0 $y_t := e^{-\lambda_0 \int_0^t k_{0,\tau}\, d\tau}$ as the state variables in our variational problem. These evolve according to $\dot{x}_t = \lambda_1 k_{1,t}$ (to which law of motion I assign the co-state μ_t) and $\dot{y}_t = -\lambda_0 k_{0,t} y_t$ (co-state γ_t), respectively. The initial values $x_0 = \ln((1-\hat{p}_0)/\hat{p}_0)$ and $y_0 = 1$ are given, and x_T and y_T are free. The agent's controls are $(k_{0,t}, k_{1,t}) \in \{(a,b) \in \mathbb{R}_+ : a+b \leq 1\}$.

With slight abuse of notation, I subsequently write ω_t as a function of x_t.[9] To ensure the piecewise continuity of $\omega_t(x_t)$ in t (for a given x_t), I henceforth assume throughout that in the continuation scheme following a first success, the principal applies a threshold number of successes m that is constant in the time of the first breakthrough t. (The proof of Lemma 1 shows that m can be chosen in this way.) To the same end, I assume throughout that $\overline{V}_0$, the lump-sum reward for the $(m+1)$st breakthrough overall, is a piecewise continuous function of t, the time of the first breakthrough.[10] These regularity conditions, together with those imposed in Section 3, guarantee that the agent has a best response, as the following lemma shows.

LEMMA 2. *The agent has a best response to any given first-stage incentives* $(\phi_t, w_t, \omega_t(\hat{p}))_{0\leq t\leq T}$ *satisfying our regularity conditions.*

To state the following proposition, I define $\epsilon_t := \check{T}(t) - t$. I say that a wage scheme is *continuous* if ϕ_t, w_t, and ϵ_t are continuous functions of t. The following proposition shows that if a wage scheme is continuous, then Pontryagin's conditions, which are exhibited in Appendix B, are not only necessary but also sufficient for it to be a best response for the agent to be honest throughout. Moreover, the proposition implies that if the wage scheme is continuous, the conditions will ensure that compliance with the principal's desire for honesty is the agent's *essentially unique* best response (i.e., except possibly for deviations on a null set, which are innocuous to the principal). While the proof of this result is a little tedious, its intuition is straightforward: if incentives at a given time t are strong enough to induce an on-path agent to be honest, any off-path agent, who will necessarily be more optimistic about the quality of arm 1 before a first breakthrough, will have strict incentives to be honest. Continuity now ensures that strict incentives for honesty prevail on an open set just before t as well, on which any off-path

[9]Recall that ω_t is the payoff that an off-path agent receives from best responding to the principal's incentive scheme following a history in which the agent had his first breakthrough on arm 0 at instant t. The agent's payoffs from different responses depend, of course, on the principal's incentive scheme as well as on the agent's private (off-path) belief. Holding the incentive scheme and the time of a first observed breakthrough t constant, one can thus write ω_t as a function of $\hat{p}_t = 1/(1+e^{x_t})$.

[10]As we have seen in Appendix A, we can write $w_t = V_m(t; \overline{V}_0; \check{T}(t))$, where $V_m(t; \overline{V}_0; \check{T}(t))$ denotes the on-path agent's expected payoff at time t given that he has to achieve m breakthroughs to collect the lump sum $\overline{V}_0$, while the game ends at time $\check{T}(t)$. We have shown that $V_m(t; \cdot; \check{T}(t))$ is continuous and strictly increasing if $t_m^* > t$ and constant if $t = t_m^*$, while t_m^* is continuous and increasing in $\overline{V}_0$, and $V_m(t; \overline{V}_0; \cdot)$ is continuous and strictly increasing. Thus, a jump in $\overline{V}_0$ is innocuous (which may be the case because $t_m^* = t$ both before and after the jump, or it is exactly counterbalanced by a jump in $\check{T}(t)$) or it leads to a jump in w_t. Because w_t is piecewise continuous, it follows that there exists a piecewise continuous time path of lump sums $\overline{V}_0(t)$ (as a function of the date of the first breakthrough t) delivering w_t.

agent thus must be honest to satisfy Pontryagin's conditions. Thus, if honesty satisfies the necessary conditions for a best response, it will do so uniquely. This is summarized in the following proposition.

PROPOSITION 3. *Suppose that $k_{1,t}=1$ for all t satisfies Pontryagin's necessary conditions, as stated in Appendix B, even for the upper bound on ω_t given by Proposition 2. Suppose furthermore that ϕ_t, w_t, and ϵ_t are continuous functions of the first breakthrough time t. Then, if $(k_{0,t},k_{1,t})_{0\le t\le T}$ is a best response, it is the case that $k_{1,t}=1$ for a.a. t.*

Our strategy for the rest of this section is to find the cheapest possible schemes that satisfy the agent's necessary conditions for being honest to be a best response. In a second step, we shall examine whether one of these schemes is continuous. If it is, then it must be optimal, because any cheaper scheme would violate the necessary conditions for honesty obtained in our first step.

The principal's problem

We now turn to the problem of the principal, who will take the agent's incentive constraints into account when designing his incentive scheme with a view toward implementing $k_{1,t}=1$ for almost all $t\in[0,T]$. As we have shown in Appendix B, for the agent to best respond by setting $k_{1,t}=1$ at a.a. t, it is necessary that there exist absolutely continuous functions μ_t and γ_t satisfying

$$\dot{\mu}_t=-\dot{\gamma}_t=e^{-rt-x_t}\lambda_1(\phi_t+w_t) \tag{1}$$

for a.a. t as well as the transversality conditions $\mu_T=\gamma_T=0$. Moreover, $x_t=x_0+\lambda_1 t$ and $y_t=1$ for all t. Furthermore, it must be the case that

$$e^{-rt}[e^{-x_t}\lambda_1(\phi_t+w_t)-(1+e^{-x_t})s]\ge-\mu_t\lambda_1 \tag{2}$$

and

$$e^{-rt}\big[e^{-x_t}\lambda_1(\phi_t+w_t)-(1+e^{-x_t})\lambda_0(\phi_t+\omega_t(x_t))\big]\ge-\mu_t(\lambda_1-\lambda_0) \tag{3}$$

for a.a. t.

Clearly, the principal can only gain from keeping ω_t low, which for any given w_t, can be achieved by virtue of the construction in Section 4. For the rest of this section, we will therefore neglect the component $\omega_t(\hat{p})$ in the first-stage incentives and focus on the principal's choice of $(\phi_t,w_t)_{0\le t\le T}$ (with $(\phi_t,w_t)\in[0,L]^2$ at all t for some $L>0$, which is chosen to be large enough) to minimize

$$\int_0^T e^{-rt-\lambda_1\int_0^t p_\tau\,d\tau}p_t\lambda_1(\phi_t+w_t)\,dt$$

subject to the constraints $x_t=x_0+\lambda_1 t$, $y_t=1$, (1), (2), and (3), and the transversality conditions $\mu_T=\gamma_T=0$.

Neglecting constant factors, one can rewrite the principal's objective in terms of the log-likelihood ratio as

$$\int_0^T e^{-(r+\lambda_1)t}(\phi_t + w_t)\,dt.$$

While this expression of the principal's objective is independent of the parties' initial belief p_0, the solution will, of course, depend on the parties' belief via the constraints. Indeed, by (1), we have that

$$\begin{aligned}\mu_t = -\gamma_t &= -\lambda_1 e^{-rt-x_t}\int_t^T e^{-(r+\lambda_1)(\tau-t)}(\phi_\tau + w_\tau)\,d\tau \\ &= -\frac{\lambda_1 p_t}{1-p_t}e^{-rt}\int_t^T e^{-(r+\lambda_1)(\tau-t)}(\phi_\tau + w_\tau)\,d\tau.\end{aligned} \tag{4}$$

Thus, $-\mu_t$ measures the agent's opportunity costs from possibly forgone future rewards. As in Hörner and Samuelson (2013), these future rewards adversely impact today's incentives. Indeed, by pulling arm 1 today, the agent risks having his first breakthrough today, thereby forfeiting the chance to collect the reward offered for achieving a first breakthrough tomorrow. Hence, generous rewards are doubly expensive for the principal: on the one hand, he must pay out more in the case of a breakthrough today; on the other hand, by paying more today, he might make it attractive for the agent to procrastinate at previous points in time in the hopes of winning today's reward. To counteract this effect, the principal must offer higher rewards at previous times to keep incentives intact, which is the effect captured by μ_t. The strength of this effect is proportional to the instantaneous probability of achieving a breakthrough today, $p_t\lambda_1\,dt$. Future rewards are discounted by the rate $r+\lambda_1$, as a higher λ_1 implies a correspondingly lower probability of players' reaching any given future period τ without a breakthrough having previously occurred. This dynamic effect becomes small as players become impatient. Because $\mu_t = -\gamma_t$ for all $t \in [0, T]$, we henceforth only keep track of μ_t.

The following proposition will give a superset of all optimal schemes and exhibit an optimal scheme. Further, it will show that optimality uniquely pins down the principal's wage costs. In the class of schemes with $\phi_t = 0$ for all t, the optimal scheme is essentially unique. This characterization relies on the fact that it never pays for the principal to provide strict rather than weak incentives for the agent to do the right thing, because if he did, he could lower his expected wage bill while still providing adequate incentives. This means that, given that he will do the right thing tomorrow, at any given instant t, the agent is indifferent between doing the right thing and using arm 1, on the one hand, and his next best outside option on the other hand. Yet, the wage scheme we have constructed in Section 4 ensures that if $\phi_t = 0$ for all t, the agent's best outside option can never be arm 0. Indeed, in this case, playing arm 0 yields the agent approximately $p_t w_t$ after a breakthrough, which occurs with an instantaneous probability of $\lambda_0\,dt$ if arm 0 is pulled over a time interval of infinitesimal length dt. Arm 1, by contrast, yields w_t in the case of a breakthrough, which occurs with an instantaneous probability of $p_t\lambda_1\,dt$; thus, as $\lambda_1 > \lambda_0$, arm 1 dominates arm 0. Hence, w_t is pinned down by the binding incentive constraint for the safe arm.

To facilitate the exposition of the following proposition, we define the function $\tilde{w}$ according to

$$\tilde{w}(t) := \begin{cases} \frac{s}{\lambda_1 p_t} + \frac{s}{r}(1 - e^{-r(T-t)}) + \frac{1-p_t}{p_t}\frac{s}{r-\lambda_1}(1 - e^{-(r-\lambda_1)(T-t)}) & \text{if } r \neq \lambda_1 \\ \frac{s}{\lambda_1 p_t} + \frac{s}{r}(1 - e^{-r(T-t)}) + \frac{1-p_t}{p_t} s(T-t) & \text{if } r = \lambda_1. \end{cases}$$

As is readily verified by substituting $\mu_t = -\lambda_1 \int_t^T e^{-r\tau - x_\tau}(\phi_\tau + w_\tau)\, d\tau$ into (2), the incentive constraint for the safe arm, $\tilde{w}(t)$ is the reward that an agent with the belief p_t must be offered at time t to make him exactly indifferent between using arm 1 and the safe arm, given that he will continue to use arm 1 in the future until time T. The first term $s/(\lambda_1 p_t)$ signifies the compensation that the agent must receive for forgoing the immediate flow of $s\,dt$; yet, with an instantaneous probability of $p_t \lambda_1\, dt$, the agent has a breakthrough, and play moves into the continuation phase, which we analyzed in Section 4. In the case of such a success, the agent must be compensated for the forgone access to the safe arm that he would have enjoyed in the absence of a breakthrough. This function is performed by the second term, $(s/r)(1 - e^{-r(T-t)})$. The third term is the *procrastination rent*, i.e., the extra payment the agent must receive to counteract the allure of future incentives. Indeed, not being myopic, the agent takes into account that if he has his first success today, he will forgo his chance of having his first success tomorrow. Thus, the procrastination rent is increasing in the remainder of time, $T - t$, and is arbitrarily small for very impatient agents. We are now ready to characterize the principal's optimal wage schemes.

PROPOSITION 4. *If a wage scheme is optimal, the process $(\phi_t, w_t)_{0\le t\le T}$ it induces is in the set $\mathcal{E}$, with*

$$\mathcal{E} := \left\{ (\phi_t, w_t)_{0\le t\le T} : 0 \le (1-p_t)\phi_t < s\left(\frac{1}{\lambda_0} - \frac{1}{\lambda_1}\right) \text{ and } \phi_t + w_t = \tilde{w}(t) \text{ } t\text{-a.s.} \right\}.$$

If a scheme is in $\mathcal{E}$ and continuous, it is optimal. One optimal wage scheme is given by $\phi_t = 0$ and $w_t = \tilde{w}(t)$ for all $t \in [0, T]$.

PROOF. By construction of $\tilde{w}$, (2) binds at a.a. t for all schemes in $\mathcal{E}$. Algebra shows that (3) holds given that (2) binds if and only if

$$\frac{e^{x_t}}{1 + e^{x_t}}\phi_t + \omega_t(x_t) \le \frac{w_t}{1 + e^{x_t}} + s\left(\frac{1}{\lambda_0} - \frac{1}{\lambda_1}\right). \tag{5}$$

As by Proposition 2, $\omega_t(p_t) > p_t w_t$ yet arbitrarily close to $p_t w_t$, condition (5) is equivalent to the inequality in the definition of $\mathcal{E}$.[11] Clearly, (5) is satisfied for $\phi_t = 0$ and $(s/r)(1 - e^{-r\epsilon_t}) \le s(1/\lambda_0 - 1/\lambda_1)$. As $w_t = \tilde{w}(t)$ is continuous, there exists a continuous ϵ_t satisfying this constraint and delivering $w_t = \tilde{w}(t)$ in the continuation scheme we have constructed in Section 4.

[11] If λ_0 is so low that the construction of Proposition 2 goes through for $m = 1$, the inequality $(1 - p_t)\phi_t \le s(1/\lambda_0 - 1/\lambda_1)$ is weak rather than strict. The same holds true if $w_t = 0$.

By the construction of $\tilde{w}$, any scheme that is not in $\mathcal{E}$ yet satisfies the constraints (1), (2), and (3) a.s., as well as the transversality condition, is more expensive to the principal than any scheme in $\mathcal{E}$. Proposition 3 thus immediately implies that if a scheme is in $\mathcal{E}$ and is continuous, it is optimal. As we have discussed, the scheme given by $\phi_t = 0$ and $w_t = \tilde{w}(t)$ for all t can be made continuous through a judicious choice of ϵ_t. This implies that any scheme outside of $\mathcal{E}$ is dominated by $\phi_t = 0$ and $w_t = \tilde{w}(t)$ for all t and, hence, cannot be optimal. □

Note that this result implies that it is without loss for the principal to restrict himself to schemes that never reward the agent for his first breakthrough even though the first breakthrough is all the principal is interested in. The intuition for this is that by Proposition 2, the principal can ensure that an increase in w_t translates into a smaller increase in ω_t, whereas ϕ_t is paid out indiscriminately to on-path and off-path agents alike. Hence, incentive provision can only be helped when incentives are given through the continuation game rather than through immediate lump-sum payments.

A further immediate implication of the preceding proposition is that the wage payments $\phi_t + w_t$ are a.s. uniquely pinned down. Indeed, the continuation scheme we have constructed in Section 4 makes arm 0 so unattractive to the agent that it is even dominated by the safe arm. This is because, *conditional on* $\theta = 1$, arm 1 dominates arm 0. As we have seen in Section 4, *for a given cost to the principal*, this advantage conditional on the state $\theta = 1$ can be made arbitrarily large for the agent through a judicious choice of the target number of successes m. Due to our assumption of a finite horizon, the agent's belief is bounded away from 0 after all histories so that the *expected* payoff advantage of arm 1 over arm 0 can also be made arbitrarily large, even for the most pessimistic agent. Increasing the payoff advantage of arm 1 over the safe arm, by contrast, is costly to the principal. Therefore, it is the incentive constraint for the safe arm that will bind at the optimum, while the incentive constraint for arm 0 will be slack.

Clearly, optimal wage costs $\tilde{w}$ are decreasing in r, implying that incentives become cheaper to provide the more impatient is the agent. As the agent becomes myopic ($r \to \infty$), wage costs tend to $s/(\lambda_1 p_t)$ because, in the limit, he must now only be compensated for the immediate flow cost of forgoing the safe arm. As the agent becomes infinitely patient ($r \downarrow 0$), wage costs tend to $s/(\lambda_1 p_T) + s(T - t)$. Concerning the evolution of rewards over time, as in Bonatti and Hörner (2011), there are two countervailing effects: On the one hand, the agent becomes more pessimistic over time, so rewards will have to increase to make him willing to work nonetheless; on the other hand, as the deadline approaches, the idea of kicking back and waiting for a future success progressively loses its allure, which should allow the principal to reduce wages somewhat in the here and now. Which effect ultimately dominates depends on the parameters: If players have very high discount rates r, the dynamic effect favoring decreasing rewards becomes very small, and the rewards will be increasing. For a very small r, by contrast, the dynamic effect dominates and the rewards will decrease over time. The discounted rewards $e^{-rt}\tilde{w}(t)$, by contrast, are always strictly decreasing. Furthermore, $(s/r)(1 - e^{-r(t_2-t_1)}) + e^{-r(t_2-t_1)}\tilde{w}(t_2) - \tilde{w}(t_1) < 0$ for all $t_1 \in [0, T)$, $t_2 \in (t_1, T]$. Thus, the

agent would never have an incentive to hide a breakthrough that occurred on arm 1 at instant t_1 and pretend that it in fact happened at some later instant $t_2 > t_1$.[12]

Another immediate implication is the importance of delivering rewards via an "off-line" mechanism, i.e., by means of the continuation game. Indeed, whenever $p_t\lambda_1 \leq \lambda_0$ at a time $t < T$, it is impossible to implement the use of arm 1 on the mere strength of immediate lump-sum rewards. This is easily seen to follow from condition (3), the incentive constraint for arm 0, because $\phi_t \geq 0$ by limited liability:

$$e^{-rt}(p_t\lambda_1 - \lambda_0)\phi_t \geq -\mu_t(1-p_t)(\lambda_1 - \lambda_0) > 0. \tag{6}$$

Conversely, whenever $p_t\lambda_1 > \lambda_0$, it is always possible to increase ϕ_t to make the incentive constraints hold. However, even in this case, it may well be suboptimal for the principal to restrict himself to immediate rewards. As is directly implied by Proposition 4, a necessary condition for immediate rewards to be consistent with optimality at a generic time t is that $\tilde{w}(t) < (s/(1-p_t))(1/\lambda_0 - 1/\lambda_1)$, which one can show is a strictly more stringent condition than $p_t\lambda_1 > \lambda_0$. The reason for this is in the right-hand side of (6), the procrastination rent: if an agent has a success now, he forgoes the chance to obtain the rewards of potential future successes. Note that (3) and, hence, (6) ensure that the agent prefers arm 1 over arm 0, *given that he uses arm 1 at all future times*. With respect to these future rewards, obtaining a success on arm 1 now is bad news in the sense that the agent learns that he indeed would have stood a good chance of obtaining a success on arm 1 at some point in the future, whereas a success on arm 0 conveys no such information. Therefore, while it is always possible to increase ϕ_t to the point that the immediate rewards crowd out this effect, doing so might be costlier than would be necessary to make the agent (weakly) prefer arm 1 over the safe arm. Hence, restricting the principal to instantaneous rewards might be costly even for beliefs $p_t > \lambda_0/\lambda_1$.

6. The optimal stopping time

In the previous section, we have shown that the presence of a cheating option does not lead to any distortions in the players' payoffs but that, nevertheless, the agent must be left a procrastination rent to counteract the allure of future rewards. In this section, we investigate the impact of this distortion in a setting in which we also allow the principal to choose the end date $T \in [0, \overline{T})$ to commit to at the outset of the game (with $\overline{T} < \infty$ chosen to be suitably large). As the first-best benchmark, I use the solution given by the hypothetical situation in which the principal operates the bandit himself and decides when to stop using arm 1, which he pulls at a flow cost of s, conditional on not having obtained a success thus far. Thus, the principal, who obtains a payoff of Π at the first breakthrough, chooses T to maximize

$$\int_0^T \{e^{-rt-\lambda_1\int_0^t p_\tau\,d\tau}(p_t\lambda_1\Pi - s)\}\,dt \tag{7}$$

[12] A similar observation applies to the continuation scheme we constructed in Section 4. There as well, the agent would want to advance the time of the $(m+1)$st breakthrough as much as possible. Yet, if the agent could hide a success on arm 0 and reveal it at a time of his choosing, this conceivably makes first breakthroughs on arm 0 more attractive. A full exploration of a setting in which an agent can hide breakthroughs exceeds the scope of this paper.

subject to $\dot{p}_t = -\lambda_1 p_t(1-p_t)$ for all $t \in (0,T)$. Clearly, the integrand is positive if and only if $p_t\lambda_1\Pi \geq s$, i.e., as long as $p_t \geq s/(\lambda_1\Pi) =: p^m$. As the principal is only interested in the first breakthrough, information has no value for him, meaning that in contrast to the classical bandit literature, he is not willing to forgo current payoffs to *learn* something about the state of the world. In other words, he will behave myopically, i.e., as though the future was of no consequence to him, and stops playing risky at his myopic cutoff belief p^m, which is reached at time $T^{FB} = (1/\lambda_1)\ln[(p_0/(1-p_0))((1-p^m)/p^m)]$.

Regarding the second-best situation wherein the principal delegates the investigation to an agent, I shall compute the optimal end date T, assuming that the principal is restricted to implementing arm 1 a.s. before time T, i.e., his goal is to commit to an end date T to maximize

$$\int_0^T \left\{e^{-rt-\lambda_1\int_0^t p_\tau\,d\tau}\, p_t\lambda_1(\Pi-\tilde{w}(t))\right\} dt \tag{8}$$

subject to $\dot{p}_t = -\lambda_1 p_t(1-p_t)$ for all $t \in (0,T)$.

Thus, all that changes with respect to the first-best problem (7) is that the opportunity cost flow s is now replaced by the optimal wage cost $\tilde{w}(t)$ (see Proposition 4). These, of course, must only be paid out in the case of a success, which occurs with an instantaneous probability of $p_t\lambda_1\,dt$. After substituting for $\tilde{w}(t)$, one finds that the first-order derivative of the objective with respect to T is given by

$$\underbrace{e^{-(r+\lambda_1)T}\left(\lambda_1\Pi - \frac{s}{p_T}\right)}_{\text{Marginal effect}} \underbrace{- e^{-rT}\frac{s}{p_T}(1-e^{-\lambda_1 T})}_{\text{Intramarginal effect}}. \tag{9}$$

The marginal effect captures the benefit the principal could collect by extending experimentation for an additional instant at time T. Yet, as we discussed in Section 5, the choice of an end date T also entails an intramarginal effect at times $t < T$. Indeed, we have seen that to be willing to use arm 1 at time t, the agent must be compensated for the opportunity cost of the potentially forgone rewards associated with a first breakthrough at some future date, an effect that is the stronger the more future remains, i.e., the more distant is the end date T. Hence, by marginally increasing T, the principal also marginally raises his wage liabilities at times $t < T$. Thus, as the following proposition shows, the principal gives up on the project too soon, an effect similar to that found in Hörner and Samuelson (2013) and Bergemann and Hege (2005).[13]

[13]In Hörner and Samuelson (2013), the principal has all the bargaining power, as in this paper. In Bergemann and Hege (2005), by contrast, the agent has all the bargaining power and thus can keep the principal down to his reservation utility of 0. In both papers, the principal has no commitment power. Here, we see that even with full commitment power, the principal cannot overcome this "procrastination effect" due to the dynamic allure of future incentives. While the agent has all the bargaining power in Bergemann and Hege (1998), parties can commit to long-term contracts. In this case, the project may, but need not, be terminated inefficiently early (see their Proposition 5).

PROPOSITION 5. *Let $p_0 > p^m$. The principal stops the game at time $T^* \in (0, T^{FB})$ when $p_{T^*} = p^m e^{\lambda_1 T^*}$, i.e.,*

$$T^* = \frac{1}{\lambda_1} \ln\left(\frac{-p^m p_0 + \sqrt{(p^m p_0)^2 + 4p^m p_0(1-p_0)}}{2p^m(1-p_0)} \right).$$

PROOF. The formula for p_{T^*} is obtained by setting the expression (9) to 0 and verifying that the second-order condition holds. Now, T^* is the unique root of $(p_0 e^{-\lambda_1 T^*})/(p_0 e^{-\lambda_1 T^*} + 1 - p_0) = p^m e^{\lambda_1 T^*}$. □

The stopping times T^{FB} and T^* are both increasing in the players' optimism p_0 as well as in the stakes at play as measured by the ratio $1/p^m = (\lambda_1 \Pi)/s$. Thus, the more optimistic the principal initially is about the agent's ability to produce a real breakthrough and the more important such a breakthrough is to him, the longer he is willing to bear with the agent.

The size of the distortion can be measured by the ratio $(p_{T^*} - p^m)/p^m$, which is also increasing in the stakes at play. This is because of the intramarginal effect we have discussed *supra*: as the stakes increase and the principal consequently extends the deadline T^*, the agent's incentives for procrastination are exacerbated at intramarginal points in time. This, in turn, increases the agent's wages $\tilde{w}(t)$ at these intramarginal points, so the principal can only appropriate part of any increase in the overall pie. Yet, the wedge $(p_{T^*} - p^m)/p^m$ is also increasing in players' optimism, as measured by p_0. Because p^m is independent of p_0, this implies that the threshold belief p_{T^*} is increasing in p_0. This means that the more highly the principal initially thinks of the agent, the higher is the bar to which he will optimally hold him. Although at any time t, wage costs $\tilde{w}(t)$ are decreasing in p_0 and, hence, T^* is increasing in p_0, there is a countervailing effect in the principal–agent game that is absent from the first-best problem. On the one hand, the agent's propensity to procrastinate $|\mu_t|$ is increasing in p_0, i.e., an agent who is more optimistic about his abilities is more likely to "take it easy" and bet on achieving a success tomorrow. On the other hand, similarly to the case of rising stakes, any increase in the end date compounds the agent's proclivity for procrastination. The following proposition summarizes these comparative statics:

PROPOSITION 6. *The stopping time T^* and the wedge $(p_{T^*} - p^m)/p^m$ are increasing in the stakes at play $(\lambda_1 \Pi)/s$ and in players' optimism p_0.*

Yet, recall from the preceding sections that given the optimal incentive scheme we have computed, the principal only needs to compensate the agent for his outside option of using the safe arm. Put differently, the presence of a cheating action, arm 0, does *not* give rise to any distortions; the only distortions that arise are due to the fact that high future rewards cannibalize, to some extent, today's rewards. However, in many applications, the principal's access may not be restricted to a single agent; rather, he might be able to hire several agents sequentially if he chooses. Now, in the limit, if the principal

can hire agents for a mere infinitesimal instant dt, he can completely eliminate the intramarginal effect we discussed above.[14] Indeed, if we assume that subsequent agents observe preceding agents' efforts (so that the agent hired at instant t will have a belief of p_t rather than p_0), we can see from the formula for $\tilde{w}$ that an agent who is only hired for an instant of length dt would have to be promised the reward for a breakthrough given by $(s/(p_t\lambda_1))(1+\lambda_1\,dt)+o(dt)$. Hence, it pays for the principal to continue the project as long as $p_t\lambda_1[\Pi-(s/(p_t\lambda_1))(1+\lambda_1\,dt)]\,dt+o(dt)=p_t\lambda_1[\Pi-s/(p_t\lambda_1)]\,dt+o(dt)>0$, i.e., he stops at the first-best efficient stopping time, a result I summarize in the following proposition.

PROPOSITION 7. *If the principal has access to a sequence of different agents, he stops the delegated project at time T^{FB} when $p_{T^{FB}}=p^m$.*

Thus, while delegating the project to an agent forces the principal to devise quite a complicated incentive scheme, it only induces him to stop the exploration inefficiently early because of the agent's propensity to procrastinate rather than his temptation to cheat. This problem can be overcome, however, if the principal has access to a sequence of many agents. To summarize, if $\lambda_0\geq\lambda_1$, the option to cheat makes it impossible to incentivize the agent to use arm 1; if $\lambda_0<\lambda_1$, by contrast, incentives are optimally structured to obviate any impact of the cheating option on the players' payoffs. When he has access to a sequence of many agents, the principal can completely shut down the procrastination effect, rendering him willing even to implement the efficient amount of experimentation.

7. CONCLUSION

The present paper introduces the question of optimal incentive design into a dynamic single-agent model of experimentation with bandits. I have shown that although the principal cares only about the first breakthrough, it is without loss for him only to reward later successes. Thus, even though the agent will be honest in equilibrium and, hence, the first observed breakthrough reveals everything the principal wants to know, committing to reward only the $(m+1)$st breakthrough can be a potent means of keeping the agent honest in the first place. This is because an agent who has not cheated on his first success is more optimistic about his ability to generate a large number of later successes. Structuring incentives appropriately in this fashion precludes distortions arising from the agent's option to cheat whenever the cheating option does not render the provision of incentives completely impossible.

[14] Intuitively, one might think that hiring *one* particularly myopic agent might remedy the problem as well. However, while it is true that the impact of future rewards on today's incentives, and hence the intramarginal effect of an extended end time T, becomes arbitrarily small as the players become very impatient, the same holds true for the marginal benefit of extending play for an instant after a given time $T>0$, so overall, the distortion is independent of the players' discount rate. If one were to relax the assumption that the players share the same discount rate, the problem could conceivably be addressed by the principal hiring an agent who is much more impatient than himself. The analysis of players with differing discount factors is beyond the scope of this paper.

My incentive scheme is admittedly somewhat extreme, as it relies on several extreme assumptions. First, the agent is endowed with a perhaps unrealistically perfect cheating technology in that the principal can never ex post verify the true value of a breakthrough. However, I show that provided that the parties are risk neutral, it is nevertheless possible to provide incentives for honesty at such a cost to the principal as if the cheating option were not available (provided that $\lambda_1 > \lambda_0$). Thus, the principal's willingness to pay for alternative verification technologies (e.g., to have a second agent check the first agent's breakthrough) would be zero.

This paper should thus be understood more as an exploration of what is theoretically possible in a setting in which an agent can produce auspicious-looking signals by undesirable means than as a prediction of what one would expect to observe in reality. Indeed, the underlying assumptions of risk neutrality and unbounded transfers are unlikely to be satisfied in many real-world situations, so in reality, we will rarely observe incentives as steep as those predicted here. Arguably, the structure of our incentive scheme is somewhat reminiscent of the situation prevailing in some professional sports, where the most lucrative contracts are, from an ex ante perspective, extremely unlikely to be won and tend be awarded to athletes with a great number of successes under their belts.[15]

The scheme I propose heavily relies on the parties' risk neutrality in that the agent is only paid in case of very rare events, yet whenever a payment is made, it will be enormously large. If the agent were risk averse, my scheme clearly would no longer be optimal. Moreover, it may well no longer be optimal for the principal to provide all the incentives via a continuation scheme, which exposes the agent to additional risk. Furthermore, as the provision of incentives via the continuation scheme imposes an additional cost on the principal, it need no longer be the case that the agent will remain indifferent between arm 1 and the safe arm, i.e., the conclusion that the availability of a cheating option does not lead to any distortions in players' payoffs is unlikely to generalize to the case of a risk-averse agent. In addition, as the principal is now averse to fluctuations in the agent's income, it might even be optimal for him to pay the agent in the absence of a breakthrough. What is clear, however, is that the implementation of honesty becomes more expensive if the agent is risk averse, as he must be obliged to bear some risk—lest he use the safe arm. This introduces additional distortion, making the principal end the project even earlier. I recommend a more thorough investigation of these issues in future work.

As my analysis demonstrates, the possibility of essentially unbounded transfers is quite powerful.[16] In fact, one natural conjecture would be that they should allow the

[15]I am indebted to Sergei Severinov for this observation. For instance, in the 2013 Forbes list of the highest paid active athletes (see http://www.forbes.com/athletes/list/ [accessed on June 6, 2014]), the top two spots were held by Tiger Woods ($78.1 mil/yr) and Roger Federer ($71.5 mil/yr). Tiger Woods won his first major in 1997 and has accumulated a rather impressive record of victories since. Roger Federer won his first Grand Slam Singles title at Wimbledon in 2003, and at the time of this writing, has accumulated an additional 16 Grand Slam wins since, making him the all-time record holder for Grand Slam Singles titles.

[16]If there were a binding bound on transfer payments, this would interfere with the construction underlying Proposition 2 by restricting the principal's choice of m and $\check{T}(t)$.

implementation of the principal's desired action as long as the signal structure allowed the principal to statistically discriminate between his desired action and other actions that would be attractive to the agent in the absence of transfers.[17] With uncertainty over the state of the world, this could mean that in some state of the world occurring with positive probability, the agent could produce a distribution of observable signals using the principal's desired action that he could not replicate, in any state of the world occurring with positive probability, by a deviation that he would find profitable under zero transfers. In my setting, this would mean that honesty was implementable as long as the supremum of the realizations of λ_1 to which the agent attached strictly positive probability was strictly higher than λ_0.[18]

In my model, the principal only employs a single agent at any given moment in time. While intuition would suggest that the rationale for rewarding only later breakthroughs should carry over to the case of several agents' simultaneously investigating the same hypothesis, a full investigation of this case would constitute an interesting avenue for future exploration.

Appendix A: Construction of an optimal continuation scheme

A.1 *Construction*

The purpose of this section of the appendix is to show, by virtue of what is essentially a continuity argument, that given $\check{T}(t)$ and m, $\overline{V}_0$ can be chosen in a way that ensures that the on-path agent receives exactly what he is supposed to, namely w_t. To do so, given m, $\check{T}(t)$, and $\overline{V}_0$, I now recursively define the auxiliary functions $V_i(\cdot;\overline{V}_0):[t,\check{T}(t)]\to\mathbb{R}$ for $i=1,\dots,m$ according to

$$V_i(\tilde{t};\overline{V}_0):=\max_{\{k_{i,\tau}\}\in\mathcal{M}(\tilde{t})}\int_{\tilde{t}}^{\check{T}(t)} e^{-r(\tau-\tilde{t})-\lambda_1\int_{\tilde{t}}^{\tau}k_{1,\chi}\,d\chi}\big[s+k_{i,\tau}(\lambda_1 V_{i-1}(\tau;\overline{V}_0)-s)\big]\,d\tau,$$

where $\mathcal{M}(\tilde{t})$ denotes the set of measurable functions $k_i:[\tilde{t},\check{T}(t)]\to[0,1]$, and I set $V_0(\tau;\overline{V}_0):=\overline{V}_0+(s/r)(1-e^{-r(\check{T}(t)-\tau)})$. Thus, $V_i(\tilde{t};\overline{V}_0)$ denotes the agent's continuation value at time $\tilde{t}$ given the agent knows that $\theta=1$ and that he has i breakthroughs to go before being able to collect the lump sum $\overline{V}_0$. I summarize the upshot of this section in the following proposition.

Proposition A.1. *(i) If $w_t>\lim_{\overline{V}_0\downarrow s/\lambda_1}V_m(t;\overline{V}_0)$, there exists a lump sum $\overline{V}_0>s/\lambda_1$ such that $w_t=V_m(t;\overline{V}_0)$.*

(ii) If $w_t\le\lim_{\overline{V}_0\downarrow s/\lambda_1}V_m(t;\overline{V}_0)$, there exists a lump sum $\overline{V}_0>s/\lambda_1$ and an end date $\check{\check{T}}(t)\in(t,\check{T}(t))$ such that $w_t=V_m(t;\overline{V}_0)$ given that the end date is $\check{\check{T}}(t)$.

[17] See Rahman (2010), who shows that under a full support assumption, an allocation is implementable if and only if all deviations that are profitable under zero transfers are statistically detectable.

[18] I am indebted to Philippe Jehiel for this suggestion.

The proof of statement (i) relies on certain properties of the V_i functions, which are exhibited in Lemma A.1 below. The proof of statement (ii) additionally uses another auxiliary function f, which is also introduced *infra* and some properties of which are stated in Lemma A.2 below. The proof is therefore provided in Section A.2 after the proofs of Lemmas A.1 and A.2.

As already mentioned, the following lemma is central to the proof of Proposition A.1. It assumes a fixed end date $\check{T}(t) \leq t + \epsilon$ and notes that once the agent knows that $\theta = 1$, a best response for him is given by a cutoff time t_i^* at which he switches to the safe arm given that he has i breakthroughs to go. It also takes note of some useful properties of the functions V_i.

LEMMA A.1. *Let $\overline{V}_0 > s/\lambda_1$. A best response for the agent is given by a sequence of cutoff times $t_m^* \leq \cdots \leq t_2^* < t_1^* = \check{T}(t)$ (with all inequalities strict if $t_{m-1}^* > t$) such that he uses arm 1 at all times $\tilde{t} \leq t_i^*$ and the safe arm at times $\tilde{t} > t_i^*$ when he still has i breakthroughs to go before collecting the lump sum $\overline{V}_0$. The cutoff time t_i^* ($i = 1, \ldots, m$) is increasing in $\overline{V}_0$; moreover, for $i = 2, \ldots, m$, there exists a constant C_i such that for $\overline{V}_0 > C_i$, the cutoff time t_i^* is strictly increasing in $\overline{V}_0$. The functions $V_i(\cdot; \overline{V}_0)$ are of class C^1 and strictly decreasing; $V_i(\tilde{t}; \cdot)$ is continuous and (strictly) increasing (on $(\overline{V}_0, \infty)$ for $\tilde{t} < t_i^*(\overline{V}_0)$).*[19] *Moreover, $\lim_{\overline{V}_0 \to \infty} t_i^* = \check{T}(t)$ and $\lim_{\overline{V}_0 \to \infty} V_i(\tilde{t}; \overline{V}_0) = \infty$ for any $\tilde{t} \in [t, \check{T}(t))$. The functions V_i satisfy*

$$V_i(\tilde{t}; \overline{V}_0) = \max_{\hat{t} \in [\tilde{t}, \check{T}(t)]} \int_{\tilde{t}}^{\hat{t}} e^{-(r+\lambda_1)(\tau - \tilde{t})} \lambda_1 V_{i-1}(\tau; \overline{V}_0)\, d\tau + \frac{s}{r} e^{-(r+\lambda_1)(\hat{t} - \tilde{t})}(1 - e^{-r(\check{T}(t) - \hat{t})}),$$

and $V_i(\tilde{t}; \overline{V}_0) \leq V_{i-1}(\tilde{t}; \overline{V}_0)$, with the inequality strict for $\tilde{t} < t_i^*$.

See Section A.2 for the proof.

The lemma thus immediately implies that if $w_t > \lim_{\overline{V}_0 \downarrow s/\lambda_1} V_m(t; \overline{V}_0)$ for the given end date $\check{T}(t)$, we can find an appropriate $\overline{V}_0 > s/\lambda_1$ ensuring that $w_t = V_m(t; \overline{V}_0)$, as we note in statement (i) of Proposition A.1.

If $w_t \leq V_m(t; s/\lambda_1)$, we need to lower the end date $\check{T}(t)$ further, as implied by statement (ii) in Proposition A.1. For this purpose, it turns out to be useful to define another auxiliary function $f : [t, \overline{T}) \times (s/\lambda_1, \infty) \to \mathbb{R}$ by $f(\check{T}(t), \overline{V}_0) = V_m(t; \overline{V}_0; \check{T}(t))$, where, in a slight abuse of notation, for any $i = 1, \ldots, m$, I write $V_i(t; \overline{V}_0; \check{T}(t))$ for $V_i(t; \overline{V}_0)$ given that the end date is $\check{T}(t)$. Thus, $f(\check{T}(t), \overline{V}_0)$ maps the choice of the stopping time $\check{T}(t)$ into the on-path agent's time-t expected payoff given the reward $\overline{V}_0 > s/\lambda_1$. The following lemma takes note of some properties of f.

LEMMA A.2. *The function $f(\cdot, \overline{V}_0)$ is continuous and strictly increasing with $f(t; \overline{V}_0) = 0$.*

See Section A.2 for the proof.

As we note in the proof of Proposition A.1, it immediately follows from Lemma A.2 that we can choose a lump sum $\hat{\overline{V}}_0 > s/\lambda_1$ and an end date $\check{\check{T}}(t) < t + \epsilon$ so that $w_t =$

[19] I write $t_i^*(\overline{V}_0)$ for the cutoff t_i^* given that the lump-sum reward is $\overline{V}_0$.

$f(\check{\check{T}}(t), \hat{\overline{V}}_0)$. As one and the same m can be used for all $\check{T}(t)$ and $\hat{\overline{V}}_0$, w_t is piecewise continuous, and $f(\cdot, \overline{V}_0)$ is continuous, it immediately follows that there exists a piecewise continuous $t \mapsto \check{T}(t)$ such that $w_t = f(\check{T}(t); \hat{\overline{V}}_0)$.

A.2 *Proofs of results in Section A.1*

Proof of Lemma A.1 To analyze the agent's best responses, I shall make use of Bellman's principle of optimality. For a given $k_{1,\tilde{t}}$, the Hamilton–Jacobi–Bellman (HJB) equation is given by

$$V_i(\tilde{t}; \overline{V}_0) = \big[s + k_{1,\tilde{t}}(\lambda_1 V_{i-1}(\tilde{t}; \overline{V}_0) - s)\big]\, dt + (1 - r\,dt)(1 - k_{1,\tilde{t}}\lambda_1\, dt)(V_i(\tilde{t}; \overline{V}_0) + \dot{V}_i(\tilde{t}; \overline{V}_0)\, dt) + o(dt).$$

Thus, neglecting terms of order dt^2 and higher and rearranging gives us

$$rV_i(\tilde{t}; \overline{V}_0) = s + \dot{V}_i(\tilde{t}; \overline{V}_0) + k_{1,\tilde{t}}\big[\lambda_1(V_{i-1}(\tilde{t}; \overline{V}_0) - V_i(\tilde{t}; \overline{V}_0)) - s\big]. \tag{10}$$

Hence, $k_{1,\tilde{t}} = 1$ solves the HJB equation if and only if

$$V_{i-1}(\tilde{t}; \overline{V}_0) - V_i(\tilde{t}; \overline{V}_0) \geq \frac{s}{\lambda_1}; \tag{11}$$

it is the unique solution if and only if this inequality is strict.

For $i = 1$, setting $k_{1,\tau} = 1$ for all $\tau \in [\tilde{t}, \check{T}(t)]$ implies

$$V_1(\tilde{t}; \overline{V}_0) = \frac{\lambda_1}{\lambda_1 + r}(1 - e^{-(r+\lambda_1)(\check{T}(t)-\tilde{t})})\left(\overline{V}_0 + \frac{s}{r}\right) - \frac{s}{r}e^{-r(\check{T}(t)-\tilde{t})}(1 - e^{-\lambda_1(\check{T}(t)-\tilde{t})}).$$

Because $\overline{V}_0 > s/\lambda_1$, the derivative $\dot{V}_1$ satisfies

$$\dot{V}_1(\tilde{t}; \overline{V}_0) = -\lambda_1 e^{-(r+\lambda_1)(\check{T}(t)-\tilde{t})}\overline{V}_0 - se^{-r(\check{T}(t)-\tilde{t})}(1 - e^{-\lambda_1(\check{T}(t)-\tilde{t})}) \leq -se^{-r(\check{T}(t)-\tilde{t})} < 0.$$

By simple algebra, one finds that

$$V_0(\tilde{t}; \overline{V}_0) - V_1(\tilde{t}; \overline{V}_0) = \left(\frac{r}{r+\lambda_1} + \frac{\lambda_1}{r+\lambda_1}e^{-(r+\lambda_1)(\check{T}(t)-\tilde{t})}\right)\overline{V}_0 + \frac{s}{r+\lambda_1}(1 - e^{-(r+\lambda_1)(\check{T}(t)-\tilde{t})}),$$

which one shows strictly to exceed s/λ_1 for all $\tilde{t} \in (t, \check{T}(t)]$ if $\overline{V}_0 > s/\lambda_1$. We conclude that $V_1(\cdot; \overline{V}_0)$ is of class C^1 and solves the HJB equation. Hence, V_1 is the value function[20] and a cutoff strategy with $t_1^* = \check{T}(t)$ is optimal. Furthermore, $V_1(\cdot; \overline{V}_0)$ is strictly decreasing with $\dot{V}_1(\tilde{t}; \overline{V}_0) \leq -se^{-r(\check{T}(t)-\tilde{t})}$ for all $\tilde{t}$.

Now let $i > 1$. As my induction hypothesis, I posit that V_{i-1} is of the structure

$$V_{i-1}(\tilde{t}; \overline{V}_0) = \int_{\tilde{t}}^{t_{i-1}^*} e^{-(r+\lambda_1)(\tau-\tilde{t})}\lambda_1 V_{i-2}(\tau; \overline{V}_0)\, d\tau + e^{-(r+\lambda_1)(t_{i-1}^*-\tilde{t})}\frac{s}{r}(1 - e^{-r(\check{T}(t)-t_{i-1}^*)})$$

[20] This follows from a standard verification argument; one can, for instance, apply Proposition 2.1 in Bertsekas (1995, p. 93).

if $\tilde{t} \leq t_{i-1}^*$ and

$$V_{i-1}(\tilde{t}; \overline{V}_0) = \frac{s}{r}(1 - e^{-r(\check{T}(t)-\tilde{t})})$$

if $\tilde{t} > t_{i-1}^*$, for some $t_{i-1}^* \leq \check{T}(t)$. It is furthermore assumed that $V_{i-1}(\cdot; \overline{V}_0)$ is C^1 and that $\dot{V}_{i-1}(\tilde{t}; \overline{V}_0) \leq -se^{-r(\check{T}(t)-\tilde{t})}$ for all $\tilde{t} \in (t, \check{T}(t))$.

Now, if $V_{i-1}(t; \overline{V}_0) < s/\lambda_1 + (s/r)(1 - e^{-r(\check{T}(t)-t)})$, I set $t_i^* = t$. Otherwise, I define t_i^* as the lowest t^* satisfying $V_{i-1}(t^*; \overline{V}_0) = s/\lambda_1 + (s/r)(1 - e^{-r(\check{T}(t)-t^*)})$. Because $\dot{V}_{i-1}(\tilde{t}; \overline{V}_0) \leq -se^{-r(\check{T}(t)-\tilde{t})}$ for all $\tilde{t} \in (t, \check{T}(t))$, $V_{i-1}(\cdot; \overline{V}_0)$ is continuous, and $V_{i-1}(\check{T}(t); \overline{V}_0) = 0$, it is the case that t_i^* exists, and $t_i^* < \check{T}(t)$.

Fix an arbitrary $\tilde{t} \in (t, \check{T}(t))$. If $V_{i-1}(\tilde{t}; \overline{V}_0) \leq s/\lambda_1 + (s/r)(1 - e^{-r(\check{T}(t)-\tilde{t})})$, i.e., $\tilde{t} \geq t_i^*$, $k_{1,\hat{\tau}} = 0$ for all $\hat{\tau} \in [\tilde{t}, \check{T}(t)]$, and its corresponding payoff function $V_i(\hat{\tau}; \overline{V}_0) = (s/r)(1 - e^{-r(\check{T}(t)-\hat{\tau})})$ solves the HJB equation. Indeed, the payoff function $V_i(\hat{\tau}; \overline{V}_0) = (s/r)(1 - e^{-r(\check{T}(t)-\hat{\tau})})$ is of class C^1, and because $\dot{V}_{i-1}(\hat{\tau}; \overline{V}_0) \leq -se^{-r(\check{T}(t)-\hat{\tau})}$, we have that $V_{i-1}(\hat{\tau}; \overline{V}_0) - V_i(\hat{\tau}; \overline{V}_0) \leq s/\lambda_1$ at all times $\hat{\tau} \in [\tilde{t}, \check{T}(t)]$. This establishes that V_i is indeed the value function and that $k_{1,\tilde{t}} = 0$ is a best response for all $\tilde{t} \geq t_i^*$.[21]

Now let us assume that $V_{i-1}(\tilde{t}; \overline{V}_0) > s/\lambda_1 + (s/r)(1 - e^{-r(\check{T}(t)-\tilde{t})})$. I shall now show that $k_{1,\hat{\tau}} = 1$ for all $\hat{\tau} \in [\tilde{t}, t_i^*]$, $k_{1,\hat{\tau}} = 0$ for all $\hat{\tau} \in (t_i^*, \check{T}(t)]$, and its appertaining payoff function

$$V_i(\hat{\tau}; \overline{V}_0) = \begin{cases} \int_{\hat{\tau}}^{t_i^*} e^{-(r+\lambda_1)(\tau-\hat{\tau})} \lambda_1 V_{i-1}(\tau; \overline{V}_0)\, d\tau + e^{-(r+\lambda_1)(t_i^*-\hat{\tau})} \frac{s}{r}(1 - e^{-r(\check{T}(t)-t_i^*)}) & \text{if } \hat{\tau} \leq t_i^* \\ \frac{s}{r}(1 - e^{-r(\check{T}(t)-\hat{\tau})}) & \text{if } \hat{\tau} > t_i^* \end{cases}$$

for $\hat{\tau} \in [\tilde{t}, \check{T}(t)]$, solve the HJB equation. To do so, it is sufficient to show that V_i is C^1 and that $V_{i-1}(\hat{\tau}; \overline{V}_0) - V_i(\hat{\tau}; \overline{V}_0) \geq s/\lambda_1$ for all $\hat{\tau} \in [\tilde{t}, t_i^*]$, while $V_{i-1}(\hat{\tau}; \overline{V}_0) - V_i(\hat{\tau}; \overline{V}_0) \leq s/\lambda_1$ for all $\hat{\tau} \in (t_i^*, \check{T}(t)]$.

First, let $\hat{\tau} \leq t_i^*$. Using the fact that, by absolute continuity of $V_{i-1}(\cdot; \overline{V}_0)$, we have that for $\tau \geq \hat{\tau}$,

$$V_{i-1}(\tau; \overline{V}_0) = V_{i-1}(\hat{\tau}; \overline{V}_0) + \int_{\hat{\tau}}^{\tau} \dot{V}_{i-1}(\sigma; \overline{V}_0)\, d\sigma \leq V_{i-1}(\hat{\tau}; \overline{V}_0) - \frac{s}{r} e^{-r(\check{T}(t)-\hat{\tau})}(e^{r(\tau-\hat{\tau})} - 1)$$

by our induction hypothesis, one shows that the following condition is sufficient for $V_{i-1}(\hat{\tau}; \overline{V}_0) - V_i(\hat{\tau}; \overline{V}_0) \geq s/\lambda_1$:

$$\left[\frac{r}{r+\lambda_1} + \frac{\lambda_1}{r+\lambda_1} e^{-(r+\lambda_1)(t_i^*-\hat{\tau})}\right]\left[V_{i-1}(\hat{\tau}; \overline{V}_0) + \frac{s}{r} e^{-r(\check{T}(t)-\hat{\tau})}\right] - \frac{s}{r} e^{-(r+\lambda_1)(t_i^*-\hat{\tau})} - \frac{s}{\lambda_1} \geq 0. \tag{12}$$

[21] If $V_{i-1}(\tilde{t}; \overline{V}_0) = s/\lambda_1 + (s/r)(1 - e^{-r(\check{T}(t)-\tilde{t})})$, we have just argued that the value function is given by $V_i(\tilde{t}; \overline{V}_0) = (s/r)(1 - e^{-r(\check{T}(t)-\tilde{t})})$. In this case, any $k_{1,\tilde{t}} \in [0, 1]$ is a best response. *Infra*, it is shown that this indifference can only occur at t_i^*.

As $\hat{\tau} \le t_i^*$, we have that $V_{i-1}(\hat{\tau}; \overline{V}_0) \ge s/\lambda_1 + (s/r)(1 - e^{-r(\check{T}(t)-\hat{\tau})})$, which implies that (12) holds because

$$\left[\frac{r}{r+\lambda_1} + \frac{\lambda_1}{r+\lambda_1} e^{-(r+\lambda_1)(t_i^*-\hat{\tau})}\right]\left[\frac{s}{\lambda_1} + \frac{s}{r}\right] - \frac{s}{r} e^{-(r+\lambda_1)(t_i^*-\hat{\tau})} - \frac{s}{\lambda_1} = 0.$$

Moreover, we have that $\dot{V}_i(\hat{\tau}; \overline{V}_0) = -se^{-r(\check{T}(t)-\hat{\tau})}$ if $\hat{\tau} > t_i^*$ and

$$\dot{V}_i(\hat{\tau}; \overline{V}_0) = -\lambda_1 e^{-(r+\lambda_1)(t_i^*-\hat{\tau})} V_{i-1}(t_i^*; \overline{V}_0) + \frac{r+\lambda_1}{r} e^{-(r+\lambda_1)(t_i^*-\hat{\tau})}(1 - e^{-r(\check{T}(t)-t_i^*)})s + \lambda_1 \int_{\hat{\tau}}^{t_i^*} e^{-(r+\lambda_1)(\tau-\hat{\tau})} \dot{V}_{i-1}(\tau; \overline{V}_0)\, d\tau$$

for $\hat{\tau} < t_i^*$. Hence, using $V_{i-1}(t_i^*; \overline{V}_0) = s/\lambda_1 + (s/r)(1 - e^{-r(\check{T}(t)-t_i^*)})$, one shows that $\lim_{\hat{\tau}\uparrow t_i^*} \dot{V}_i(\hat{\tau}; \overline{V}_0) = -se^{-r(\check{T}(t)-t_i^*)} = \lim_{\hat{\tau}\downarrow t_i^*} \dot{V}_i(\hat{\tau}; \overline{V}_0)$, implying that $V_i(\cdot; \overline{V}_0)$ is of class C^1. Thus, I have shown that $k_{1,\hat{\tau}} = 1$ for all $\hat{\tau} \in [\tilde{t}, t_i^*]$, $k_{1,\hat{\tau}} = 0$ for all $\hat{\tau} \in (t_i^*, \check{T}(t)]$, and

$$V_i(\tilde{t}; \overline{V}_0) = \begin{cases} \int_{\tilde{t}}^{t_i^*} e^{-(r+\lambda_1)(\tau-\tilde{t})} \lambda_1 V_{i-1}(\tau; \overline{V}_0)\, d\tau + e^{-(r+\lambda_1)(t_i^*-\tilde{t})} \frac{s}{r}(1 - e^{-r(\check{T}(t)-t_i^*)}) & \text{if } \tilde{t} \le t_i^* \\ \frac{s}{r}(1 - e^{-r(\check{T}(t)-\tilde{t})}) & \text{if } \tilde{t} > t_i^* \end{cases}$$

solve the HJB equation. Hence, V_i is indeed the value function.

It remains to prove that $\dot{V}_i(\tilde{t}; \overline{V}_0) \le -se^{-r(\check{T}(t)-\tilde{t})}$ for $\tilde{t} < t_i^*$. Yet, this is easily shown to follow from the fact that, by induction hypothesis, $\dot{V}_{i-1}(\tilde{t}; \overline{V}_0) \le -se^{-r(\check{T}(t)-\tilde{t})}$, and hence

$$\lambda_1 \int_{\tilde{t}}^{t_i^*} e^{-(r+\lambda_1)(\tau-\tilde{t})} \dot{V}_{i-1}(\tau; \overline{V}_0)\, d\tau \le -se^{-r(\check{T}(t)-\tilde{t})}(1 - e^{-\lambda_1(t_i^*-\tilde{t})}),$$

which completes the induction step.

Now, consider some $i \in \{1, \dots, m-1\}$. Having established that the agent's best response is given by a cutoff strategy, I shall now show that $t_{i+1}^* \le t_i^*$. Consider an arbitrary time $\tilde{t} \ge t_i^*$ and suppose the agent still has $i+1$ breakthroughs to go. By stopping at an arbitrary time $t^* \in (\tilde{t}, \check{T}(t)]$, the agent can collect

$$\begin{aligned} &\int_{\tilde{t}}^{t^*} \lambda_1 \frac{s}{r} e^{-(r+\lambda_1)(\tau-\tilde{t})}(1 - e^{-r(\check{T}(t)-\tau)})\, d\tau + \frac{s}{r} e^{-(r+\lambda_1)(t^*-\tilde{t})}(1 - e^{-r(\check{T}(t)-t^*)}) \\ &\quad = \frac{s}{r}\left[\frac{\lambda_1}{\lambda_1+r}(1 - e^{-(r+\lambda_1)(t^*-\tilde{t})}) - e^{-r(\check{T}(t)-\tilde{t})}(1 - e^{-\lambda_1(t^*-\tilde{t})})\right] \\ &\qquad\qquad + \frac{s}{r} e^{-(r+\lambda_1)(t^*-\tilde{t})}(1 - e^{-r(\check{T}(t)-t^*)}). \end{aligned}$$

By stopping immediately at time $\tilde{t}$, he can collect $(s/r)(1 - e^{-r(\check{T}(t)-\tilde{t})})$. Thus, because

$$\begin{aligned} 1 - e^{-r(\check{T}(t)-\tilde{t})} &> \frac{\lambda_1}{\lambda_1+r}(1 - e^{-(r+\lambda_1)(t^*-\tilde{t})}) \\ &\qquad - e^{-r(\check{T}(t)-\tilde{t})}(1 - e^{-\lambda_1(t^*-\tilde{t})}) + e^{-(r+\lambda_1)(t^*-\tilde{t})}(1 - e^{-r(\check{T}(t)-t^*)}) \\ \iff \quad 1 &> \frac{\lambda_1}{r+\lambda_1} + \frac{r}{r+\lambda_1} e^{-(r+\lambda_1)(t^*-\tilde{t})}, \end{aligned}$$

the agent strictly prefers to stop immediately at $\tilde{t}$. For $\tilde{t} = t_i^*$, in particular, we can conclude that $t_{i+1}^* \leq t_i^*$; if $t_i^* > t$, we have that $t_{i+1}^* < t_i^*$.

Clearly, if $\hat{\overline{V}}_0 > \overline{V}_0$, we have that $V_i(\tilde{t}; \hat{\overline{V}}_0) \geq V_i(\tilde{t}; \overline{V}_0)$ for all $\tilde{t} \in [t, \check{T}(t)]$ and all $i = 1, \dots, m$, as the agent can always use the strategy that was optimal given the reward $\overline{V}_0$, and be no worse off when the reward is $\hat{\overline{V}}_0$ instead. Moreover, $V_1(\tilde{t}; \cdot)$ is strictly increasing for all $\tilde{t} < t_1^* = \check{T}(t)$, with $\lim_{\overline{V}_0 \to \infty} V_1(\tilde{t}; \overline{V}_0) = \infty$. I posit the induction hypothesis that for all $\overline{V}_0 \in (s/\lambda_1, \infty)$ and all $\tilde{t} < t_{i-1}^*(\overline{V}_0)$, we have that $V_{i-1}(\tilde{t}; \cdot)$ is strictly increasing on $(\overline{V}_0, \infty)$, with $\lim_{\hat{\overline{V}}_0 \to \infty} V_{i-1}(\tilde{t}; \hat{\overline{V}}_0) = \infty$. As playing a cutoff strategy with the old cutoff $t_i^*(\overline{V}_0)$ is always a feasible strategy for the agent, we can conclude that for $\tilde{t} < t_i^*(\overline{V}_0) < t_{i-1}^*(\overline{V}_0)$ and $\hat{\overline{V}}_0 > \overline{V}_0$,

$$\begin{aligned} V_i(\tilde{t}; \hat{\overline{V}}_0) &\geq \int_{\tilde{t}}^{t_i^*(\overline{V}_0)} \lambda_1 e^{-(r+\lambda_1)(\tau - \tilde{t})} V_{i-1}(\tau; \hat{\overline{V}}_0)\, d\tau + \frac{s}{r} e^{-(r+\lambda_1)(t_i^*(\overline{V}_0) - \tilde{t})} (1 - e^{-r(\check{T}(t) - t_i^*(\overline{V}_0))}) \\ &> V_i(\tilde{t}; \overline{V}_0), \end{aligned}$$

with the last inequality following from the fact that $\tilde{t} < t_i^*(\overline{V}_0) < t_{i-1}^*(\overline{V}_0)$, implying by our induction hypothesis that $V_{i-1}(\tau; \hat{\overline{V}}_0) > V_{i-1}(\tau; \overline{V}_0)$ for all $\tau \in [\tilde{t}, t_i^*(\overline{V}_0)]$. By the same token, our induction hypothesis implies that $V_{i-1}(\tau; \hat{\overline{V}}_0) \to \infty$ as $\hat{\overline{V}}_0 \to \infty$ for all $\tau \in [\tilde{t}, t_i^*(\overline{V}_0)]$, so that we can conclude that $\lim_{\hat{\overline{V}}_0 \to \infty} V_i(\tilde{t}; \hat{\overline{V}}_0) = \infty$. To sum, $V_i(\tilde{t}; \cdot)$ is increasing, and strictly increasing on $(\overline{V}_0, \infty)$ with $\lim_{\overline{V}_0 \to \infty} V_i(\tilde{t}; \overline{V}_0) = \infty$ if $\tilde{t} < t_i^*(\overline{V}_0)$ for all $i = 1, \dots, m$.

Suppose $t_{i+1}^*(\overline{V}_0) > t$. Then $t_{i+1}^*(\overline{V}_0)$ is defined as the smallest root to $V_i(t_{i+1}^*; \overline{V}_0) = s/\lambda_1 + (s/r)(1 - e^{-r(\check{T}(t) - t_{i+1}^*)})$. As $t_i^*(\overline{V}_0) > t_{i+1}^*(\overline{V}_0)$, we furthermore know by our previous step that $V_i(t_{i+1}^*(\overline{V}_0); \cdot)$ is strictly increasing on $(\overline{V}_0, \infty)$ at $t_{i+1}^*(\overline{V}_0)$. Hence, we have that $t_{i+1}^*(\hat{\overline{V}}_0) > t_{i+1}^*(\overline{V}_0)$ for all $\hat{\overline{V}}_0 > \overline{V}_0$. We conclude that the cutoff $t_{i+1}^*(\cdot)$ is strictly increasing on $(\overline{V}_0, \infty)$.

Now, suppose that $t_{i+1}^*(\overline{V}_0) = t$. Then $V_i(t; \overline{V}_0) \leq s/\lambda_1 + (s/r)(1 - e^{-r(\check{T}(t) - t)})$. Let $j := \min\{\iota \in \{1, \dots, m\} : t_\iota^*(\overline{V}_0) = t\}$. Because $t_1^* = \check{T}(t) > t$, we have that $j \geq 2$. Now $V_{j-1}(t; \cdot)$ is strictly increasing in $(\overline{V}_0, \infty)$ with $\lim_{\hat{\overline{V}}_0 \to \infty} V_{j-1}(t; \hat{\overline{V}}_0) = \infty$. Hence, there exists a constant C_{j-1} such that for $\hat{\overline{V}}_0 > C_{j-1}$, we have that $V_{j-1}(t; \hat{\overline{V}}_0) > s/\lambda_1 + (s/r)(1 - e^{-r(\check{T}(t) - t)})$, and hence $t_j^*(\hat{\overline{V}}_0) > t$. Iterated application of this argument yields the existence of a constant C_i such that $\overline{V}_0 > C_i$ implies that $t_{i+1}^*(\overline{V}_0) > t$. Hence, by our previous step, t_{i+1}^* is strictly increasing in $\overline{V}_0$ for $\overline{V}_0 > C_i$.

Now consider arbitrary $\tilde{t} \in [t, \check{T}(t))$ and $i \in \{1, \dots, m\}$. Let σ be defined by $\sigma := \max\{\iota \in \{1, \dots, m\} : t_\iota^*(\overline{V}_0) > \tilde{t}\}$. As $\tilde{t} < \check{T}(t) = t_1^*$, $\sigma \geq 1$. As $\tilde{t} < t_\sigma^*(\overline{V}_0)$, $V_\sigma(\tilde{t}; \cdot)$ is strictly increasing in $(\overline{V}_0, \infty)$, with $\lim_{\hat{\overline{V}}_0 \to \infty} V_\sigma(\tilde{t}; \hat{\overline{V}}_0) = \infty$. Hence, there exists a constant $\tilde{C}_\sigma$ such that $\hat{\overline{V}}_0 > \tilde{C}_\sigma$ implies $V_\sigma(\tilde{t}; \hat{\overline{V}}_0) > s/\lambda_1 + (s/r)(1 - e^{-r(\check{T}(t) - \tilde{t})})$, and hence $t_{\sigma+1}^*(\hat{\overline{V}}_0) > \tilde{t}$. Iterated application of this argument yields the existence of a constant

$\tilde{C}_{i-1}$ ($i \in \{1,\dots,m\}$) such that $\overline{V}_0 > \tilde{C}_{i-1}$ implies $t_i^*(\overline{V}_0) > \tilde{t}$. As $\tilde{t} \in [t, \check{T}(t))$ was arbitrary, we conclude that $\lim_{\overline{V}_0 \to \infty} t_i^*(\overline{V}_0) = \check{T}(t)$, and that $\lim_{\overline{V}_0 \to \infty} V_i(\tilde{t}; \overline{V}_0) = \infty$ for any $\tilde{t} \in [t, \check{T}(t))$, $i \in \{1,\dots,m\}$.

For $\tilde{t} \geq t_i^*$, we have that $V_i(\tilde{t}; \overline{V}_0) = (s/r)(1 - e^{-r(\check{T}(t)-\tilde{t})}) \leq V_{i-1}(\tilde{t}; \overline{V}_0)$. It remains to be shown that for $\tilde{t} < t_i^*$, $V_i(\tilde{t}; \overline{V}_0) < V_{i-1}(\tilde{t}; \overline{V}_0)$. Because V_{i-1} is strictly decreasing, we have that

$$\begin{aligned} V_i(\tilde{t}; \overline{V}_0) &= \int_{\tilde{t}}^{t_i^*} e^{-(r+\lambda_1)(\tau-\tilde{t})} \lambda_1 V_{i-1}(\tau; \overline{V}_0)\, d\tau + e^{-(r+\lambda_1)(t_i^*-\tilde{t})} \frac{s}{r}(1 - e^{-r(\check{T}(t)-t_i^*)}) \\ &\leq \frac{\lambda_1}{\lambda_1 + r} V_{i-1}(\tilde{t}; \overline{V}_0)(1 - e^{-(r+\lambda_1)(t_i^*-\tilde{t})}) + e^{-(r+\lambda_1)(t_i^*-\tilde{t})} \frac{s}{r}(1 - e^{-r(\check{T}(t)-t_i^*)}). \end{aligned}$$

Now suppose that $V_i(\tilde{t}; \overline{V}_0) \geq V_{i-1}(\tilde{t}; \overline{V}_0)$. Then the above inequality implies that

$$\left(\frac{r}{r+\lambda_1} + \frac{\lambda_1}{r+\lambda_1} e^{-(r+\lambda_1)(t_i^*-\tilde{t})}\right) V_i(\tilde{t}; \overline{V}_0) \leq e^{-(r+\lambda_1)(t_i^*-\tilde{t})} \frac{s}{r}(1 - e^{-r(\check{T}(t)-t_i^*)}).$$

Yet as $V_i(\tilde{t}; \overline{V}_0) \geq (s/r)(1 - e^{-r(\check{T}(t)-\tilde{t})}) > (s/r)(1 - e^{-r(\check{T}(t)-t_i^*)})$, this implies that

$$\frac{r}{r+\lambda_1} + \frac{\lambda_1}{r+\lambda_1} e^{-(r+\lambda_1)(t_i^*-\tilde{t})} < e^{-(r+\lambda_1)(t_i^*-\tilde{t})},$$

a contradiction.

It remains to be shown that the functions V_i are continuous functions of $\overline{V}_0$. Here we will in fact show the slightly stronger statement that the functions V_i are jointly continuous in $(\tilde{t}, \overline{V}_0)$. For $i = 1$, this immediately follows from the explicit expression for V_1. By our explicit expression for V_i, all that remains to be shown is that t_i^* is a continuous function of $\overline{V}_0$. For $t_1^* = \check{T}(t)$, this is immediate. Before we are ready to do the appertaining induction step, we first make two preliminary observations.

First, it is the case that for all i, $\dot{V}_i(\tilde{t}; \overline{V}_0) < -se^{-r(\check{T}(t)-\tilde{t})}$ for $\tilde{t} < t_i^*$. Indeed, for $i = 1$, this is immediate. For $i > 1$, the induction step follows as *supra* by noting that if $\tilde{t} \in [t, t_i^*)$, we have that $t < t_i^* < t_{i-1}^*$. Second, this immediately implies that if $t_i^* > t$, the equation $V_{i-1}(\hat{t}; \overline{V}_0) - (s/r)(1 - e^{-r(\check{T}(t)-\hat{t})}) - s/\lambda_1 = 0$ has in fact $\hat{t} = t_i^*$ as its unique root.

Our induction hypothesis is that $t_{i-1}^*(\overline{V}_0)$ and $V_{i-1}(\tilde{t}; \overline{V}_0)$ are continuous. Let $\check{\overline{V}}_0 \in (s/\lambda_1, \infty)$ be arbitrary. I shall now argue that our induction hypothesis implies that $t_i^*(\overline{V}_0)$ (and hence V_i) is continuous at $\check{\overline{V}}_0$. To do so, it is convenient to define an auxiliary function $h(\overline{V}_0, \tilde{t}) := V_{i-1}(\tilde{t}; \overline{V}_0) - (s/r)(1 - e^{-r(\check{T}(t)-\tilde{t})}) - s/\lambda_1$; we note that h is continuous by induction hypothesis.

First, assume that $\check{\overline{V}}_0$ is such that $h(\check{\overline{V}}_0, t) < 0$. Because h is continuous, it follows that $h(\overline{V}_0, t) < 0$, and hence $t_i^*(\overline{V}_0) = t$, for all $\overline{V}_0$ in some neighborhood of $\check{\overline{V}}_0$.

Now let $h(\check{\overline{V}}_0, t) = 0$. Then $t_{i-1}^*(\check{\overline{V}}_0) > t = t_i^*(\check{\overline{V}}_0)$. We must show that for every $\tilde{\epsilon} > 0$ there exists a $\tilde{\delta} > 0$ such that for all $\overline{V}_0$ satisfying $|\check{\overline{V}}_0 - \overline{V}_0| < \tilde{\delta}$ we have that $|t_i^*(\overline{V}_0) - t| < \tilde{\epsilon}$. Fix an arbitrary $\tilde{\epsilon} \in (0, \check{T}(t) - t]$ (if $\tilde{\epsilon} > \check{T}(t) - t$, the statement trivially holds for all $\tilde{\delta} > 0$) and consider the date $\tilde{t} := t + \tilde{\epsilon}/2$. As $t_{i-1}^*(\check{\overline{V}}_0) > t$, we have

that $h(\check{\overline{V}}_0, \tilde{t}) < 0$. As $h(\cdot, \tilde{t})$ is continuous (by induction hypothesis) and, as we have shown, increasing in $\overline{V}_0$ with $\lim_{\overline{V}_0 \to \infty} h(\overline{V}_0, \tilde{t}) = \infty$, we know that there exists a $\tilde{\overline{V}}_0 > \check{\overline{V}}_0$ such that $h(\tilde{\overline{V}}_0, \tilde{t}) = 0$. Moreover, by monotonicity of $h(\cdot, \tilde{t})$, we have that $h(\overline{V}_0, \tilde{t}) \leq 0$ for all $\overline{V}_0 \leq \tilde{\overline{V}}_0$, and, hence, $t_i^*(\overline{V}_0) \leq \tilde{t} < t + \tilde{\epsilon}$. Defining $\tilde{\delta} := \tilde{\overline{V}}_0 - \check{\overline{V}}_0 > 0$ completes the step.

Finally, suppose that $h(\check{\overline{V}}_0, t) > 0$. In this case, $t_{i-1}^*(\check{\overline{V}}_0) > t_i^*(\check{\overline{V}}_0) > t$. Because t_{i-1}^* is continuous in $\overline{V}_0$ by our induction hypothesis, there exist $\tilde{\epsilon}, \tilde{\delta} > 0$ such that $t_i^*(\check{\overline{V}}_0) + \tilde{\epsilon} < t_{i-1}^*(\overline{V}_0)$ for all $\overline{V}_0 \in (\check{\overline{V}}_0 - \tilde{\delta}, \check{\overline{V}}_0 + \tilde{\delta})$. This implies that for any $\tilde{t} \in (t_i^*(\check{\overline{V}}_0) - \tilde{\epsilon}, t_i^*(\check{\overline{V}}_0) + \tilde{\epsilon})$ and any fixed $\overline{V}_0 \in (\check{\overline{V}}_0 - \tilde{\delta}, \check{\overline{V}}_0 + \tilde{\delta})$, we have that $\dot{V}_{i-1}(\tilde{t}; \overline{V}_0) < -se^{-r(\check{T}(t)-\tilde{t})}$ and, hence, $(\partial h/\partial \tilde{t})(\overline{V}_0, \tilde{t}) < 0$. (We have shown above that $V_{i-1}(\cdot; \overline{V}_0)$ and, hence, $h(\overline{V}_0, \cdot)$, is C^1.) By the implicit function theorem,[22] continuity of $t_i^*(\overline{V}_0)$ at $\check{\overline{V}}_0$ now follows from the fact that $t_i^*(\overline{V}_0)$ is defined by $h(\overline{V}_0, t_i^*(\overline{V}_0)) = 0$.

Proof of Lemma A.2 That $f(t; \overline{V}_0) = 0$ immediately follows from the fact that $V_m(\check{T}(t); \overline{V}_0) = 0$ for any $\check{T}(t) \in [t, \overline{T})$. Strict monotonicity of $V_i(\tilde{t}; \overline{V}_0; \check{T}(t))$ $(i = 1, \ldots, m)$ in $\check{T}(t)$ is immediately implied by the observation that for any fixed $\tilde{t} \leq \check{T}_1$ and $\overline{V}_0 > s/\lambda_1$, and given the end date $\check{T}_2 > \check{T}_1$, the agent can always guarantee himself a payoff of $V_i(\tilde{t}; \overline{V}_0; \check{T}_1) + (s/r)e^{-r(\check{T}_1 - \tilde{t})}(1 - e^{-r(\check{T}_2 - \check{T}_1)}) > V_i(\tilde{t}; \overline{V}_0; \check{T}_1)$ by following the strategy that was optimal for the end date $\check{T}_1$ in the time interval $[t, \check{T}_1]$ and playing safe for sure on $(\check{T}_1, \check{T}_2]$. As $f(\cdot, \overline{V}_0) = V_m(t; \overline{V}_0; \cdot)$, this shows the monotonicity property of f that we claimed.

It remains to prove continuity of $f(\cdot; \overline{V}_0)$. By Lemma A.1, we have that

$$f(\check{T}(t), \overline{V}_0) = \begin{cases} \int_t^{t_m^*} e^{-(r+\lambda_1)(\tau - t)} \lambda_1 V_{m-1}(\tau; \overline{V}_0; \check{T}(t))\, d\tau + e^{-(r+\lambda_1)(t_m^* - t)} \frac{s}{r}(1 - e^{-r(\check{T}(t) - t_m^*)}) \\ \qquad\qquad \text{if } t < t_m^* \\ \frac{s}{r}(1 - e^{-r(\check{T}(t) - t)}) \quad \text{if } t = t_m^* \end{cases}$$

and that

$$V_i(\tilde{t}; \overline{V}_0; \check{T}(t)) = \begin{cases} \int_{\tilde{t}}^{t_i^*} e^{-(r+\lambda_1)(\tau - \tilde{t})} \lambda_1 V_{i-1}(\tau; \overline{V}_0; \check{T}(t))\, d\tau + e^{-(r+\lambda_1)(t_i^* - \tilde{t})} \frac{s}{r}(1 - e^{-r(\check{T}(t) - t_i^*)}) \\ \qquad\qquad \text{if } \tilde{t} \leq t_i^* \\ \frac{s}{r}(1 - e^{-r(\check{T}(t) - \tilde{t})}) \quad \text{if } \tilde{t} > t_i^* \end{cases}$$

for all $i = 1, \ldots, m$, and $\tilde{t} \in [t, \check{T}(t)]$. Moreover, we have that

$$V_1(\tilde{t}; \overline{V}_0; \check{T}(t)) = \frac{\lambda_1}{\lambda_1 + r}(1 - e^{-(r+\lambda_1)(\check{T}(t) - \tilde{t})})\left(\overline{V}_0 + \frac{s}{r}\right) - \frac{s}{r} e^{-r(\check{T}(t) - \tilde{t})}(1 - e^{-\lambda_1(\check{T}(t) - \tilde{t})}),$$

i.e., for any given $\overline{V}_0$, V_1 is jointly continuous in $(\tilde{t}, \check{T}(t))$; moreover, $t_1^*(\check{T}(t)) = \check{T}(t)$ is trivially continuous in $\check{T}(t)$.

[22] Most versions of the implicit function theorem would require $V_{i-1}(\tilde{t}; \overline{V}_0)$ to be C^1 rather than just C^0. However, there are nondifferentiable versions of the theorem; here, one can, for instance, use the version in Kudryavtsev (2001).

The rest of the proof closely follows our proof of the continuity of $V_i(\tilde{t};\overline{V}_0)$ in $\overline{V}_0$ in Lemma A.1. In particular, our induction hypothesis is that $t^*_{i-1}(\check{T}(t))$ and $V_{i-1}(\tilde{t};\overline{V}_0;\check{T}(t))$ are continuous (for a given fixed $\overline{V}_0$). Let $\check{T}^* \in [t,\overline{T})$ be arbitrary. I shall now argue that our induction hypothesis implies that $t^*_i(\check{T}(t))$ is continuous at $\check{T}^*$; by our explicit expression for V_i, this implies that V_i is continuous in $(\tilde{t},\check{T}(t))$ for given $\overline{V}_0$. Again, we define an auxiliary function $\check{h}(\check{T},\tilde{t}) := V_{i-1}(\tilde{t};\overline{V}_0;\check{T}) - (s/r)(1 - e^{-r(\check{T}-\tilde{t})}) - s/\lambda_1$. We recall from our proof of Lemma A.1 that $t^*_i(\check{T})$ is implicitly defined by $\check{h}(\check{T},t^*_i(\check{T})) = 0$ if $\check{h}(\check{T},t) \geq 0$; otherwise, $t^*_i(\check{T}) = t$. We note that $\check{h}$ is continuous by induction hypothesis; we furthermore know that $\check{h}$ is decreasing in $\tilde{t}$ and strictly decreasing if $\tilde{t} < t^*_{i-1}(\check{T})$. By our argument at the beginning of this proof, we also know that as we increase $\check{T}$ to some arbitrary $\check{T}' > \check{T}$, V_{i-1} at $\tilde{t} \leq \check{T}$ increases by at least $(s/r)e^{-r(\check{T}-\tilde{t})}(1 - e^{-r(\check{T}'-\check{T})})$. Hence, we can conclude that $\check{h}(\cdot,\tilde{t})$ is weakly increasing.

First, assume that $\check{T}^*$ is such that $\check{h}(\check{T}^*,t) < 0$. Because $\check{h}$ is continuous, it follows that $\check{h}(\check{T},t) < 0$ and, hence, $t^*_i(\check{T}) = t$, for all $\check{T}$ in some neighborhood of $\check{T}^*$.

Now assume that $\check{h}(\check{T}^*,t) = 0$. This implies that $\check{T}^* \geq t^*_{i-1}(\check{T}^*) > t = t^*_i(\check{T}^*)$. We must show that for every $\epsilon > 0$ there exists a $\delta > 0$ such that $|\check{T} - \check{T}^*| < \delta$ implies $|t^*_i(\check{T}) - t| < \epsilon$. Fix an arbitrary $\epsilon > 0$ and consider the date $\tilde{t} = t + \kappa\epsilon$, with $\kappa \in (0,1)$ being chosen so that $\tilde{t} < \check{T}^*$. As $t_{i-1}(\check{T}^*) > t$, we have that $\check{h}(\check{T}^*,\tilde{t}) < 0$. Now suppose there exists a $\check{T}^{**} \in (\check{T}^*,\overline{T})$ such that $\check{h}(\check{T}^{**},\tilde{t}) = 0$. Because $\check{h}(\cdot,\tilde{t})$ is increasing, this implies that for all $\check{T} \in [t,\check{T}^{**}]$, we have that $t^*_i(\check{T}) \leq \tilde{t} < t + \epsilon$. In this case, setting $\delta = \check{T}^{**} - \check{T}^* > 0$ does the job. However, it could also be the case that $\check{h}(\check{T},\tilde{t}) < 0$ for all $\check{T} \in [\check{T}^*,\overline{T})$. In this case, $t^*_i(\check{T}) < \tilde{t} < t + \epsilon$ for all $\check{T} \in [t,\overline{T})$. Hence, any $\delta > 0$, for instance $\delta = \frac{1}{2}(\overline{T} - \check{T}^*)$, will do.

Finally, suppose that $\check{h}(\check{T}^*,t) > 0$. In this case, $t^*_{i-1}(\check{T}^*) > t^*_i(\check{T}^*) > t$. Because t^*_{i-1} is continuous in $\check{T}$ by our induction hypothesis, there exist $\tilde{\epsilon},\tilde{\delta} > 0$ such that $t^*_i(\check{T}^*) + \tilde{\epsilon} < t^*_{i-1}(\check{T})$ for all $\check{T} \in (\check{T}^* - \tilde{\delta},\check{T}^* + \tilde{\delta})$. This implies that for any $\tilde{t} \in (t^*_i(\check{T}^*) - \tilde{\epsilon}, t^*_i(\check{T}^*) + \tilde{\epsilon})$ and any fixed $\check{T} \in (\check{T}^* - \tilde{\delta},\check{T}^* + \tilde{\delta})$, we have that $\dot{V}_{i-1}(\tilde{t};\overline{V}_0;\check{T}) < -se^{-r(\check{T}-\tilde{t})}$ and, hence, $(\partial\check{h}/\partial\tilde{t})(\check{T},\tilde{t}) < 0$. As $\check{T}^*$ is an interior point (as $\check{h}(t,t) = -s/\lambda_1 < 0$), we can again apply the implicit function theorem to conclude that $t^*_i(\check{T})$ is continuous at $\check{T}^*$ because $t^*_i(\check{T})$ is defined by $\check{h}(\check{T},t^*_i(\check{T})) = 0$.

Thus, we have shown that for all $i = 1,\dots,m$, $t^*_i(\check{T})$ is continuous and, hence, $V_i(\tilde{t};\overline{V}_0;\check{T})$ is jointly continuous in $(\tilde{t},\check{T})$. In particular, this implies that $f(\check{T}(t);\overline{V}_0) = V_m(t;\overline{V}_0;\check{T}(t))$ is continuous in $\check{T}(t)$.

Proof of Proposition A.1 By Lemma A.1, we know that $V_m(t;\cdot)$ is continuous and (weakly) increasing; moreover, we know that there exists a constant C_m such that $\overline{V}_0 > C_m$ implies that $V_m(t;\cdot)$ is strictly increasing, with $\lim_{\overline{V}_0\to\infty} V_m(t;\overline{V}_0) = \infty$. Hence, statement (i) follows.

With respect to statement (ii), we first choose some lump sum $\hat{\overline{V}}_0 > s/\lambda_1$ such that $w_t < f(\check{T}(t);\hat{\overline{V}}_0)$. (The existence of such a $\hat{\overline{V}}_0$ is immediate by an argument analogous to the above.) Continuity and monotonicity of f (see Lemma A.2) now immediately imply the existence of some $\check{\check{T}}(t) \in (t,\check{T}(t))$ such that $w_t = f(\check{\check{T}}(t);\hat{\overline{V}}_0)$.

Appendix B: Pontryagin's conditions for the agent's problem

Neglecting a constant factor, the Hamiltonian $\mathfrak{H}_t$ for the agent's problem is given by

$$\begin{aligned}\mathfrak{H}_t = {} & e^{-rt} y_t \big[(1-k_{0,t}-k_{1,t})s + k_{0,t}\lambda_0(\phi_t+\omega_t(x_t))\big] \\ & + y_t e^{-rt-x_t}\big[(1-k_{0,t}-k_{1,t})s + k_{0,t}\lambda_0(\phi_t+\omega_t(x_t)) + k_{1,t}\lambda_1(\phi_t+w_t)\big] \\ & + \mu_t\lambda_1 k_{1,t} - \gamma_t\lambda_0 k_{0,t} y_t.\end{aligned}$$

By the maximum principle,[23] the existence of absolutely continuous functions μ_t and γ_t respectively satisfying the equations (13) and (14) a.e., as well as (15), which has to be satisfied for a.a. t, together with the transversality conditions $\gamma_T = \mu_T = 0$, are necessary for the agent's behaving optimally by setting $k_{1,t} = 1$ at any time t:[24]

$$\begin{aligned}\dot{\mu}_t = {} & e^{-rt} y_t \big\{ e^{-x_t}\big[(1-k_{0,t}-k_{1,t})s + k_{0,t}\lambda_0(\phi_t+\omega_t(x_t)) + k_{1,t}\lambda_1(\phi_t+w_t)\big] \\ & - k_{0,t}\lambda_0(1+e^{-x_t})\omega_t'(x_t)\big\}\end{aligned} \tag{13}$$

$$\begin{aligned}\dot{\gamma}_t = {} & -e^{-rt}\big\{\big[(1-k_{0,t}-k_{1,t})s + k_{0,t}\lambda_0(\phi_t+\omega_t(x_t))\big] \\ & + e^{-x_t}\big[(1-k_{0,t}-k_{1,t})s + k_{0,t}\lambda_0(\phi_t+\omega_t(x_t)) + k_{1,t}\lambda_1(\phi_t+w_t)\big]\big\} \\ & + \gamma_t\lambda_0 k_{0,t}\end{aligned} \tag{14}$$

$$\begin{aligned}& e^{-rt} y_t[e^{-x_t}\lambda_1(\phi_t+w_t) - (1+e^{-x_t})s] + \mu_t\lambda_1 \\ & \qquad \geq \max\big\{0, e^{-rt} y_t(1+e^{-x_t})\big[\lambda_0(\phi_t+\omega_t(x_t)) - s\big] - \gamma_t\lambda_0 y_t\big\}.\end{aligned} \tag{15}$$

Now, setting $k_{1,t} = 1$ at a.a. times t implies $x_t = x_0 + \lambda_1 t$ and $y_t = 1$ for all t. Thus, we can rewrite (13) and (14) as

$$\dot{\mu}_t = -\dot{\gamma}_t = e^{-rt-x_t}\lambda_1(\phi_t+w_t),$$

which is (1) in the main text. Furthermore, we can rewrite (15) as the two joint conditions

$$e^{-rt}[e^{-x_t}\lambda_1(\phi_t+w_t) - (1+e^{-x_t})s] \geq -\mu_t\lambda_1$$

and

$$e^{-rt}\big[e^{-x_t}\lambda_1(\phi_t+w_t) - (1+e^{-x_t})\lambda_0(\phi_t+\omega_t(x_t))\big] \geq -\mu_t(\lambda_1-\lambda_0),$$

which are (2) and (3) in the main text.

[23] See Theorem 2 in Seierstad and Sydster (1987, p. 85). One verifies that the relaxed regularity conditions in footnote 9, p. 132, are satisfied by observing that $\omega_t(\hat{p})$ is convex in $\hat{p}$, hence continuous for $\hat{p} \in (0,1)$. As $x = \ln[(1-\hat{p})/\hat{p}]$ is a continuous one-to-one transformation of $\hat{p}$, the relevant continuity requirements in Seierstad and Sydster (1987, footnote 9, p. 132) are satisfied.

[24] By standard arguments, the value function $\omega_t(\hat{p})$ is convex given any t; hence, it admits left and right derivatives with respect to $\hat{p}$ anywhere and is differentiable a.e. Because x is a differentiable transformation of $\hat{p}$, $\omega_t'(x)$ exists as a proper derivative for a.a. x. If x_t is one of those (countably many) points x at which it does not, $\omega_t'(x_t)$ is to be understood as the right derivative (because x_t can only ever increase over time).

Appendix C: Other proofs

Proof of Lemma 1

Fix an arbitrary $\check{T}(t) \in (t, \overline{T})$, $\tilde{t} \in (t, \check{T}(t)]$, $\hat{p}_{\tilde{t}} \in [p_{\tilde{t}}, p_0]$, and $\overline{V}_0 > 0$. Consider the restricted problem in which the agent can only choose between arms 0 and 1. Then the agent's time-$\tilde{t}$ expected reward is given by

$$\int_{\tilde{t}}^{\check{T}(t)} e^{-r(\tau_m - \tilde{t})} \left(\overline{V}_0 + \frac{s}{r}(1 - e^{-r(\check{T}(t) - \tau_m)}) \right) dF,$$

where F is the distribution over τ_m, the time of the mth breakthrough after time $\tilde{t}$. As the integrand is decreasing in τ_m, all that remains to be shown is that $F^*(\cdot; \hat{p}_{\tilde{t}})$ (where $F^*(\tau; \hat{p}_{\tilde{t}})$ denotes the probability of m breakthroughs up to time $\tau \in (\tilde{t}, \check{T}(t)]$ when the agent always pulls arm 1) is first-order stochastically dominated by the distribution of the mth breakthrough for *any* alternative strategy, which I shall denote by $\tilde{F}(\cdot; \hat{p}_{\tilde{t}})$. Fix an arbitrary $\tau \in (\tilde{t}, \check{T}(t)]$. Now

$$F^*(\tau; \hat{p}_{\tilde{t}}) = \hat{p}_{\tilde{t}} \frac{\lambda_1^m}{m!} (\tau - \tilde{t})^m e^{-\lambda_1(\tau - \tilde{t})}.$$

Whatever the alternative strategy under consideration may be, $\tilde{F}$ can be written as

$$\tilde{F}(\tau; \hat{p}_{\tilde{t}}) = \int_0^1 F_\alpha(\tau; \hat{p}_{\tilde{t}}) \mu(d\alpha),$$

with

$$F_\alpha(\tau; \hat{p}_{\tilde{t}}) = \hat{p}_{\tilde{t}} \frac{[\alpha\lambda_1 + (1-\alpha)\lambda_0]^m}{m!} (\tau - \tilde{t})^m e^{-(\alpha\lambda_1 + (1-\alpha)\lambda_0)(\tau - \tilde{t})} + (1 - \hat{p}_{\tilde{t}}) \frac{[(1-\alpha)\lambda_0]^m}{m!} (\tau - \tilde{t})^m e^{-(1-\alpha)\lambda_0(\tau - \tilde{t})}$$

for some probability measure μ on $\alpha \in [0, 1]$. The weight α can be interpreted as the fraction of the time interval $[\tilde{t}, \tau]$ devoted to arm 1; of course, because the agent's strategy allows him to condition his action on the entire previous history, α will generally be stochastic. Therefore, the strategy of the proof is to find an m such that, for any $\tilde{t} \in (t, \overline{T})$, $\tau \in (\tilde{t}, \overline{T})$ and $\hat{p}_{\tilde{t}} \in [p_{\overline{T}}, p_0]$, it is the case that

$$F^*(\tau; \hat{p}_{\tilde{t}}) > F_\alpha(\tau; \hat{p}_{\tilde{t}}) \tag{16}$$

uniformly for all $\alpha \in [0, 1)$.

To do so, I introduce the auxiliary function $\xi(q) := q^m e^{-q(\tau - \tilde{t})}$ (for $q \in [0, \lambda_1]$, $(\tau - \tilde{t}) \in (0, \overline{T}]$).[25] Note that ξ is (strictly) increasing and (strictly) convex if $(m - (\tau - \tilde{t})q)^2 > m > (\tau - \tilde{t})\lambda_1$ (for $q > 0$). Therefore, by choosing m such that

$$(m - \overline{T}\lambda_1)^2 > m > \overline{T}\lambda_1, \tag{17}$$

we can ensure that ξ is increasing and convex on its entire domain for any $(\tau - \tilde{t}) \in (0, \overline{T}]$.

[25] I am indebted to an anonymous referee for suggesting this argument, which replaces a more convoluted one in earlier versions.

Now, consider arbitrary $(\tau-\tilde{t})\in(0,\overline{T}]$ and $\alpha\in[0,1)$. By convexity of ξ, we have that

$$(\alpha\lambda_1+(1-\alpha)\lambda_0)^m e^{-(\alpha\lambda_1+(1-\alpha)\lambda_0)(\tau-\tilde{t})}\leq \alpha\lambda_1^m e^{-\lambda_1(\tau-\tilde{t})}+(1-\alpha)\lambda_0^m e^{-\lambda_0(\tau-\tilde{t})}$$

and

$$(\alpha 0+(1-\alpha)\lambda_0)^m e^{-(\alpha 0+(1-\alpha)\lambda_0)(\tau-\tilde{t})}\leq \alpha 0+(1-\alpha)\lambda_0^m e^{-\lambda_0(\tau-\tilde{t})}.$$

Now, adding $\hat{p}_{\tilde{t}}$ times the first inequality to $1-\hat{p}_{\tilde{t}}$ times the second inequality yields

$$\begin{aligned}\hat{p}_{\tilde{t}}(\alpha\lambda_1+(1-\alpha)\lambda_0)^m e^{-(\alpha\lambda_1+(1-\alpha)\lambda_0)(\tau-\tilde{t})}&+(1-\hat{p}_{\tilde{t}})((1-\alpha)\lambda_0)^m e^{-(1-\alpha)\lambda_0(\tau-\tilde{t})}\\ &\leq \hat{p}_{\tilde{t}}\alpha\lambda_1^m e^{-\lambda_1(\tau-\tilde{t})}+(1-\alpha)\lambda_0^m e^{-\lambda_0(\tau-\tilde{t})}\\ &=\hat{p}_{\tilde{t}}\lambda_1^m e^{-\lambda_1(\tau-\tilde{t})}-(1-\alpha)\lambda_1^m e^{-\lambda_1(\tau-\tilde{t})}\left[\hat{p}_{\tilde{t}}-\left(\frac{\lambda_0}{\lambda_1}\right)^m e^{(\lambda_1-\lambda_0)(\tau-\tilde{t})}\right].\end{aligned}$$

Therefore, by choosing m large enough so that

$$p_{\overline{T}}\left(\frac{\lambda_1}{\lambda_0}\right)^m > e^{(\lambda_1-\lambda_0)\overline{T}}, \tag{18}$$

we can ensure that the expression in square brackets is strictly positive, so that because $\alpha<1$, we have

$$\begin{aligned}\hat{p}_{\tilde{t}}(\alpha\lambda_1+(1-\alpha)\lambda_0)^m e^{-(\alpha\lambda_1+(1-\alpha)\lambda_0)(\tau-\tilde{t})}&+(1-\hat{p}_{\tilde{t}})((1-\alpha)\lambda_0)^m e^{-(1-\alpha)\lambda_0(\tau-\tilde{t})}\\ &<\hat{p}_{\tilde{t}}\lambda_1^m e^{-\lambda_1(\tau-\tilde{t})}.\end{aligned}$$

Thus, we choose an $m\in\mathbb{N}\cap[2,\infty)$ large enough so that both (17) and (18) are satisfied. Note that the choice of m is independent of α, $\tilde{t}$, $\tau>\tilde{t}$, $\check{T}(t)$, and $\hat{p}_{\tilde{t}}\in[p_{\overline{T}},p_0]$. Choosing m in this manner ensures that

$$F^*(\tau;\hat{p}_{\tilde{t}})>F_\alpha(\tau;\hat{p}_{\tilde{t}})$$

for all $\alpha\in[0,1)$. Hence, for such an m, it is the case that for any $\tilde{t}$, $\tau>\tilde{t}$, and $\hat{p}_{\tilde{t}}\in[p_{\overline{T}},p_0]$, we have that $F^*(\tau;\hat{p}_{\tilde{t}})>\tilde{F}(\tau;\hat{p}_{\tilde{t}})$ for any $\tau>\tilde{t}$ whenever $\mu\neq\delta_1$, where δ_1 denotes the Dirac measure associated with the strategy of always pulling arm 1, whatever may befall.

It remains to be shown that the preference ordering does not change if the agent also has access to the safe arm. In this case, his goal is to maximize

$$\begin{aligned}\int_{\tilde{t}}^{\check{T}(t)}\Bigg\{&(1-k_\tau)e^{-r(\tau-\tilde{t})}s\\ &+\int_{\tilde{t}}^{\check{T}(t)}e^{-r(\tau_m-\tilde{t})}\left(\overline{V}_0+\frac{s}{r}(1-e^{-r(\check{T}(t)-\tau_m)})\right)d\tilde{F}^{\{k_\tau\}}(\tau_m;\hat{p}_{\tilde{t}})\Bigg\}\,d\nu(\{k_\tau\}_{\tilde{t}\leq\tau\leq\check{T}(t)})\end{aligned}$$

over probability measures $\tilde{F}^{\{k_\tau\}}$ and ν, with the process $\{k_\tau\}$ satisfying $0\leq k_\tau\leq 1$ for all $\tau\in[\tilde{t},\check{T}(t)]$.

I now show that for any such process $\{k_\tau\}$ and $\hat{p}_{\tilde{t}} \in [p_{\overline{T}}, p_0]$, it is the case that if $\int_{\tilde{t}}^{\check{T}(t)} k_\sigma \, d\sigma = 0$, then $\tilde{F}^{\{k_\tau\}}(\cdot; \hat{p}_{\tilde{t}}) = 0$, and if $\int_{\tilde{t}}^{\check{T}(t)} k_\sigma \, d\sigma > 0$, then $(\tilde{F}^{\{k_\tau\}})^*$, the distribution over the mth breakthrough that ensues from the agent's never using arm 0, is first-order stochastically dominated by all other distributions $\tilde{F}^{\{k_\tau\}} \neq (\tilde{F}^{\{k_\tau\}})^*$. Arguing as above, we can write

$$\tilde{F}^{\{k_\tau\}}(\tau; \hat{p}_{\tilde{t}}) = \int_0^1 F_\alpha^{\{k_\tau\}}(\tau; \hat{p}_{\tilde{t}}) \mu(d\alpha)$$

for

$$F_\alpha^{\{k_\tau\}}(\tau; \hat{p}_{\tilde{t}}) = \hat{p}_{\tilde{t}} \frac{[\alpha\lambda_1 + (1-\alpha)\lambda_0]^m}{m!} \left(\int_{\tilde{t}}^{\tau} k_\sigma \, d\sigma \right)^m e^{-(\alpha\lambda_1 + (1-\alpha)\lambda_0) \int_{\tilde{t}}^{\tau} k_\sigma \, d\sigma} + (1 - \hat{p}_{\tilde{t}}) \frac{[(1-\alpha)\lambda_0]^m}{m!} \left(\int_{\tilde{t}}^{\tau} k_\sigma \, d\sigma \right)^m e^{-(1-\alpha)\lambda_0 \int_{\tilde{t}}^{\tau} k_\sigma \, d\sigma}$$

and some probability measure μ. Because all that changes with respect to our calculations above is for $\tau - \tilde{t} > 0$ to be replaced by $\int_{\tilde{t}}^{\tau} k_\sigma \, d\sigma \in [0, \tau - \tilde{t}]$, and our previous τ was arbitrary, the previous calculations continue to apply if $\int_{\tilde{t}}^{\tau} k_\sigma \, d\sigma > 0$. (Otherwise, $\tilde{F}^{\{k_\tau\}} = 0$ for all measures μ.) In particular, any $m \geq 2$ satisfying conditions (17) and (18) ensures that if $\int_{\tilde{t}}^{\tau} k_\sigma \, d\sigma > 0$, $(\tilde{F}^{\{k_\tau\}})^*$ is first-order stochastically dominated by any $\tilde{F}^{\{k_\tau\}} \neq (\tilde{F}^{\{k_\tau\}})^*$. As $e^{-r(\tau_m - \tilde{t})}(\overline{V}_0 + (s/r)(1 - e^{-r(\check{T}(t) - \tau_m)}))$ is decreasing in τ_m, we can conclude that setting $\alpha = 1$ with probability 1 is (strictly) optimal for all $\{k_\sigma\}_{\tilde{t} \leq \sigma \leq \check{T}(t)}$ (with $\int_{\tilde{t}}^{\check{T}(t)} k_\sigma \, d\sigma > 0$).

Proof of Lemma 2

Because m is constant over time, piecewise continuity of $\check{T}(t)$ and of the lump-sum reward $\overline{V}_0(t)$ (as a function of the date of the first breakthrough t) imply the piecewise continuity in t of the value $\omega_t(x)$. As $\omega_t(x)$ is furthermore continuous in x (see footnote 23), the regularity conditions required for Filippov–Cesari's existence theorem (Theorem 8 in Seierstad and Sydster 1987, p. 132) are satisfied.[26]

Clearly, $\check{\mathcal{U}} := \{(a, b) \in \mathbb{R}_+^2 : a + b \leq 1\}$ is closed, bounded, and convex, the set of admissible policies is nonempty, and the state variables are bounded. Using in addition the linearity of the objective and the laws of motion in the choice variables (see Appendix B), one can show that the conditions of Filippov–Cesari's theorem are satisfied.

Proof of Proposition 3

Suppose that in addition to the path implied by $k_{1,t} = 1$ for all t, there is an alternative path $(\hat{k}_{0,t}, \hat{k}_{1,t})_{0 \leq t \leq T}$, with $\hat{k}_{1,t} \neq 1$ on a set of positive measure, that satisfies Pontryagin's conditions. I denote the associated state and co-state variables by $\check{x}_t$, $\check{y}_t$, $\check{\mu}_t$, $\check{\gamma}_t$ for the former and $\hat{x}_t$, $\hat{y}_t$, $\hat{\mu}_t$, $\hat{\gamma}_t$ for the latter path. Moreover, I define $\hat{t} := \sup\{t \in \bigcup_i (t_i^\dagger, t_i^\ddagger) : t_i^\dagger <$

[26] See Note 17 in Seierstad and Sydster (1987, p. 133) for a statement of the regularity conditions.

$t_i^{\ddagger}$ and $\hat{k}_{1,\tau} \neq 1$ for a.a. $\tau \in (t_i^{\dagger}, t_i^{\ddagger})\}$. Because $\hat{k}_{1,t} \neq 1$ on a set of positive measure, we have that $\hat{t} > 0$.

By (13) and the transversality condition $\hat{\mu}_T = \check{\mu}_T = 0$, we have that $(e^{\hat{x}_{\hat{t}}}\hat{\mu}_{\hat{t}})/\hat{y}_{\hat{t}} = e^{\check{x}_{\hat{t}}}\check{\mu}_{\hat{t}}$; moreover, we know that by Pontryagin's principle, the mappings $t \mapsto (e^{\hat{x}_t}\hat{\mu}_t)/\hat{y}_t$ and $t \mapsto e^{\check{x}_t}\check{\mu}_t$ are continuous. Now consider an $\eta > 0$ such that $\hat{k}_{1,\tau} \neq 1$ for a.a. $\tau \in (\hat{t} - \eta, \hat{t})$. (Such an $\eta > 0$ exists because $\hat{k}_{1,t} \neq 1$ on a set of positive measure.) Then we have that

$$\lambda_1(\phi_t + w_t) - (1 + e^{\hat{x}_t})s > \lambda_1(\phi_t + w_t) - (1 + e^{\check{x}_t})s$$

for all $t \in [\hat{t} - \frac{1}{2}\eta, \hat{t}]$ because $\hat{x}_t < \check{x}_t$ there. Moreover, because $k_{1,t} = 1$ for all t satisfies Pontryagin's conditions, we have that

$$\lambda_1(\phi_t + w_t) - (1 + e^{\check{x}_t})s \geq -\lambda_1 e^{\check{x}_t + rt}\check{\mu}_t$$

for a.a. $t \in [0, T]$, and, hence, by continuity,

$$\lambda_1(\phi_{\hat{t}} + w_{\hat{t}}) - (1 + e^{\hat{x}_{\hat{t}}})s > \lambda_1(\phi_{\hat{t}} + w_{\hat{t}}) - (1 + e^{\check{x}_{\hat{t}}})s \geq -\lambda_1 e^{\check{x}_{\hat{t}} + r\hat{t}}\check{\mu}_{\hat{t}} = -\lambda_1 \frac{e^{\hat{x}_{\hat{t}} + r\hat{t}}}{\hat{y}_{\hat{t}}}\hat{\mu}_{\hat{t}}. \quad (19)$$

The path $(\hat{k}_{0,t}, \hat{k}_{1,t})_{0 \leq t \leq T}$ satisfies Pontryagin's necessary conditions. Thus, in particular, $(\hat{k}_{0,t}, \hat{k}_{1,t})$ maximizes the Hamiltonian $\mathfrak{H}_t$ at a.a. $t \in [0, T]$. This implies that $\hat{k}_{0,t} + \hat{k}_{1,t} = 1$ at a.a. t at which we have

$$\lambda_1(\phi_t + w_t) - (1 + e^{\hat{x}_t})s > -\lambda_1 \frac{e^{\hat{x}_t + rt}}{\hat{y}_t}\hat{\mu}_t.$$

We can therefore conclude that $\hat{k}_{0,\tau} + \hat{k}_{1,\tau} = 1$ for a.a. τ in some left neighborhood of $\hat{t}$, as both sides of inequality (19) are continuous.

Furthermore, by conditions (13) and (14) and the transversality condition $\check{\mu}_T = \hat{\mu}_T = \check{\gamma}_T = \hat{\gamma}_T = 0$, we have that $-\lambda_0 e^{\hat{x}_{\hat{t}}}\hat{\gamma}_{\hat{t}} - (\lambda_1 e^{\hat{x}_{\hat{t}}}\hat{\mu}_{\hat{t}})/\hat{y}_{\hat{t}} = -(\lambda_1 - \lambda_0)e^{\check{x}_{\hat{t}}}\check{\mu}_{\hat{t}}$. Again, by Pontryagin's conditions, the mapping $t \mapsto -\lambda_0 e^{\hat{x}_t}\hat{\gamma}_t - (\lambda_1 e^{\hat{x}_t}\hat{\mu}_t)/\hat{y}_t$ is continuous. Moreover, we have that

$$\begin{aligned}
&\lambda_1(\phi_t + w_t) - \lambda_0(1 + e^{\hat{x}_t})(\phi_t + \omega_t(\hat{x}_t)) \\
&\qquad \geq \lambda_1(\phi_t + w_t) - \lambda_0\left[w_t + (1 + e^{\hat{x}_t})\left(\phi_t + \frac{s}{r}(1 - e^{-r\epsilon_t})\right)\right] \\
&\qquad > \lambda_1(\phi_t + w_t) - \lambda_0\left[w_t + (1 + e^{\check{x}_t})\left(\phi_t + \frac{s}{r}(1 - e^{-r\epsilon_t})\right)\right]
\end{aligned}$$

for all $t \in [\hat{t} - \frac{1}{2}\eta, \hat{t}]$, with the first inequality being implied by Proposition 2. Moreover, by continuity and the fact that $k_{1,t} = 1$ for all $t \in [0, T]$ satisfies Pontryagin's necessary conditions, we have that $\phi_t + w_t \geq (s/\lambda_1)(1 + e^{x_0}) > 0$ and, hence, $\phi_t + (s/r)(1 - e^{-r\epsilon_t}) > 0$ for all $t \in [0, T]$. Hence, because $\hat{x}_t < \check{x}_t$, the second inequality also holds for all $t \in [\hat{t} - \frac{1}{2}\eta, \hat{t}]$. The fact that $k_{1,t} = 1$ for all $t \in [0, T]$ satisfies Pontryagin's conditions even for the upper bound on ω_t given by Proposition 2 furthermore implies that

$$\lambda_1(\phi_t + w_t) - \lambda_0\left[w_t + (1 + e^{\check{x}_t})\left(\phi_t + \frac{s}{r}(1 - e^{-r\epsilon_t})\right)\right] \geq -(\lambda_1 - \lambda_0)e^{\check{x}_t + rt}\check{\mu}_t$$

for a.a. $t \in [0, T]$, and, hence, by continuity,

$$\lambda_1(\phi_{\hat{t}} + w_{\hat{t}}) - \lambda_0\left[w_{\hat{t}} + (1 + e^{\check{x}_{\hat{t}}})\left(\phi_{\hat{t}} + \frac{s}{r}(1 - e^{-r\epsilon_{\hat{t}}})\right)\right] \geq -(\lambda_1 - \lambda_0)e^{\check{x}_{\hat{t}} + r\hat{t}}\check{\mu}_{\hat{t}}.$$

This implies that

$$\begin{aligned}\lambda_1(\phi_{\hat{t}} + w_{\hat{t}}) - \lambda_0(1 + e^{\hat{x}_{\hat{t}}})(\phi_{\hat{t}} + \omega_{\hat{t}}(\hat{x}_{\hat{t}}))& \\ \geq \lambda_1(\phi_{\hat{t}} + w_{\hat{t}}) - \lambda_0\left[w_{\hat{t}} + (1 + e^{\hat{x}_{\hat{t}}})\left(\phi_{\hat{t}} + \frac{s}{r}(1 - e^{-r\epsilon_{\hat{t}}})\right)\right]& \\ > -(\lambda_1 - \lambda_0)e^{\check{x}_{\hat{t}} + r\hat{t}}\check{\mu}_{\hat{t}} = -e^{r\hat{t}}\left[\lambda_0 e^{\hat{x}_{\hat{t}}}\hat{\gamma}_{\hat{t}} + \lambda_1\frac{e^{\hat{x}_{\hat{t}}}}{\hat{y}_{\hat{t}}}\hat{\mu}_{\hat{t}}\right].&\end{aligned}$$

Because $(\hat{k}_{0,t}, \hat{k}_{1,t})_{0\leq t\leq T}$ satisfies Pontryagin's necessary conditions, so in particular, $(\hat{k}_{0,t}, \hat{k}_{1,t})$ maximizes the Hamiltonian $\mathfrak{H}_t$ at a.a. $t \in [0, T]$, we can conclude by continuity that $\hat{k}_{0,\tau} = 0$ for a.a. τ in some left neighborhood of $\hat{t}$. Yet, because by our previous step $\hat{k}_{0,\tau} + \hat{k}_{1,\tau} = 1$ for a.a. τ in some left neighborhood of $\hat{t}$, we conclude that there exists some left neighborhood of $\hat{t}$ such that $\hat{k}_{1,\tau} = 1$ for a.a. τ in this left neighborhood, a contradiction to our definition of $\hat{t}$. We can thus conclude that there does not exist an alternative path $(\hat{k}_{0,t}, \hat{k}_{1,t})_{0\leq t\leq T}$, with $\hat{k}_{1,t} \neq 1$ on a set of positive measure, that satisfies Pontryagin's conditions.

Proof of Proposition 6

For $1/p^m$, the claim immediately follows from the explicit expressions for T^*,

$$T^* = \frac{1}{\lambda_1}\ln\left(\frac{-p_0 + \sqrt{p_0^2 + \frac{4}{p^m}p_0(1 - p_0)}}{2(1 - p_0)}\right),$$

and for the wedge

$$\frac{p_{T^*} - p^m}{p^m} = \frac{p_0 - 2 + \sqrt{p_0^2 + 4\frac{p_0}{p^m}(1 - p_0)}}{2(1 - p_0)}.$$

For p_0, one shows that the sign of $\partial T^*/\partial p_0$ is equal to the sign of

$$2(1 - p_0) + p_0 p^m - \sqrt{(p^m p_0)^2 + 4p^m p_0(1 - p_0)},$$

which is strictly positive if and only if

$$0 < (2(1 - p_0))^2.$$

This immediately implies that the wedge $(p_{T^*} - p^m)/p^m = e^{\lambda_1 T^*} - 1$ is increasing in p_0.

References

Arie, Guy (2014), "Dynamic costs and moral hazard: A duality based approach." Simon School Working Paper FR 12-11. [777]

Bergemann, Dirk and Ulrich Hege (1998), "Venture capital financing, moral hazard, and learning." *Journal of Banking & Finance*, 22, 703–735. [776, 791]

Bergemann, Dirk and Ulrich Hege (2005), "The financing of innovation: Learning and stopping." *RAND Journal of Economics*, 36, 719–752. [776, 791]

Bertsekas, Dimitri P. (1995), *Dynamic Programming and Optimal Control*, Vol. I. Athena Scientific, Belmont, Massachusetts. [797]

Bhaskar, Venkataraman (2012), "Dynamic moral hazard, learning and belief manipulation." CEPR Discussion Paper DP8948. [777]

Bonatti, Alessandro and Johannes Hörner (2011), "Collaborating." *American Economic Review*, 101, 632–663. [778, 789]

Cremer, Jacques and Richard P. McLean (1988), "Full extraction of the surplus in Bayesian and dominant strategy auctions." *Econometrica*, 56, 1247–1257. [775]

Fong, Kyna (2007), "Evaluating skilled experts: Optimal scoring rules for surgeons." SIEPR Discussion Paper 07-43. [777]

Francis, Bill B., Iftekhar Hasan, and Zenu Sharma (2011), "Incentives and innovation: Evidence from CEO compensation contracts." Bank of Finland Research Discussion Paper 17. [774]

Garfagnini, Umberto (2011), "Delegated experimentation." Unpublished paper. [778]

Guo, Yingni (forthcoming), "Dynamic delegation of experimentation." *American Economic Review*. [778]

Halac, Marina C., Qingmin Liu, and Navin Kartik (forthcoming), "Optimal contracts for experimentation." *Review of Economic Studies*. [777]

Holmstrom, Bengt and Paul Milgrom (1987), "Aggregation and linearity in the provision of intertemporal incentives." *Econometrica*, 55, 303–328. [776]

Holmstrom, Bengt and Paul Milgrom (1991), "Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design." *Journal of Law, Economics, & Organization*, 7, 24–52. [776]

Hörner, Johannes and Larry Samuelson (2013), "Incentives for experimenting agents." *RAND Journal of Economics*, 44, 632–663. [775, 776, 787, 791]

Keller, Godfrey, Sven Rady, and Martin Cripps (2005), "Strategic experimentation with exponential bandits." *Econometrica*, 73, 39–68. [778]

Kudryavtsev, Lev D. (2001), "Implicit function." In *Encyclopedia of Mathematics* (Michiel Hazewinkel, ed.), Springer, New York. Available at https://www.encyclopediaofmath.org/index.php/Implicit_function. [802]

Kwon, Suehyun (2015), "Dynamic moral hazard with persistent states." Unpublished paper, Department of Economics, University College London. [777]

Manso, Gustavo (2011), "Motivating innovation." *Journal of Finance*, 66, 1823–1860. [777]

Radner, Roy (1985), "Repeated principal–agent games with discounting." *Econometrica*, 53, 1173–1198. [775]

Rahman, David (2010), "Dynamic implementation." Unpublished paper. [795]

Seierstad, Alte and Knut Sydster (1987), *Optimal Control Theory With Economic Applications*. Elsevier Science, New York. [804, 807]

Co-editor George J. Mailath handled this manuscript.

Submitted 2014-7-3. Final version accepted 2015-8-13. Available online 2015-8-14.