The Importance of Being Honest*

Nicolas Klein†

This version: October 25, 2011

Abstract

This paper analyzes the case of a principal who wants to give an agent proper incentives to investigate a hypothesis which 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: Experimentation, Bandit Models, Poisson Process, Bayesian Learning, Principal-Agent Models, Optimal Incentive Scheme.

JEL Classification Numbers: C79, D82, D83, O32.

*I thank Johannes Hörner and Sven Rady for their advice, patience, and encouragement, as well as Dirk Bergemann, Tri-Vi Dang, Federico Echenique, Reinoud Joosten, Daniel Krähmer, Lucas Maestri, Thomas Mariotti, Benny Moldovanu, Pauli Murto, Frank Rosar, Dinah Rosenberg, Francisco Ruiz-Aliseda, Andy Skrzypacz, Eilon Solan, Bruno Strulovici, Juuso Välimäki, Tom Wiseman, Jianjun Wu, as well as seminar audiences at Berlin, Bonn, Paris, the 2011 North American Summer Meeting of the Econometric Society in St. Louis, the 2011 Canadian Economic Theory Conference in Vancouver, the 2011 Midwestern Economic Theory Conference at Notre Dame University, the 2010 International Conference on Game Theory at Stony Brook, and the 2010 Workshop on Stochastic Methods in Game Theory at Erice, 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. Financial support from the National Research Fund of Luxembourg and the German Research Fund through SFB TR-15 is gratefully acknowledged.

†University of Bonn, Lennéstr. 37, D-53113 Bonn, Germany; email: kleinnic@yahoo.com.

1 Introduction

Enormous amounts of money are being spent on the financing of scientific research. For instance, the National Cancer Institute spent $105 billion on the "War on Cancer" from 1971 to 2009. According to some, though, the grant system by which the research is to a large extent financed could be improved.[1] Some scientists bemoan that grant boards seem to favor low-risk, low-yield projects, at the expense of more promising, yet uncertain, prospects, it is reported. For instance, $100,000 over two years were spent on a study investigating whether people who were especially fond of good-tasting food had a harder time staying on a diet, while more fundamental proposals went unfunded.[2]

This paper proposes a stylized model of an alternative way of incentivizing innovation.[3] It leaves project selection to the scientist himself; the grant board only determines the length of funding, as well as the prizes a scientist can earn, as a function of the observable history. To capture the idea that scientists will typically be in a better position to ascertain the characteristics of a highly specialized research project, it is assumed that the principal (grant board) can only observe the occurrence of a success, such as e.g. a publication in a highly regarded peer-reviewed journal; yet, he cannot observe the characteristics of the project leading to the observed success.

Moreover, it is often the case that the true value of a scientific discovery can only be ascertained after a considerable amount of time has elapsed, an insight that has apparently altered the early practices of the Nobel Prize Committee, for instance. Whereas Alfred Nobel's will mandated that the prize be awarded for discoveries made "during the preceding year," several of the putative achievements recognized by the first Nobel awards were later discredited. In response, the committee moved toward recognizing discoveries that had withstood the test of time; Subrahmanyan Chandrasekhar e.g. shared the 1983 physics prize in recognition of discoveries made in the 1930s.[4] In order to capture this aspect of scientific investigation, I assume that the true quality of a breakthrough will only become obvious in the distant future, so that it will not be possible to condition the scientist's incentives on this future revelation.

[1] See e.g. the *New York Times* of June 28, 2009. I am indebted to Jianjun Wu for alerting me to these problems.

[2] The idea behind the study was that obesity is related to higher risk of cancer; hence, the discovery of better weight-management methods could potentially reduce the incidence of cancer; see the *New York Times* of June 28, 2009.

[3] This question has been addressed in the literature from a great variety of angles, see e.g. Holmström (1989) or Manso (2011).

[4] See e.g. The Titi Tudorancea Bulletin, http://www.tititudorancea.com/z/nobel_prize.htm (as of October, 13, 2011).

In particular, it is assumed that at any point in time, the scientist has a choice between two projects. One project's endeavor is merely to get publications out of old, established, knowledge; it yields apparent "successes," which are not socially valuable, according to a commonly known distribution. The other project, which is socially valuable, involves the investigation of a hypothesis which is uncertain. It is furthermore assumed that, being concerned with advancing scientific knowledge, the principal's interest in the matter is in finding out that the uncertain hypothesis is true; yet, when faced with an observable success such as a publication, for instance, the principal does not know, or cannot contract upon, whether it is old knowledge or whether it is a truly new discovery. Moreover, the agent could also shirk exerting effort, which gives him some private flow benefit, but in which case he will never achieve an observable success. The agent's effort choice is also unobservable to the principal. This paper shows how to implement honest investigation of the uncertain hypothesis, subject to the afore-mentioned informational restrictions. Specifically, the principal's objective is to minimize the wage costs of implementing honesty up to the first success with probability 1 on the equilibrium path. However, the principal only observes the occurrence, and timing, of successes; he does not observe whether a given success was a cheat or was achieved by honest means.

As is well known from the principal-agent literature, when his actions cannot easily be monitored, an agent's pay must be made contingent on his performance, so as to provide proper incentives for him to exert effort. Thus, the agent will get paid a substantial bonus if, and only if, he proves his hypothesis. While this may well provide him with the necessary incentives to work, unfortunately, it might also tempt him to try and fake a 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, Hasan and Sharma (2009). Using data from ExecuComp firms for the period 1992–2002, they show that the performance sensitivity of CEO pay has no impact on a firm's innovation performance, as measured by the number of patents taken out, or by the number of citations to patents.

In case even the investigation of a correct hypothesis yields breakthroughs at a lower frequency than manipulation, honesty is not implementable at all, i.e. it is impossible to get the scientist to pursue a low-yield high-risk project. In the more interesting case when the high-risk project is also the high-yield project, I show by what schemes the principal can make sure that the agent is always honest up to the first breakthrough at least. These optimal schemes all share the property that cheating is made so unattractive that it is dominated even by shirking. Hence, the agent only needs to be compensated for his forgone benefit of being lazy; put differently, the presence of a cheating action creates no distortions in players' values. Still, when the principal can additionally choose the end date of the interaction conditional on no breakthrough having occurred, he stops the project inefficiently early. The reason

for this is that future rewards adversely impact today's incentives: If the agent is paid a lot for achieving his first success tomorrow, he is loath to "risk" having his first success today, thereby forgoing the possibility of collecting tomorrow's reward. This distortion, however, could easily be overcome if the principal could hire different agents sequentially, as a means of counteracting the dynamic allure of future rewards. If agents could be hired for a mere instant, then in the limit the principal would end the project at the first-best optimal stopping time.

While investigating the hypothesis, the agent increasingly grows pessimistic about its being true as long as no breakthrough arrives. As an honest investigation can never show a false hypothesis to be true, all uncertainty is resolved at the first breakthrough, and the agent will know for sure that the proposition is true. Whereas the principal has no learning motive since he is only interested in the *first* breakthrough the agent achieves on arm 1, making information valuable to the agent provides an expedient way of giving incentives. Whereas 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 m being chosen appropriately large, in order to deter him from engaging in manipulation, which otherwise might seem expedient to him in the short term. Think e.g. of an investor who is wary of potentially being presented with fake evidence purporting to prove that an asset is good. Therefore, he will write a contract committing himself only to pay the analyst for the $(m+1)$-st piece of evidence presented, even though, in equilibrium, the agent is known to be honest with probability 1, so that the first piece of evidence presented already constitutes full proof that the asset is good. This commitment only to reward the $(m+1)$-st breakthrough is in turn what keeps the agent honest in equilibrium.

Now, the threshold number of successes m will be chosen high enough that even for an off-path agent, who has achieved his first breakthrough via manipulation, m breakthroughs are so unlikely to be achieved by cheating that he prefers to be honest after his first breakthrough. This puts a cheating off-path agent at a distinct disadvantage, as, in contrast to an honest on-path agent, he has not had a discontinuous jump in his belief. Thus, only an honest agent has a high level of confidence about his ability to navigate the continuation scheme devised by the principal; therefore, the agent will want to make sure he only enters the continuation regime after an honest success. Indeed, an agent who has had an honest success will be more optimistic about being able to curry favor with the principal by producing many additional successes in the future, while a cheating off-path agent, fully aware of his dishonesty, will be comparatively very pessimistic about his ability to produce a large number of future successes in the continuation game following the first success. Hence, the importance of being honest arises endogenously as a tool for the principal to give incentives in the cheapest possible way, as this difference in beliefs between on-path and off-path agents

is leveraged by the principal, who enjoys full commitment power.

This finding is consistent with empirical observations emphasizing that commitment to long-term compensation schemes is crucial in spurring innovation. Thus, Francis, Hasan, Sharma (2009) show that while performance sensitivity of CEO pay has no impact on innovation output, skewing incentives toward the long term via vested and unvested options does entail a positive and significant impact on both patents and citations to patents. Examining the impact of corporate R&D heads' incentives on innovation output, Lerner & Wulf (2007) find that long-term incentives lead to more patents, more extensively cited patents, and patents of greater originality.

In order to provide adequate incentives in the cheapest way possible, it is best for the principal to give a low value to a dishonest off-path agent after a first breakthrough, given the promised continuation value to the on-path agent. While paying only for the $(m+1)$-st breakthrough ensures that off-path agents do not persist in cheating, they will nevertheless continue to update their beliefs after their first success. 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. One way for the principal to handle this challenge is for him to end the game suitably soon after the first breakthrough, thereby curtailing the time the agent has access to the safe arm, thus correspondingly reducing the option value associated with it. Then, given this end date, the reward for the $(m+1)$-st breakthrough is chosen appropriately to give the intended continuation value to the on-path agent.

Alternatively, the model could be interpreted as one of an agent who is hired expressly to investigate a given hypothesis, yet who has the possibility of producing "fake breakthroughs." Think e.g. of a pharmaceutical firm hiring a scientist to produce a certain drug in a commercially viable way. Yet, it is common knowledge that there exists a commercially non-viable method of producing the drug, of which the scientist could surreptitiously avail himself to fake a breakthrough.[5] Generally speaking, these fake breakthroughs might be thought of as the pursuit of a research agenda that does not advance the public interest, or as an effort to massage or manipulate the data, with a view toward creating an erroneous impression that the hypothesis was proved.[6] There are studies suggesting that the problem of such scientific

[5] I am indebted to blogger "*afinetheorem*" for this example, cf. http://afinetheorem.wordpress.com/2010/09/12/the-importance-of-being-honest-n-klein-2010/ (as of October 13, 2011)

[6] A case in point, where a scientist's untoward behavior was eventually discovered, might be provided by (in)famous South Korean stem cell researcher Hwang Woo-Suk. Mr. Hwang was considered one of the world's foremost authorities in the field of stem-cell research, and was even designated his country's first "top scientist" by the South Korean Government. He purported to have succeeded in creating patient-matched stem cells, which would have been a major breakthrough that had raised high hopes for new cures for hitherto

misconduct is quite widespread indeed. In a survey of appertaining investigations, Fanelli (2009) concludes that one out of seven scientists admitted to their colleagues' having falsified data at least once, whereas only 1.97% admitted to having done so themselves. One third admitted to having engaged themselves in arguably less serious forms of misconduct, while 72% reported that their colleagues were guilty of such misconduct.[7]

The rest of the paper is set up as follows: Section 2 reviews some relevant literature; Section 3 introduces the model; Section 4 analyzes the provision of a certain continuation value; Section 5 analyzes the optimal mechanisms before a first breakthrough; Section 6 analyzes when the principal will optimally elect to stop the project conditional on no success having occurred, and Section 7 concludes. Proofs not provided immediately in the text are given in the Appendix.

2 Related Literature

Holmström & Milgrom (1991) analyze a case where, not unlike in my model, the agent performs several tasks, some of which may be undesirable from the principal's point of view. The principal may be able to monitor certain activities more accurately than others. They show that in the limiting case with two activities, one of which cannot be monitored at all, incentives will only be given for the activity which can in fact be monitored; if the activities are substitutes (complements) in the agent's private cost function, incentives are more muted (steeper) than in the single task case. While their model could be extended to a dynamic model with the agent controlling the drift rate of a Brownian Motion signal,[8] 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 had a discontinuous jump in his beliefs.

Bergemann & Hege (1998, 2005), as well as Hörner & Samuelson (2009), examine a venture capitalist's provision of funds for an investment project of initially uncertain quality;

hard-to-treat diseases, and that I am told had been the source of considerable pride in South Korea. Yet, a university panel found that "the laboratory data for 11 stem cell lines that were reported in the 2005 paper were all data made using two stem cell lines in total," forcing Mr. Hwang to resign in disgrace, and causing quite a shock to people in South Korea and throughout the scientific community. I am indebted to Tri-Vi Dang for alerting me to this story; see e.g. the report by the Associated Press from December, 23, 2005.

[7]Fanelli (2009) only looks at such forms of misconduct that distort overall scientific knowledge, such as data manipulation. Other forms of professional misconduct, as e.g. plagiarism, would not be included in the figures. He argues that his would be rather conservative estimates of the extent of the problem, based as they are on scientists' own responses to questionnaires featuring some possibly awkward queries.

[8]See Holmström & Milgrom (1987).

the project is managed by an entrepreneur, who might divert the funds to his private ends. The investor cannot observe the entrepreneur's allocation of the funds, so that, 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. While Bergemann & Hege (2005) and Hörner & Samuelson (2009) analyze the game without commitment, Bergemann & Hege (1998) investigate the problem under full commitment. These papers differ from my model chiefly in that there is no way for the entrepreneur to "fake" a success; any success that is publicly observed will have been achieved by honest means alone.

Gerardi & Maestri (2008) investigate the case of a principal who, in order to find out about the binary state of the world, has to employ an agent. The agent can decide to incur private costs to exert effort to acquire an informative binary signal, one realization of which is only possible in the good state. As for the principal, he can monitor neither the agent's effort choice nor the realization of the signal. The game ends as soon as the agent announces that he has had conclusive evidence in favor of the good state. They show that the agent needs to be left an information rent because of both the Moral Hazard and the Adverse Selection problems. In my model, by contrast, the game does not end after the first breakthrough; much to the contrary, I show that, in order to give optimal incentives in my model, it is absolutely vital that they be provided via the continuation game that follows the first breakthrough rather than via an immediate transfer.

One paper that is close in spirit to mine is Manso (2011), who analyzes a two-period model where an agent can either shirk, try to produce in some established manner with a known success probability, or experiment with a risky alternative. He shows that, in order 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 artefact of the discrete structure of the model and the limited signal space; indeed, in Manso's (2011) model, early failure can be a very informative signal that the agent has not exploited the known technology, but has rather chosen the risky, unknown alternative. By contrast, while confirming Manso's (2011) central intuition that it is better to give incentives through later rewards, I show that as the signal space becomes rich enough, the presence of the alternative production method does not distort the players' payoffs. Indeed, in continuous time, 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.

In Barraquer & Tan (2011), agents tend to congregate in those projects that are most informative about their underlying ability as market competition increases, making for a

potential source of inefficiency. In their model, the market observes in which project a success has been achieved. In my model, this is not observed by the principal; on the contrary, it is his goal to design incentives in such a way as to induce the agent to use the informative method of investigation.

Shan (2011) analyzes a contracting problem between a principal and an agent who is supposed to complete a multi-stage R & D project. The agent can shirk but not produce any fake successes. In his model, there is no uncertainty about the underlying state of the world, so that in the optimal contract, payments to the agent decrease continuously over time in the absence of a success, and jump to a higher level after each success. In my model, by contrast, we shall see that the evolution of rewards over time depends on the parameters on account of the countervailing effect of the agent's evolving belief about the state of the world.

De Marzo & Sannikov (2008) also incorporate private learning on the agent's part into their model, in which current output depends both on the firm's inherent profitability and on the agent's effort, which is unobservable to the principal. Thus, off-path, the agent's private belief about the firm's productivity will differ from the public belief. Specifically, if the agent withholds effort, this depresses the drift rate of the firm's Brownian motion cash flow. They show that the firm optimally accumulates cash as fast as it can until it reaches some target level, after which it starts paying out dividends; the firm is liquidated as soon as it runs out of cash. De Marzo & Sannikov (2008) show that one optimal way of providing incentives is to give the agent an equity stake in the firm, which is rescindable in the case of liquidation, and that liquidation decisions are efficient, agency problems notwithstanding.

To capture the learning aspect of the agent's problem, I model it as a bandit problem.[9] Bandit problems have been used in economics to study the trade-off between experimentation and exploitation since Rothschild's (1974) discrete-time single-agent model. The single-agent two-armed exponential model, a variant of which I am using, has first been analyzed by Presman (1990). Strategic interaction among several agents has been analyzed in the models by Bolton & Harris (1999, 2000), Keller, Rady, Cripps (2005), Keller & Rady (2010), who all investigate the case of perfect positive correlation between players' two-armed bandit machines, as well as by Klein & Rady (2011), who investigate the cases of perfect, as well as imperfect, negative correlation. Klein (2011) analyzes the case where bandits have three arms, with the two risky ones being perfectly negatively correlated. While the afore-mentioned papers all assumed that players' actions, as well as the outcomes of their actions, were perfectly publicly observable, Rosenberg, Solan, Vieille (2007), as well as Murto & Välimäki (2011), analyze the case where actions are observable, while outcomes are not. Bonatti & Hörner

[9] See Bergemann & Välimäki (2008) for an overview of this literature.

(2011) analyze the case where actions are not observable, while outcomes are. Bergemann & Välimäki (1996, 2000) consider strategic experimentation in buyer-seller interactions.

Rahman (2009, 2010) deals with the question of implementability in dynamic contexts, and finds that, under a full support assumption, a necessary and sufficient condition for implementability is for all non-detectable deviations to be unprofitable under zero transfers. The issue of implementability turns out to be quite simple in my model, and is dealt with in Proposition 3.1.

3 The Model

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

The principal's objective is to ensure at minimal cost that it is a best response for the agent to use arm 1 up to the first breakthrough with probability 1. He chooses an end date $\check{T}(t) \in [t, \overline{T})$ (where $\overline{T} \in (T, \infty)$ is arbitrary), in case the first breakthrough occurs at time t. Conditional on there having been no breakthrough, the game ends at time $T < \infty$. Once the game ends, utilities are realized. In the first half of the paper, the horizon T is exogenous. In the second half, when I let the principal choose the end date T, the first breakthrough achieved on arm 1 at time t gives him a payoff of $e^{-rt}\Pi$.[10]

Formally, the number of breakthroughs achieved on arm i up to, and including, time t defines the point processes $\{N_t^i\}_{0\leq t\leq \overline{T}}$ (for $i \in \{0,1\}$). In addition, let the point process $\{N_t\}_{0\leq t\leq \overline{T}}$ be defined by $N_t := N_t^0 + N_t^1$ for all t. Moreover, let $\mathfrak{F} := \{\mathfrak{F}_t\}_{0\leq t\leq \overline{T}}$ and $\mathfrak{F}^N := \left\{\mathfrak{F}_t^N\right\}_{0\leq t\leq \overline{T}}$ denote the filtrations generated by the processes $\{(N_t^0, N_t^1)\}_{0\leq t\leq \overline{T}}$ and

[10] I am following Grossman & Hart's (1983) classical approach to principal-agent problems in that I first solve for the optimal incentive scheme given an arbitrary T (Sections 4 and 5), and then let the principal optimize over T (Section 6).

$\{N_t\}_{0\leq t\leq \overline{T}}$, respectively.

By choosing which arm to pull, the agent affects the probability 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 $\theta\left(1 - e^{-\lambda_1 k_1 \Delta}\right)$.

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

A *wage scheme* offered by the principal is a non-negative, non-decreasing 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 the principal has consciously made to the agent up to, and including, time t. I assume the agent is protected by limited liability; hence $\{\mathcal{W}_t\}_{0\leq t\leq \overline{T}}$ is non-negative and non-decreasing.[11] 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 a schedule of end dates $\{\check{T}(t)\}_{t\in[0,T]}$, at the outset of the game. In order to ensure that the agent have a best response, I restrict the principal to choosing a piecewise continuous function $t \mapsto \check{T}(t)$.

Over and above the payments he gets as a function of breakthroughs, the agent can secure himself a safe payoff flow of s from the principal by pulling the safe arm, which is unobservable to the principal. The idea is that society cannot observe scientists shirking in real time, as it were; only after the lab e.g. is shut down, such information might come to light, and society only finds out *ex post* that it has been robbed of the payoff flow of s during the operation of the research lab. Thus, even though there is no explicit cost to the principal's provision of the bandit in my model, this assumption ensures that implied flow costs from doing so are at least s.

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

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

implementing a full-experimentation strategy, it is never a good idea for him to pay the agent in the absence of a breakthrough; moreover, since the principal is only interested in the first breakthrough, the notation can be simplified somewhat. Let $\{\mathcal{W}_t\}_{0\leq t\leq \overline{T}}$ be the principal's wage scheme, and t the time of the first breakthrough: In the rest of the paper, I shall write h_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 $\lim_{\tau\uparrow t} N_\tau = 0$, we can write $h_t := e^{rt}\left(\mathcal{W}_t - \lim_{\tau\uparrow t}\mathcal{W}_\tau\right)$. 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 will behave optimally in the future; formally,

$$w_t := \sup_{\{(k_{0,\tau},k_{1,\tau})\}_{t<\tau\leq\check{T}(t)}} E\left[e^{rt}\left(\mathcal{W}_{\check{T}(t)} - \mathcal{W}_t\right)\right.$$
$$\left.+s\int_t^{\check{T}(t)} e^{-r(\tau-t)}\left(1-k_{0,\tau}-k_{1,\tau}\right)\,d\tau\,|\mathfrak{F}_t, N_t^1 = 1, \lim_{\tau\uparrow t} N_\tau^1 = 0, N_t^0 = 0, \{(k_{0,\tau},k_{1,\tau})\}_{t<\tau\leq\check{T}(t)}\right],$$

i.e. the expectation conditions on the agent's knowledge that the first breakthrough has been achieved on arm 1 at time t. Again, I impose piecewise continuity of the mappings $t\mapsto h_t$ and $t\mapsto w_t$. The corresponding expected continuation payoff of an off-path agent, who achieves his first breakthrough on arm 0 at time t, I denote by ω_t; formally,

$$\omega_t := \sup_{\{(k_{0,\tau},k_{1,\tau})\}_{t<\tau\leq\check{T}(t)}} E\left[e^{rt}\left(\mathcal{W}_{\check{T}(t)} - \mathcal{W}_t\right)\right.$$
$$\left.+s\int_t^{\check{T}(t)} e^{-r(\tau-t)}\left(1-k_{0,\tau}-k_{1,\tau}\right)\,d\tau\,|\mathfrak{F}_t, N_t^0 = 1, \lim_{\tau\uparrow t} N_\tau^0 = 0, N_t^1 = 0, \{(k_{0,\tau},k_{1,\tau})\}_{0\leq\tau\leq\check{T}(t)}\right].$$

At the top of Section 5, I shall impose assumptions guaranteeing the piecewise continuity of the mapping $t\mapsto\omega_t$.

The state of the world is uncertain; clearly, whenever the agent uses arm 1, he gets new information about its quality; this *learning* is captured in the evolution of his (private) belief $\hat{p}_t$ that arm 1 is good. Formally, $\hat{p}_t := E\left[\theta|\mathfrak{F}_t, \{(k_{0,\tau},k_{1,\tau})\}_{0\leq\tau<t}\right]$. On the equilibrium path, the principal will correctly anticipate $\hat{p}_t$; formally, $p_t = \hat{p}_t$, where p_t is defined by $p_t := E\left[\hat{p}_t|\mathfrak{F}_t^N, \mathbf{k}\in\mathcal{K}\right]$.

The evolution of beliefs is easy to describe, since 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 $\hat{p}_t = 1$ forever thereafter.

As, in equilibrium, the agent will always operate arm 1 until the first breakthrough, it is clear that if on the equilibrium path $N_t \geq 1$, then $p_{t+\Delta} = 1$ for all $\Delta > 0$. If $N_t = 0$, Bayes' rule implies that

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

Now, before the first breakthrough, given an arbitrary incentive scheme $\mathbf{g} := (h_t, w_t)_{0 \leq t \leq T}$, the agent seeks to choose $\mathbf{k}_T \in \mathcal{U}_T$ so as to maximize

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

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 3.1 *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 $\lambda_0 \geq \lambda_1$, and suppose 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 $\tilde{k}_{0,t} = \frac{p_0 e^{-\lambda_1 t}}{p_0 e^{-\lambda_1 t} + 1 - p_0} \frac{\lambda_1}{\lambda_0}$ before the first breakthrough. After a first breakthrough, he sets $\tilde{k}_{0,t} = k_{0,t} + \frac{\lambda_1}{\lambda_0} k_{1,t} \leq k_{0,t} + k_{1,t}$, 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 $\mathbf{k}$; yet, the agent strictly prefers $\tilde{\mathbf{k}}$ as it gives him 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 the paper, I shall therefore assume that $\lambda_1 > \lambda_0$. When we denote the set of solutions to the agent's problem that are implemented by an incentive scheme $\mathbf{g}$ as $\mathbf{K}^*(\mathbf{g})$, the principal's problem is to choose $\mathbf{g} = (h_t, w_t)_{0 \leq t \leq T}$ so as to minimize his wage bill

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

subject to $p_t = \frac{p_0 e^{-\lambda_1 t}}{p_0 e^{-\lambda_1 t} + 1 - p_0}$ and $\mathbf{K}^*(\mathbf{g}) \cap \mathcal{K}_T \neq \emptyset$. It turns out that the solution to this problem coincides with the solution to the problem in which $\mathbf{K}^*(\mathbf{g}) = \mathcal{K}_T$ is additionally imposed; i.e. it is no costlier to the principal to implement full experimentation in *any* Nash equilibrium than to ensure that there exist a Nash equilibrium in which the agent employs a full-experimentation strategy (see Section 5).

In the next two sections, the end date T is given. In Section 6, the principal will optimally choose this end date T. Thus far, we have been silent on *how* the continuation value of w_t is delivered to the agent after his first breakthrough. It will turn out, though, that the manner by which the principal gives the agent his continuation value will matter greatly, as we will see in the next section.

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 of $w_t > 0$ given a first breakthrough has occurred at time t. His goal will be to find a scheme which maximally discriminates between an agent who has achieved his breakthrough on arm 1, as he was supposed to, and an agent who has been "cheating," i.e. who has achieved the breakthrough on arm 0. Put differently, for any given promise w_t to the on-path agent, it is the principal's goal to push the off-path agent's continuation value ω_t down, as this will give him a bigger bang for his buck in terms of incentives. As an off-path agent always has the option of imitating the on-path agent's strategy, we know that $\omega_t \geq \hat{p}_t w_t$, with $\hat{p}_t \in [p_t, p_0]$ denoting his (off-path) belief at time t. The following proposition summarizes the main result of this section; it shows that, as a function of $\hat{p}_t$, ω_t can be pushed arbitrarily close to this lower bound.

Proposition 4.1 *For every $\epsilon > 0$, $w_t > 0$, there exists a continuation scheme such that $\omega_t(\hat{p}_t) \leq \hat{p}_t w_t + \frac{s}{r}(1 - e^{-r\epsilon})$ for all $\hat{p}_t \in [p_t, p_0]$.*

Proof: Proof is by construction, see Subsection 4.2. ∎

The construction of this wage scheme relies on the assumption that $\lambda_1 > \lambda_0$, implying 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 in such a way as to reward realizations in the number of later breakthroughs that are "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 pull arm 0. Yet, now, in contrast to the off-path agents, an on-path agent will know for sure that his arm 1 is good, and therefore 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 right incentives to use arm 1 rather than arm 0 before the first breakthrough occurs.

4.2 Construction of An Optimal Continuation Scheme

Since ω_t would coincide with its lower bound $\hat{p}_t w_t$ if an on-path agent always played arm 1 after a first breakthrough, and off-path agents had no better option than to imitate the former's behavior, the purpose of the construction is to approximate such a situation. Since $\lambda_1 > \lambda_0$, 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 hold true for all off-path agents also. To this effect, the principal will only pay the agent for the m-th breakthrough after time t, where m is chosen large enough that any, even the most pessimistic of 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 above by $\frac{s}{r}(1 - e^{-r\epsilon})$. Then, given the end date $\check{T}(t)$, the reward is chosen appropriately so that the on-path agent exactly receive his promised continuation value of w_t in expectation.

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

Lemma 4.2 *There exists an integer m such that if the agent is only paid a lump sum reward $\overline{V}_0 > 0$ for the m-th breakthrough, arm 1 dominates arm 0 for any type of off-path agent whenever he still has m breakthroughs to go before collecting the lump sum reward.*

PROOF: See Appendix. ∎

Intuitively, the likelihood ratio of m breakthroughs being achieved on arm 1 vs. arm 0 in the time interval $(t, \check{T}(t)]$, $\hat{p}_t \left(\frac{\lambda_1}{\lambda_0}\right)^m e^{-(\lambda_1 - \lambda_0)(\check{T}(t) - t)}$, is unbounded in m. The proof now shows, by virtue of a first-order stochastic dominance argument, that when m exceeds certain thresholds, it indeed never pays for the agent to use arm 0.

Thus, Lemma 4.2 shows that we can ensure that off-path agents will never continue to use arm 0 after time t. Ending the game suitably soon after a first breakthrough, namely at some time $\check{T}(t) \in (t, t + \epsilon]$, bounds off-path agents' option values from having access to the safe arm by $\frac{s}{r}(1 - e^{-r\epsilon})$. Hence, an off-path agent of type $\hat{p}_t$ can indeed at most get $\hat{p}_t w_t + \frac{s}{r}(1 - e^{-r\epsilon})$.

The purpose of the rest of this subsection is to show that, given $\check{T}(t)$ and m, $\overline{V}_0$ can be chosen in a way that ensures that the on-path agent exactly get what he is supposed to get,

namely w_t. In order to do so, given m, $\check{T}(t)$, and $\overline{V}_0$, I now recursively define the auxiliary functions $V_i(\cdot;\overline{V}_0) : [t,\check{T}(t)] \longrightarrow \mathbb{R}$ for $i = 1,\cdots,m$ according to

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

where $\mathcal{M}(\tilde{t})$ denotes the set of measurable functions $k_i : [\tilde{t},\check{T}(t)] \to [0,1]$, and I set $V_0(\tau;\overline{V}_0) := \overline{V}_0 + \frac{s}{r}\left(1 - e^{-r(\check{T}(t)-\tau)}\right)$. 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 the rest of this section in the following proposition:

Proposition 4.3 *(1.) If $w_t > \lim_{\overline{V}_0\downarrow\frac{s}{\lambda_1}} V_m(t;\overline{V}_0)$, there exists a lump sum $\overline{V}_0 > \frac{s}{\lambda_1}$ such that $w_t = V_m(t;\overline{V}_0)$.*

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

PROOF: The proof of statement (1.) relies on certain properties of the V_i functions, which are exhibited in Lemma 4.4 below. The proof of statement (2.) additionally uses another auxiliary function f, which is also introduced *infra*, and some properties of which are stated in Lemma 4.5 below. The proof is therefore provided in the appendix after the proofs of Lemmas 4.4 and 4.5. ∎

As already mentioned, the following lemma is central to the proof of Proposition 4.3. 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 he has i breakthroughs to go. It also takes note of some useful properties of the functions V_i:

Lemma 4.4 *Let $\overline{V}_0 > \frac{s}{\lambda_1}$. A best response for the agent is given by a sequence of cutoff times $t_m^* \geq \cdots \geq t_2^* > t_1^* = \check{T}(t)$ (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^* $(i = 1,\cdots,m)$ is increasing in $\overline{V}_0$; moreover, for $i = 2,\cdots,m$, 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 $\tilde{t} < t_i^*(\overline{V}_0)$).*[12] *Moreover,*

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

$\lim_{\overline{V}_0\to\infty} t_i^* = \check{T}(t)$, *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*

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

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

Proof: See Appendix. ∎

The lemma thus immediately implies that if $w_t > \lim_{\overline{V}_0\downarrow\frac{s}{\lambda_1}} V_m(t;\overline{V}_0)$ for the given end date $\check{T}(t)$, we can find an appropriate $\overline{V}_0 > \frac{s}{\lambda_1}$ ensuring that $w_t = V_m(t;\overline{V}_0)$, as we note in statement (1.) of Proposition 4.3.

If $w_t \leq V_m(t;\frac{s}{\lambda_1})$, we need to lower the end date $\check{T}(t)$ further, as statement (2.) in Proposition 4.3 implies. For this purpose, it turns out to be useful to define another auxiliary function $f : [t,\overline{T}) \times (\frac{s}{\lambda_1},\infty) \longrightarrow \mathbb{R}$ 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 $i = 1,\cdots,m$, I write $V_i(t;\overline{V}_0;\check{T}(t))$ for $V_i(t;\overline{V}_0)$ given 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 $\overline{V}_0 > \frac{s}{\lambda_1}$. The following lemma takes note of some properties of f:

Lemma 4.5 $f(\cdot,\overline{V}_0)$ *is continuous and strictly increasing with* $f(t;\overline{V}_0) = 0$.

Proof: See Appendix. ∎

As we note in the proof of Proposition 4.3, it immediately follows from Lemma 4.5 that we can choose a lump sum $\hat{\overline{V}}_0 > \frac{s}{\lambda_1}$ and an end date $\check{\check{T}}(t) < t+\epsilon$, so that $w_t = f(\check{\check{T}}(t),\hat{\overline{V}}_0)$. As one and the same m can be used for all $\check{T}(t)$ and $\hat{\overline{V}}_0$, and 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 $w_t = f(\check{T}(t);\hat{\overline{V}}_0)$.

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 in order to harness maximal incentive power at a given cost. On the one hand, it must make sure off-path agents never continue to play arm 0; this is achieved by only rewarding the m-th breakthrough after time t, with m being chosen appropriately high. On the other hand, the mechanism must preclude the more pessimistic off-path agents from collecting an excessive option value from being able to switch between the safe arm and arm 1. This is achieved by ending the game soon enough after a first breakthrough.

5 Incentive Provision Before A Breakthrough

Whereas in the previous section, I have investigated how the principal would optimally deliver a given *continuation* value w_t, the purpose of this section is to understand how optimally to provide incentives before a first breakthrough. I shall show that thanks to the continuation scheme we have constructed in the previous section (see Proposition 4.1), arm 0 can be made so unattractive that in any optimal scheme it is dominated by the safe arm. Thus, in order to induce the agent to use arm 1, he only needs to be compensated for his outside option of playing safe, which pins down the principal's wage costs (Proposition 5.3).

In order formally to analyze the optimal incentive schemes before a first breakthrough, we first have to consider the agent's best responses to a given incentive scheme $(h_t, w_t)_{0\leq t\leq T}$, in order 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 the current incentives he is facing but also on the entire path of future incentives. To the extent 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 $(\hat{p}_t, h_t, w_t)$ would be inadequate. Thus, I shall be using Pontryagin's Optimal Control approach.

The Agent's Problem

Given an incentive scheme $(h_t, w_t)_{0\leq t\leq T}$, the agent chooses $(k_{0,t}, k_{1,t})_{0\leq t\leq T}$ so as to maximize

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

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

It will turn out to be useful to work with the log-likelihood ratio $x_t := \ln\left(\frac{1-\hat{p}_t}{\hat{p}_t}\right)$, 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 $x_0 = \ln\left(\frac{1-p_0}{p_0}\right)$ and $y_0 = 1$ are given, and x_T and y_T are free . The agent's controls are $(k_{0,t}, k_{1,t}) \in \{(a,b) \in \mathbb{R}_+ : a+b \leq 1\}$.

Neglecting a constant factor, the Hamiltonian $\mathfrak{H}_t$ is now given by[13]

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

By the Maximum Principle,[14] the existence of absolutely continuous functions μ_t and γ_t respectively satisfying the equations (1) and (2) a.e., as well as (3), 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:[15]

$$\begin{aligned}\dot{\mu}_t = e^{-rt}y_t\left\{e^{-x_t}\left[(1-k_{0,t}-k_{1,t})s + k_{0,t}\lambda_0(h_t+\omega_t(x_t)) + k_{1,t}\lambda_1(h_t+w_t)\right]\right. \\ \left. -k_{0,t}\lambda_0(1+e^{-x_t})\omega_t'(x_t)\right\}, \quad (1)\end{aligned}$$

$$\begin{aligned}\dot{\gamma}_t = -e^{-rt}\{[(1-k_{0,t}-k_{1,t})s + k_{0,t}\lambda_0(h_t+\omega_t(x_t))] \\ + e^{-x_t}\left[(1-k_{0,t}-k_{1,t})s + k_{0,t}\lambda_0(h_t+\omega_t(x_t)) + k_{1,t}\lambda_1(h_t+w_t)\right]\} + \gamma_t\lambda_0 k_{0,t}\,, \quad (2)\end{aligned}$$

$$\begin{aligned}e^{-rt}y_t\left[e^{-x_t}\lambda_1(h_t+w_t) - (1+e^{-x_t})s\right] + \mu_t\lambda_1 \\ \geq \max\left\{0, e^{-rt}y_t(1+e^{-x_t})[\lambda_0(h_t+\omega_t(x_t)) - s] - \gamma_t\lambda_0 y_t\right\}. \quad (3)\end{aligned}$$

In order to ensure the piecewise continuity of $\omega_t(x_t)$ in t (for a given x_t), I shall henceforth assume that in the continuation scheme following a first success, the principal will apply a threshold number of successes m that is constant over time. (The proof of Lemma 4.2 shows that m can be chosen in this way.) Moreover, to the same end, I will be assuming that $\overline{V}_0$, the lump sum reward for the $(m+1)$-st breakthrough overall, is piecewise continuous as a function of t, the time of the first breakthrough.[16] In the following lemma, I now show that

[13] In a slight abuse of notation, I now write ω_t as a function of x_t.

[14] See Theorem 2 in Seierstad & Sydsæter, 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 $x = \ln\left(\frac{1-\hat{p}}{\hat{p}}\right)$ is a continuous one-to-one transformation of $\hat{p}$, the relevant continuity requirements in Seierstad & Sydsæter, 1987, footnote 9, p. 132, are satisfied.

[15] 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. Since 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 (since x_t can only ever increase over time).

[16] Note that $w_t = V_m(t;\overline{V}_0;\check{T}(t))$, and 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 either innocuous (which may be the case either because $t_m^* = t$ both before and after the jump, or because it is exactly counterbalanced by a jump in $\check{T}(t)$), or it leads to a jump in w_t. Since 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.

the agent indeed has a best response:

Lemma 5.1 *The agent has a best response to any given incentive scheme* $(h_t, w_t)_{0\leq t\leq T}$.

PROOF: See Appendix. ∎

To state the following proposition, I define $\epsilon_t := \check{T}(t) - t$. I shall say that a wage scheme is *continuous* if h_t, w_t and ϵ_t are continuous functions of time t. The following proposition shows that if a wage scheme is continuous, then Pontryagin's conditions are not only necessary, but also sufficient, for the agent's best-responding by being 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. but possibly for deviations on a null set, which are innocuous to the principal).

Proposition 5.2 *Suppose that* $k_{1,t} = 1$ *for all* t *satisfies Pontryagin's necessary conditions as stated above, even for the upper bound on* ω_t *given by Proposition 4.1. Suppose furthermore that* h_t, w_t, *and* ϵ_t *are continuous functions of time* t. *Then, if* $(k_{0,t}, k_{1,t})_{0\leq t\leq T}$ *is a best response, it is the case that* $k_{1,t} = 1$ *for a.a.* t.

PROOF: See Appendix. ∎

Our strategy for the rest of this section is to find the cheapest possible schemes such that the agent's necessary conditions for his best responding by being honest be satisfied. In a second step, we shall then verify if that scheme is in fact continuous. If it is, it must be optimal, since any cheaper scheme would violate even the necessary conditions for honesty by our first step.

We note that 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 (1) and (2) as

$$\dot{\mu}_t = e^{-rt-x_t}\lambda_1(h_t + w_t), \tag{4}$$

$$\dot{\gamma}_t = -\dot{\mu}_t. \tag{5}$$

Furthermore we can rewrite (3) as the following two joint conditions:

$$e^{-rt}\left[e^{-x_t}\lambda_1(h_t + w_t) - (1 + e^{-x_t})s\right] \geq -\mu_t\lambda_1, \tag{6}$$

and

$$e^{-rt}\left[e^{-x_t}\lambda_1(h_t + w_t) - (1 + e^{-x_t})\lambda_0(h_t + \omega_t(x_t))\right] \geq -\mu_t(\lambda_1 - \lambda_0). \tag{7}$$

The Principal's Problem

Now we turn to the principal's problem, 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 $t \in [0, T]$. Thus, the principal's objective is to choose $(h_t, w_t)_{0 \leq t \leq T}$ (with $(h_t, w_t) \in [0, L]^2$ at all t, for some $L > 0$ which is chosen large enough) so as to minimize

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

subject to the constraints $x_t = x_0 + \lambda_1 t$, $y_t = 1$, (3), (4), (5), and the transversality conditions $\mu_T = \gamma_T = 0$, which, as we have discussed, are necessary for the agent to best respond by being honest.

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

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

By (4) and (5), we have that

$$\mu_t = -\gamma_t = -\lambda_1 e^{-rt-x_t} \int_t^T e^{-(r+\lambda_1)(\tau-t)} (h_\tau + w_\tau) \, d\tau = -\frac{\lambda_1 p_t}{1-p_t} e^{-rt} \int_t^T e^{-(r+\lambda_1)(\tau-t)} (h_\tau + w_\tau) \, d\tau. \tag{8}$$

Thus, $-\mu_t$ measures the agent's opportunity costs from possibly forgone future rewards. They adversely impact today's incentives, as by pulling arm 1 today the agent "risks" having his first breakthrough today, thereby forfeiting his chance of collecting the rewards offered for a first breakthrough tomorrow. Hence, generous rewards are doubly expensive for the principal: On the one hand, he has to pay out more in case of a breakthrough today; yet, on the other hand, by paying a lot today, he might make it attractive for the agent to procrastinate at previous points in time in the hope of winning today's reward. In order to counteract this effect, the principal has to offer higher rewards at previous times in order to maintain 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. Since $\mu_t = -\gamma_t$ for all $t \in [0, T]$, we shall henceforth only keep track of μ_t.

Our following proposition will give a superset of all optimal schemes, as well as exhibit an optimal scheme. It will furthermore show that optimality uniquely pins down the principal's wage costs. In the class of schemes with $h_t = 0$ for all t, the optimal scheme is essentially

unique. The characterization relies on the fact that it never pays for the principal to give 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 he will do the right thing come 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 makes sure that the agent's best outside option can never be arm 0. Indeed, 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 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

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

As is readily verified by plugging $\mu_t = -\lambda_1 \int_t^T e^{-r\tau - x_\tau}(h_\tau + w_\tau)\, d\tau$ into the incentive constraint for the safe arm, $\tilde{w}(t)$ is the reward that an agent with the belief p_t has to 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 $\frac{s}{\lambda_1 p_t}$ signifies the compensation the agent must receive for forgoing the immediate flow of sdt; 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 have analyzed in Section 4. This continuation phase has to compensate the agent for the forgone access to the safe arm he would have enjoyed in the absence of a breakthrough; this function is performed by the second term, $\frac{s}{r}(1 - e^{-r(T-t)})$. The third term counteracts the allure of future incentives, which might induce the agent to procrastinate, as we have discussed *supra*. It is increasing in the remainder of time, $T-t$, and arbitrarily small for very impatient agents. We are now ready to characterize the principal's optimal wage schemes:

Proposition 5.3 *If a scheme is optimal, it is in the set $\mathcal{E}$, with*

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

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

PROOF: By construction of $\tilde{w}$, (6) binds at a.a. t for all schemes in $\mathcal{E}$. Algebra shows that (7) holds given that (6) binds if, and only if,

$$\frac{e^{x_t}}{1+e^{x_t}}h_t + \omega_t(x_t) \leq \frac{w_t}{1+e^{x_t}} + s\left(\frac{1}{\lambda_0} - \frac{1}{\lambda_1}\right). \tag{9}$$

As by Proposition 4.1, $\omega_t(p_t) > p_t w_t$, yet arbitrarily close to $p_t w_t$, condition (9) is equivalent to the inequality in the definition of $\mathcal{E}$.[17] Clearly, (9) is satisfied for $h_t = 0$ and $\frac{s}{r}\left(1 - e^{-r\epsilon_t}\right) \leq s\left(\frac{1}{\lambda_0} - \frac{1}{\lambda_1}\right)$. 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.

By the construction of $\tilde{w}$, any scheme that is not in $\mathcal{E}$ yet satisfies the constraints (4), (5), (6) and (7) a.s., as well as the transversality condition, is more expensive to the principal than any scheme in $\mathcal{E}$. Proposition 5.2 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 $h_t = 0$ 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 $h_t = 0$ 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 when paying an immediate lump sum for the first breakthrough, the principal cannot discriminate between an agent who has achieved his first breakthrough on arm 0 on the one hand, and an on-path agent on the other; the latter, though, will enjoy an informational advantage in the continuation game. Indeed, by Proposition 4.1, the principal can make sure that an increase in w_t translates into less of an increase in ω_t, whereas h_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. The fundamental reason for this is that the information released by arm 0 is a Blackwell-garbling of the information released by arm 1; hence, having had a breakthrough on arm 1 is (weakly) better, no matter what the ensuing decision problem may be. If, by contrast, the principal wanted to implement an action yielding Blackwell-garbled information, the best he could do would be to give myopic incentives; in my model, if the goal was the implementation of arm 0, the principal could without loss restrict himself to schemes with $w_t = \omega_t = 0$ for all t.

[17] If λ_0 is so low that the proof of Proposition 4.1 goes through for $m = 1$, the inequality $(1 - p_t)h_t \leq s\left(\frac{1}{\lambda_0} - \frac{1}{\lambda_1}\right)$ is weak, rather than strict. The same holds true if $w_t = 0$.

A further immediate implication of the preceding proposition is that the optimal incentive scheme is essentially unique in that the wage payments $h_t + w_t$ are a.s. uniquely pinned down. Clearly, optimal wage costs $\tilde{w}$ are decreasing in r, implying that incentives are the cheaper to provide the more impatient the agent is. As the agent becomes myopic ($r \to \infty$), wage costs tend to $\frac{s}{\lambda_1 p_t}$, since in the limit he now only has to 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 $\frac{s}{\lambda_1 p_T} + s(T-t)$. Concerning the evolution of rewards over time, there are two countervailing effects as in Bonatti & Hörner (2011): On the one hand, the agent becomes more pessimistic over time, so that rewards will have to increase to make him willing to use arm 1 nonetheless; on the other hand, as the end date approaches, the idea of 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 rewards will be increasing. For very small r, by contrast, the dynamic effect dominates, and rewards will decrease over time.

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 (7), the incentive constraint for arm 0, since $h_t \geq 0$ by limited liability:

$$e^{-rt}(p_t\lambda_1 - \lambda_0)h_t \geq -\mu_t(1-p_t)(\lambda_1 - \lambda_0) > 0. \tag{10}$$

Conversely, whenever $p_t\lambda_1 > \lambda_0$, it is always possible to blow up h_t in a way 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 5.3, a necessary condition for immediate rewards to be consistent with optimality at a generic time t is that $\tilde{w}(t) < \frac{s}{1-p_t}\left(\frac{1}{\lambda_0} - \frac{1}{\lambda_1}\right)$. To show that this is a more stringent condition than $p_t\lambda_1 > \lambda_0$, we note that optimality implies that the incentive constraint for the safe arm (6) will bind given the reward offered is $\tilde{w}(t)$, and hence that

$$-\mu_t = e^{-rt}\left\{\frac{s}{r}\frac{p_t}{1-p_t}\left(1 - e^{-r(T-t)}\right) + \frac{s}{r-\lambda_1}\left(1 - e^{-(r-\lambda_1)(T-t)}\right)\right\}$$

if $r \neq \lambda_1$. For immediate incentives to be consistent with optimality, it is necessary that $h_t = \frac{s}{1-p_t}\left(\frac{1}{\lambda_0} - \frac{1}{\lambda_1}\right)$ satisfy the incentive constraint (10) given this μ_t with slackness, which is equivalent to

$$p_t\lambda_1 - \lambda_0 > (1-p_t)\lambda_0\lambda_1\left[p_t\frac{1-e^{-r(T-t)}}{r} + (1-p_t)\frac{1-e^{-(r-\lambda_1)(T-t)}}{r-\lambda_1}\right] > 0.$$

Thus, in summary, while the implementation of arm 1 by mere immediate lump-sum incentives is feasible if, and only if, $p_t\lambda_1 - \lambda_0 > 0$, restricting himself to these immediate incentives will come at a cost to the principal whenever $p_t\lambda_1 - \lambda_0$ violates a strictly more stringent condition.[18]

6 The Optimal Stopping Time

In this section, we let the principal additionally choose to what end date $T \in [0, \overline{T})$ to commit at the outset of the game (with $\overline{T} < \infty$ chosen 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 he 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, he chooses T so as to maximize

$$\int_0^T \left\{ e^{-rt-\lambda_1 \int_0^t p_\tau \, d\tau} (p_t\lambda_1\Pi - s) \right\} dt \tag{11}$$

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 $p_t \geq \frac{s}{\lambda_1\Pi} =: p^m$. As the principal is only interested in the first breakthrough, information has no value for him, meaning that, very much in contrast to the classical bandit literature, he is not willing to forgo current payoffs in order 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 $T^{FB} = \frac{1}{\lambda_1} \ln\left(\frac{p_0}{1-p_0}\frac{1-p^m}{p^m}\right)$.

Regarding the second-best situation where 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 so as to maximize

$$\int_0^T \left\{ e^{-rt-\lambda_1 \int_0^t p_\tau \, d\tau} p_t\lambda_1 \left(\Pi - \tilde{w}(t)\right) \right\} dt \tag{12}$$

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 (11) is that the opportunity cost flow s is now replaced by the optimal wage costs $\tilde{w}(t)$ (see Proposition 5.3). These of course only have to be paid out in case of a success, which happens with an instantaneous

[18] The corresponding necessary condition for the case $r = \lambda_1$ is given by

$$p_t\lambda_1 - \lambda_0 > (1-p_t)\lambda_0\lambda_1 \left[p_t \frac{1-e^{-r(T-t)}}{r} + (1-p_t)(T-t) \right] > 0.$$

probability of $p_t\lambda_1 dt$. After plugging in for $\tilde{w}(t)$, one finds that the first-order derivative of the objective with respect to T is given by

$$\underbrace{e^{-(r+\lambda_1)T}\left(\lambda_1\Pi - \frac{s}{p_T}\right)}_{\text{Marginal effect}} \underbrace{-e^{-rT}\frac{s}{p_T}\left(1-e^{-\lambda_1 T}\right)}_{\text{Intra-marginal effect}}. \tag{13}$$

The marginal effect captures the benefit the principal could collect by extending experimentation for an additional instant at time T. Yet, as we have discussed in Section 5, the choice of an end date T also entails an intra-marginal effect at times $t < T$. Indeed, we have seen that for him to use arm 1 at time t, the agent has to be compensated for the opportunity cost of the potentially forgone rewards for having his first breakthrough at some future date, an effect that is the stronger "the more future there is," i.e. the more distant the end date T is. Hence, by marginally increasing T, the principal also marginally raises his wage liabilities at times $t < T$. This creates a distortion, so that the following proposition comes as no surprise:

Proposition 6.1 *Let $p_0 > p^m$. The principal stops the game at the 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 gotten by setting the expression (13) to 0, and verifying that the second-order condition holds. Now, T^* is the unique root of $\frac{p_0 e^{-\lambda_1 T^*}}{p_0 e^{-\lambda_1 T^*} + 1 - p_0} = p^m e^{\lambda_1 T^*}$. ∎

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 $\frac{1}{p^m} = \frac{\lambda_1 \Pi}{s}$. The size of the distortion can be measured by the ratio $\frac{p_{T^*} - p^m}{p^m}$, which is also increasing in the stakes at play. This is because of the intra-marginal effect we have discussed *supra*; as stakes increase, and the principal consequently extends the deadline T^*, the agent's incentives for procrastination are exacerbated at intra-marginal points in time. This in turn increases the agent's wages $\tilde{w}(t)$ at these intra-marginal points in time, so that the principal can only appropriate part of any increase in the overall pie. Yet the wedge $\frac{p_{T^*} - p^m}{p^m}$ is also increasing in players' optimism, as measured by p_0. Since p^m is independent of p_0, this implies that the threshold belief p_{T^*} is increasing in p_0. Whereas 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 second-order 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; on the other hand, similarly to the case of rising stakes, any increase in the

end date additionally compounds the agent's proclivity for procrastination. The following proposition summarizes these comparative statics:

Proposition 6.2 *The stopping time T^* as well as the wedge $\frac{p_{T^*}-p^m}{p^m}$ are increasing in the stakes at play $\frac{\lambda_1 \Pi}{s}$ and in players' optimism p_0.*

PROOF: See Appendix. ∎

Yet, also recall from the preceding sections that given the optimal incentive scheme we have computed there, 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 to some extent cannibalize today's rewards. Yet, 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 so chooses. Now, in the limit, if the principal can hire agents for a mere infinitesimal instant dt, he can completely shut down the intra-marginal effect we have discussed above.[19] 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 the reward an agent who is only hired for an instant of length dt would have to be promised for a breakthrough is given by $\frac{s}{\lambda_1 p_t}(1+\lambda_1 dt)+o(dt)$. Hence, it pays for the principal to go on with the project as long as $p_t\lambda_1\left(\Pi-\frac{s}{p_t\lambda_1}(1+\lambda_1 dt)\right)dt+o(dt)=p_t\lambda_1\left(\Pi-\frac{s}{p_t\lambda_1}\right)dt+o(dt)>0$, i.e. he stops at the first-best efficient stopping time, a result I summarize in the following proposition:

Proposition 6.3 *If the principal has access to a sequence of different agents, he stops the delegated project at the 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

[19] Intuitively, one might think that hiring *one* particularly myopic agent might remedy the problem as well. However, while it is true that the impact future rewards have on today's incentives, and hence the intra-marginal 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 that in sum 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's hiring an agent who is much more impatient than himself. I leave the analysis of players with differing discount factors outside the scope of this paper.

problem can be overcome, though, if the principal has access to a sequence of many agents. To sum, if $\lambda_0 \geq \lambda_1$, the option to cheat makes it impossible to make the agent use arm 1; if $\lambda_0 < \lambda_1$, by contrast, incentives are optimally structured in such a way as to obviate any impact of the cheating option on 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 on bandits. I have shown that even though the principal only cares about the first breakthrough, it is without loss for him only to reward later ones. Thus, even though the agent will be honest for sure in equilibrium, and hence the first observed breakthrough reveals everything the principal wants to know, committing to rewarding 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 ones. Structuring incentives appropriately in this fashion precludes any distortions arising from the agent's option to cheat whenever the cheating option does not render the provision of incentives completely impossible. Thus, my analysis would suggest that skewing incentives more toward rewarding longer-term performance might constitute an avenue to explore in addressing agents' propensity "to play it too safe."

In my model, the principal only employs one single agent at any given moment in time. It would be interesting to explore how the structure of the optimal incentive scheme would change if several agents were simultaneously investigating the same hypothesis. Intuition would suggest that the rationale for only rewarding later breakthroughs should carry over to that case. I leave a full exploration of these questions for future work.

Appendix

Proof of Lemma 4.2

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

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

where F is the distribution over τ_m, the time of the m-th 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 m-th 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

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

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, since 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

$$F^*(\tau; \hat{p}_{\tilde{t}}) > F_\alpha(\tau; \hat{p}_{\tilde{t}}) \tag{A.1}$$

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

Computations show that

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

Further computations show that $\frac{\partial F_\alpha}{\partial \alpha} > 0$ if and only if

$$\xi(\alpha) > 1 - \lambda_0 \frac{\tau - \tilde{t}}{m}$$

with

$$\xi(\alpha) := \frac{\hat{p}_{\tilde{t}}}{1-\hat{p}_{\tilde{t}}} e^{-\alpha\lambda_1(\tau-\tilde{t})} \left(\frac{\lambda_1}{\lambda_0} - 1\right) \left(\frac{\alpha\lambda_1 + (1-\alpha)\lambda_0}{(1-\alpha)\lambda_0}\right)^{m-1} \left[1 - (\alpha\lambda_1 + (1-\alpha)\lambda_0)\frac{\tau-\tilde{t}}{m}\right] - \alpha\lambda_0 \frac{\tau-\tilde{t}}{m}$$

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

Now, we choose $m \geq 2$ high enough that

$$(m-1)\left(\frac{\lambda_1}{\lambda_0} - \frac{\lambda_1^2}{\lambda_0}\frac{\overline{T}}{m}\right) - \frac{\lambda_1^2}{\lambda_0}\overline{T}\left(1+\frac{1}{m}\right) > \frac{1-p_{\overline{T}}}{p_{\overline{T}}}\frac{\lambda_0^2}{\lambda_1-\lambda_0} e^{\lambda_1\overline{T}}\frac{\overline{T}}{m}. \tag{A.2}$$

As the left-hand side of (A.2) is diverging to $+\infty$ for $m \to +\infty$ while the right-hand side converges to 0, such an $m \geq 2$ exists. Algebra shows that (A.2) ensures that $\lim_{\alpha\uparrow 1}\xi(\alpha) = \infty$ and $\xi'(\alpha) > 0$ for all $\alpha \in (0,1)$. Now, if $\hat{p}_{\tilde{t}}\lambda_1 \geq \lambda_0$, it is the case that $\xi(0) \geq 1 - \lambda_0\frac{\tau-\tilde{t}}{m}$, and hence $\frac{\partial F_\alpha}{\partial\alpha} > 0$ for all $\alpha \in (0,1)$, so that $F^*(\tau;\hat{p}_{\tilde{t}}) > F_\alpha(\tau;\hat{p}_{\tilde{t}})$ for all $\alpha \in [0,1)$. In case $\hat{p}_{\tilde{t}}\lambda_1 < \lambda_0$, we have that $\xi(0) < 1 - \lambda_0\frac{\tau-\tilde{t}}{m}$; now, there exists a unique $\alpha^* \in (0,1)$ such that $\frac{\partial F_\alpha}{\partial\alpha} < 0$ for all $\alpha \in (0,\alpha^*)$, $\frac{\partial F_\alpha}{\partial\alpha} > 0$ for all $\alpha \in (\alpha^*,1)$, and we can conclude that $F_\alpha(\tau;\hat{p}_{\tilde{t}})$ is maximized either by $\alpha = 1$ or $\alpha = 0$. Choosing m such that

$$p_{\overline{T}}\left(\frac{\lambda_1}{\lambda_0}\right)^m > e^{(\lambda_1-\lambda_0)\overline{T}} \tag{A.3}$$

ensures that the maximum is indeed attained at $\alpha = 1$.

In summary, there exists an $m \in \mathbb{N} \cap [2,\infty)$ satisfying both (A.2) and (A.3). Choosing m in this manner ensures that

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

for all $\alpha \in [0,1)$. Note that the choice of m is independent of $\tilde{t}$, $\tau > \tilde{t}$, and $\hat{p}_{\tilde{t}} \in [p_{\overline{T}}, p_0]$. Hence, for such an m, it is clearly 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 of always pulling arm 1, whatever 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

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

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 $\tilde{F}^{\{k_\tau\}}(\cdot;\hat{p}_{\tilde{t}}) = 0$; if $\int_{\tilde{t}}^{\check{T}(t)} k_\sigma\, d\sigma > 0$, then $(\tilde{F}^{\{k_\tau\}})^*$, the distribution over the m-th 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. Since 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 (A.2) and (A.3) 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 $e^{-r(\tau_m-\tilde{t})}\left(\overline{V}_0+\frac{s}{r}(1-e^{-r(\check{T}(t)-\tau_m)}\right)$ 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 4.4

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 HJB equation is given by

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

Thus, neglecting terms of order dt^2 and higher, and re-arranging gives us

$$rV_i(\tilde{t};\overline{V}_0) = s+\dot{V}_i(\tilde{t};\overline{V}_0)+k_{1,\tilde{t}}\left[\lambda_1\left(V_{i-1}(\tilde{t};\overline{V}_0)-V_i(\tilde{t};\overline{V}_0)\right)-s\right]. \tag{A.4}$$

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

$$V_{i-1}(\tilde{t};\overline{V}_0)-V_i(\tilde{t};\overline{V}_0)\geq\frac{s}{\lambda_1}; \tag{A.5}$$

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

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

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

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

By simple algebra, one finds that

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

which one shows strictly to exceed $\frac{s}{\lambda_1}$ for all $\tilde{t}\in(t,\check{T}(t)]$ if $\overline{V}_0>\frac{s}{\lambda_1}$. 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

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

with $t_1^* = \check{T}(t)$ is optimal. Furthermore, $V_1(\cdot;\overline{V}_0)$ is absolutely continuous, and 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 following structure:

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

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

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

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 absolutely continuous and 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 $V_{i-1}(t;\overline{V}_0) < \frac{s}{\lambda_1} + \frac{s}{r}\left(1-e^{-r(\check{T}(t)-t)}\right)$, I set $t_i^* = t$. Otherwise, I define t_i^* as the lowest t^* satisfying $V_{i-1}(t^*;\overline{V}_0) = \frac{s}{\lambda_1} + \frac{s}{r}\left(1-e^{-r(\check{T}(t)-t^*)}\right)$. Since $\dot{V}_{i-1}(\tilde{t};\overline{V}_0) \leq -se^{-r(\check{T}(t)-\tilde{t})}$ 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 $V_{i-1}(\tilde{t};\overline{V}_0) \leq \frac{s}{\lambda_1} + \frac{s}{r}\left(1-e^{-r(\check{T}(t)-\tilde{t})}\right)$, 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 $V_i(\hat{\tau};\overline{V}_0) = \frac{s}{r}\left(1-e^{-r(\check{T}(t)-\hat{\tau})}\right)$ solve the HJB equation. Indeed, the payoff function $V_i(\hat{\tau};\overline{V}_0) = \frac{s}{r}\left(1-e^{-r(\check{T}(t)-\hat{\tau})}\right)$ is of class C^1, and, since $\dot{V}_{i-1}(\hat{\tau};\overline{V}_0) \leq -se^{-r(\check{T}(t)-\hat{\tau})}$, we have that $V_{i-1}(\hat{\tau};\overline{V}_0) - V_i(\hat{\tau};\overline{V}_0) \leq \frac{s}{\lambda_1}$ 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 $V_{i-1}(\tilde{t};\overline{V}_0) > \frac{s}{\lambda_1} + \frac{s}{r}\left(1-e^{-r(\check{T}(t)-\tilde{t})}\right)$. 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,

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

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

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

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

$$\left[\frac{r}{r+\lambda_1} + \frac{\lambda_1}{r+\lambda_1}e^{-(r+\lambda_1)(t_i^*-\hat{\tau})}\right]\left[V_{i-1}(\hat{\tau};\overline{V}_0) + \frac{s}{r}e^{-r(\check{T}(t)-\hat{\tau})}\right] - \frac{s}{r}e^{-(r+\lambda_1)(t_i^*-\hat{\tau})} - \frac{s}{\lambda_1} \geq 0. \quad \text{(A.6)}$$

[21] If $V_{i-1}(\tilde{t};\overline{V}_0) = \frac{s}{\lambda_1} + \frac{s}{r}\left(1-e^{-r(\check{T}(t)-\tilde{t})}\right)$, we have just argued that the value function is given by $V_i(\tilde{t};\overline{V}_0) = \frac{s}{r}\left(1-e^{-r(\check{T}(t)-\tilde{t})}\right)$. 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 $\hat{\tau} \leq t_i^*$, we have that $V_{i-1}(\hat{\tau}; \overline{V}_0) \geq \frac{s}{\lambda_1} + \frac{s}{r}\left(1 - e^{-r(\check{T}(t)-\hat{\tau})}\right)$, which implies that (A.6) holds, since

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

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

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

for $\hat{\tau} < t_i^*$. Hence, using $V_{i-1}(t_i^*; \overline{V}_0) = \frac{s}{\lambda_1} + \frac{s}{r}\left(1 - e^{-r(\check{T}(t)-t_i^*)}\right)$, 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 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

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

solve the HJB equation. Hence, V_i is indeed the value function. As, by induction hypothesis, $V_{i-1}(\cdot; \overline{V}_0)$ is absolutely continuous, and hence of bounded variation, it immediately follows that $V_i(\cdot; \overline{V}_0)$ is also of bounded variation, and hence absolutely continuous.

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

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

which completes the induction step.

Now, consider some $i \in \{1, \cdots, m-1\}$. Having established that the agent's best response is given by a cutoff strategy, I shall now show that $t_{i+1}^* \leq t_i^*$. Consider an arbitrary time $\tilde{t} \geq t_i^*$, 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

$$\int_{\tilde{t}}^{t^*} \lambda_1 \frac{s}{r} e^{-(r+\lambda_1)(\tau-\tilde{t})}\left(1 - e^{-r(\check{T}(t)-\tau)}\right) d\tau + \frac{s}{r} e^{-(r+\lambda_1)(t^*-\tilde{t})}\left(1 - e^{-r(\check{T}(t)-t^*)}\right)$$

$$= \frac{s}{r}\left[\frac{\lambda_1}{\lambda_1+r}\left(1 - e^{-(r+\lambda_1)(t^*-\tilde{t})}\right) - e^{-r(\check{T}(t)-\tilde{t})}\left(1 - e^{-\lambda_1(t^*-\tilde{t})}\right)\right] + \frac{s}{r} e^{-(r+\lambda_1)(t^*-\tilde{t})}\left(1 - e^{-r(\check{T}(t)-t^*)}\right).$$

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

$$1 - e^{-r(\check{T}(t)-\tilde{t})} > \frac{\lambda_1}{\lambda_1+r}\left(1 - e^{-(r+\lambda_1)(t^*-\tilde{t})}\right) - e^{-r(\check{T}(t)-\tilde{t})}\left(1 - e^{-\lambda_1(t^*-\tilde{t})}\right) + e^{-(r+\lambda_1)(t^*-\tilde{t})}\left(1 - e^{-r(\check{T}(t)-t^*)}\right)$$

$$\iff 1 > \frac{\lambda_1}{r+\lambda_1} + \frac{r}{r+\lambda_1} e^{-(r+\lambda_1)(t^*-\tilde{t})},$$

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 $i = 1, \cdots, m$, 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 $\overline{V}_0 \in (\frac{s}{\lambda_1}, \infty)$, 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 $\hat{\overline{V}}_0 > \overline{V}_0$,

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

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 $V_{i-1}(\tau; \hat{\overline{V}}_0) \longrightarrow \infty$ as $\hat{\overline{V}}_0 \longrightarrow \infty$ 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 $i = 1, \cdots, m$.

Suppose $t_{i+1}^*(\overline{V}_0) > t$. Then, $t_{i+1}^*(\overline{V}_0)$ is defined as the smallest root to $V_i(t_{i+1}^*; \overline{V}_0) = \frac{s}{\lambda_1} + \frac{s}{r}\left(1 - e^{-r(\check{T}(t) - t_{i+1}^*)}\right)$. 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, $V_i(t; \overline{V}_0) \leq \frac{s}{\lambda_1} + \frac{s}{r}\left(1 - e^{-r(\check{T}(t) - t)}\right)$. Let $j := \min\left\{\iota \in \{1, \cdots, m\} : t_\iota^*(\overline{V}_0) = t\right\}$. Since $t_1^* = \check{T}(t) > t$, 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 $V_{j-1}(t; \hat{\overline{V}}_0) > \frac{s}{\lambda_1} + \frac{s}{r}\left(1 - e^{-r(\check{T}(t) - t)}\right)$, 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 $i \in \{1, \cdots, m\}$. Let σ be defined by $\sigma := \max\left\{\iota \in \{1, \cdots, m\} : t_\iota^*(\overline{V}_0) > \tilde{t}\right\}$. 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 $V_\sigma(\tilde{t}; \hat{\overline{V}}_0) > \frac{s}{\lambda_1} + \frac{s}{r}\left(1 - e^{-r(\check{T}(t) - \tilde{t})}\right)$, and hence $t_{\sigma+1}^*(\hat{\overline{V}}_0) > \tilde{t}$. Iterated application of this argument yields the existence of a constant $\tilde{C}_{i-1}$ ($i \in \{1, \cdots, m\}$) 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 $\tilde{t} \in [t, \check{T}(t))$, $i \in \{1, \cdots, m\}$.

For $\tilde{t} \geq t_i^*$, we have that $V_i(\tilde{t}; \overline{V}_0) = \frac{s}{r}\left(1 - e^{-r(\check{T}(t) - \tilde{t})}\right) \leq V_{i-1}(\tilde{t}; \overline{V}_0)$. It remains to be shown

that for $\tilde{t} < t_i^*$, $V_i(\tilde{t}; \overline{V}_0) < V_{i-1}(\tilde{t}; \overline{V}_0)$. Since V_{i-1} is strictly decreasing, we have that

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

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

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

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

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

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

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.

Firstly, 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 $t < t_i^* < t_{i-1}^*$. Secondly, this immediately implies that if $t_i^* > t$, the equation $V_{i-1}(\hat{t}; \overline{V}_0) - \frac{s}{r}\left(1 - e^{-r(\check{T}(t)-\hat{t})}\right) - \frac{s}{\lambda_1} = 0$ 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 $\check{\overline{V}}_0 \in (\frac{s}{\lambda_1}, \infty)$ 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 $h(\overline{V}_0, \tilde{t}) := V_{i-1}(\tilde{t}; \overline{V}_0) - \frac{s}{r}\left(1 - e^{-r(\check{T}(t)-\tilde{t})}\right) - \frac{s}{\lambda_1}$; we note that h is continuous by induction hypothesis.

First, assume that $\check{\overline{V}}_0$ is such that $h(\check{\overline{V}}_0, t) < 0$. Since 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 have to 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 $\tilde{t} := t + \frac{\tilde{\epsilon}}{2}$. 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, $t_{i-1}^*(\check{\overline{V}}_0) > t_i^*(\check{\overline{V}}_0) > t$. Since 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 $\frac{\partial h}{\partial \tilde{t}}(\overline{V}_0, \tilde{t}) < 0$. (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 4.5

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 $V_i(\tilde{t}; \overline{V}_0; \check{T}(t))$ $(i = 1, \cdots, m)$ in $\check{T}(t)$ is immediately implied by the observation that for any fixed $\tilde{t} \leq \check{T}_1$ and $\overline{V}_0 > \frac{s}{\lambda_1}$, and given the end date $\check{T}_2 > \check{T}_1$, the agent can always guarantee himself a payoff of $V_i(\tilde{t}; \overline{V}_0; \check{T}_1) + \frac{s}{r}e^{-r(\check{T}_1-\tilde{t})}\left(1 - e^{-r(\check{T}_2-\check{T}_1)}\right) > V_i(\tilde{t}; \overline{V}_0; \check{T}_1)$ 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 we claimed.

It remains to prove continuity of $f(\cdot; \overline{V}_0)$. By Lemma 4.4, we have that

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

and that

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

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

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

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 $\check{T}(t)$.

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 4.4. 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 $\check{h}(\check{T}, \tilde{t}) := V_{i-1}(\tilde{t}; \overline{V}_0; \check{T}) - \frac{s}{r}\left(1 - e^{-r(\check{T}-\tilde{t})}\right) - \frac{s}{\lambda_1}$. We recall from our proof of Lemma 4.4 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

[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 non-differentiable versions of the theorem; here, one can for instance use the version in Kudryavtsev (2001).

increase $\check{T}$ to some arbitrary $\check{T}' > \check{T}$, V_{i-1} at $\tilde{t} \leq \check{T}$ increases by at least $\frac{s}{r}e^{-r(\check{T}-\tilde{t})}\left(1-e^{-r(\check{T}'-\check{T})}\right)$. Hence, we can conclude that $\check{h}(\cdot,\tilde{t})$ is weakly increasing.

First, assume that $\check{T}^*$ is such that $\check{h}(\check{T}^*,t) < 0$. Since $\check{h}$ 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 have to 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 $\check{h}(\check{T}^{**},\tilde{t}) = 0$. Since $\check{h}(\cdot,\tilde{t})$ 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 $\delta = \frac{\overline{T}-\check{T}^*}{2}$, will do.

Finally, suppose that $\check{h}(\check{T}^*,t) > 0$. In this case, $t_{i-1}^*(\check{T}^*) > t_i^*(\check{T}^*) > t$. Since 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 $\frac{\partial \check{h}}{\partial \tilde{t}}(\check{T},\tilde{t}) < 0$. As $\check{T}^*$ is an interior point (as $\check{h}(t,t) = -\frac{s}{\lambda_1} < 0$), we can again apply the Implicit Function Theorem to conclude that $t_i^*(\check{T})$ is continuous at $\check{T}^*$, since $t_i^*(\check{T})$ is defined by $\check{h}(\check{T},t_i^*(\check{T})) = 0$.

Thus, we have shown that, for all $i = 1,\cdots,m$, $t_i^*(\check{T})$ 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 $f(\check{T}(t);\overline{V}_0) = V_m(t;\overline{V}_0;\check{T}(t))$ is continuous in $\check{T}(t)$. ∎

Proof of Proposition 4.3

By Lemma 4.4, 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 (1.) follows.

With respect to statement (2.), we first choose some lump sum $\hat{\overline{V}}_0 > \frac{s}{\lambda_1}$ 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 analogous argument to above.) Continuity and monotonicity of f (see Lemma 4.5) now immediately imply the existence of some $\check{\check{T}}(t) \in (t,\check{T}(t))$ such that $w_t = f(\check{\check{T}}(t);\hat{\overline{V}}_0)$. ∎

Proof of Lemma 5.1

Since 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, the regularity conditions required for Filippov-Cesari's Existence Theorem (Thm. 8 in Seierstad & Sydsæter, 1987, p. 132) are satisfied.[23]

Clearly, $\check{\mathcal{U}} := \{(a,b) \in \mathbb{R}_+^2 : a+b \leq 1\}$ is closed, bounded and convex , the set of admissible

[23] See Note 17, p. 133, in Seierstad & Sydsæter, 1987, for a statement of the regularity conditions.

policies is non-empty, and the state variables are bounded. Using in addition the linearity of the objective and the laws of motion in the choice variables, one shows that the conditions of Filippov-Cesari's Theorem are satisfied. ∎

Proof of Proposition 5.2

Suppose that besides 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, which 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 $\hat{t} := \sup\left\{\bigcup_i (t_i^\dagger, t_i^\ddagger) : t_i^\dagger < t_i^\ddagger \text{ and } \hat{k}_{1,\tau} \neq 1 \text{ for a.a. } \tau \in (t_i^\dagger, t_i^\ddagger)\right\}$. Since $\hat{k}_{1,t} \neq 1$ on a set of positive measure, we have that $\hat{t} > 0$.

By (1) and the transversality condition $\hat{\mu}_T = \check{\mu}_T = 0$, we have that $\frac{e^{\hat{x}_{\hat{t}}}}{\hat{y}_{\hat{t}}}\hat{\mu}_{\hat{t}} = e^{\check{x}_{\hat{t}}}\check{\mu}_{\hat{t}}$; moreover, we know that, by Pontryagin's principle, the mappings $t \mapsto \frac{e^{\hat{x}_t}}{\hat{y}_t}\hat{\mu}_t$ 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

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

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

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

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

$$\lambda_1(h_{\hat{t}} + w_{\hat{t}}) - (1 + e^{\hat{x}_{\hat{t}}})s > \lambda_1(h_{\hat{t}} + w_{\hat{t}}) - (1 + e^{\check{x}_{\hat{t}}})s \geq -\lambda_1 e^{\check{x}_{\hat{t}} + r\hat{t}}\check{\mu}_{\hat{t}} = -\lambda_1 \frac{e^{\hat{x}_{\hat{t}} + r\hat{t}}}{\hat{y}_{\hat{t}}}\hat{\mu}_{\hat{t}}. \tag{A.7}$$

Since $(\hat{k}_{0,t}, \hat{k}_{1,t})_{0\leq t\leq T}$ satisfies Pontryagin's necessary conditions, and in particular condition (3), which implies that $\hat{k}_{0,t} + \hat{k}_{1,t} = 1$ at a.a. t at which it holds that

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

we can 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 (A.7) are continuous.

Furthermore, by conditions (1) and (2) and the transversality condition $\check{\mu}_T = \hat{\mu}_T = \check{\gamma}_T = \hat{\gamma}_T = 0$, we have that $-\lambda_0 e^{\hat{x}_{\hat{t}}}\hat{\gamma}_{\hat{t}} - \lambda_1 \frac{e^{\hat{x}_{\hat{t}}}}{\hat{y}_{\hat{t}}}\hat{\mu}_{\hat{t}} = -(\lambda_1 - \lambda_0)e^{\check{x}_{\hat{t}}}\check{\mu}_{\hat{t}}$. Again, by Pontryagin's conditions, the mapping $t \mapsto -\lambda_0 e^{\hat{x}_t}\hat{\gamma}_t - \lambda_1 \frac{e^{\hat{x}_t}}{\hat{y}_t}\hat{\mu}_t$ is continuous. Moreover, we have that

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

for all $t \in [\hat{t} - \frac{\eta}{2}, \hat{t}]$, with the first inequality being implied by Proposition 4.1. 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 $h_t + w_t \geq \frac{s}{\lambda_1}(1 + e^{x_0}) > 0$, and hence $h_t + \frac{s}{r}(1 - e^{-r\epsilon_t}) > 0$ for all $t \in [0, T]$. Hence, since $\hat{x}_t < \check{x}_t$, the second inequality also holds for all $t \in [\hat{t} - \frac{\eta}{2}, \hat{t}]$. 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 4.1 , furthermore implies that

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

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

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

This implies that

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

Since $(\hat{k}_{0,t}, \hat{k}_{1,t})_{0 \leq t \leq T}$ satisfies Pontryagin's necessary conditions, and in particular condition (3), we can conclude by continuity that $\hat{k}_{0,\tau} = 0$ for a.a. τ in some left-neighborhood of $\hat{t}$. Yet, since 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, which satisfies Pontryagin's conditions. ∎

Proof of Proposition 6.2

For $\frac{1}{p^m}$, 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 $\frac{\partial T^*}{\partial p_0}$ 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 $\frac{p_{T^*} - p^m}{p^m} = e^{\lambda_1 T^*} - 1$ is increasing in p_0. ∎

References

The Associated Press (December 23, 2005): "Korean Scientist Resigns Over Fake Stem Cell Research," available online e.g. at http://http://www.guardian.co.uk/science/2005/dec/23/stemcells.genetics (as of October 13, 2011).

Barraquer, T. R. and X. Tan (2011): "Conspicuous Scholarship," mimeo, Stanford University.

Bergemann, D. and U. Hege (2005): "The Financing of Innovation: Learning and Stopping," *RAND Journal of Economics*, 36, 719–752.

Bergemann, D. and U. Hege (1998): "Dynamic Venture Capital Financing, Learning and Moral Hazard," *Journal of Banking and Finance*, 22, 703–735.

Bergemann, D. and J. Välimäki (2008): "Bandit Problems," in: *The New Palgrave Dictionary of Economics*, 2nd edition ed. by S. Durlauf and L. Blume. Basingstoke and New York, Palgrave Macmillan Ltd.

Bergemann, D. and J. Välimäki (2000): "Experimentation in Markets," *Review of Economic Studies*, 67, 213–234.

Bergemann, D. and J. Välimäki (1996): "Learning And Strategic Pricing," *Econometrica*, 64, 1125–1149.

Bertsekas, D. (1995): *Dynamic Programming And Optimal Control(Vol. 1)*. Athena Scientific, Belmont, Massachusetts.

Bolton, P. and C. Harris (1999): "Strategic Experimentation," *Econometrica*, 67, 349–374.

Bolton, P. and C. Harris (2000): "Strategic Experimentation: the Undiscounted Case," in: *Incentives, Organizations and Public Economics – Papers in Honour of Sir James Mirrlees*, ed. by P.J. Hammond and G.D. Myles. Oxford: Oxford University Press, 53–68.

Bonatti, A. and J. Hörner (2011): "Collaborating," *American Economic Review*, 101(2), 632–663.

De Marzo, P. and Y. Sannikov (2008): "Learning in Dynamic Incentive Contracts," mimeo, Stanford University.

Fanelli, D. (2009): "How Many Scientists Fabricate And Falsify Research? A Systematic

Review and Meta-Analysis of Survey Data," *PLoS ONE*, 4, e5738.

Francis, B., I. Hasan and Z. Sharma (2009): "Do Incentives Create Innovation? Evidence from CEO Compensation Contracts," mimeo, Rensselaer Polytechnic Institute.

Gerardi, D. and L. Maestri (2008): "A Principal-Agent Model of Sequential Testing," Cowles Foundation Discussion Paper No. 1680.

Grossman, S. and O. Hart (1983): "An Analysis of the Principal-Agent Problem," *Econometrica*, 51, 7–45.

Holmström, B. (1989): "Agency Costs and Innovation," *Journal of Economic Behavior & Organization*, 12(3), 305–27.

Holmström, B. and P. Milgrom (1987): "Aggregation and Linearity in The Provision of Intertemporal Incentives," *Econometrica*, 55, 303–28.

Holmström, B. and P. Milgrom (1991): "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *The Journal of Law, Economics & Organization*, 7, Special Issue: 24–52.

Hörner, J. and L. Samuelson (2009): "Incentives for Experimenting Agents," Cowles Foundation Discussion Paper No. 1726.

Keller, G. and S. Rady (2010): "Strategic Experimentation with Poisson Bandits," *Theoretical Economics*, 5, 275–311.

Keller, G., S. Rady and M. Cripps (2005): "Strategic Experimentation with Exponential Bandits," *Econometrica*, 73, 39–68.

Klein, N. and S. Rady (2011): "Negatively Correlated Bandits," *Review of Economic Studies*, 78(2), 693–732.

Klein, N. (2011): "Strategic Learning in Teams," working paper, available at http://ssrn.com/abstract=1652832.

Kolata, G. (2009): "Grant System Leads Cancer Researchers to Play It Safe," *The New York Times*, June 28, 2009, available at http://www.nytimes.com/2009/06/28/health/research/28cancer.html?_r=1& pagewanted=all (as of October 13, 2011).

Kudryavtsev, L. D. (2001): "Implicit Function," in: *Encyclopedia of Mathematics*, ed. by M. Hazewinkel. Springer, available at eom.springer.de/i/i050310.htm (as of October 13, 2011).

Lerner, J. and J. Wulf (2007): "Innovation And Incentives: Evidence from Corporate

R&D," *Review of Economics and Statistics*, 89, 634–644.

Manso, G. (2011): "Motivating Innovation," *Journal of Finance*, forthcoming.

Murto, P. and J. Välimäki (2011): "Learning in a Model of Exit," *Review of Economic Studies*, forthcoming.

Presman, E.L. (1990): "Poisson Version of the Two-Armed Bandit Problem with Discounting," *Theory of Probability and its Applications*, 35, 307–317.

Rahman, D. (2010): "Detecting Profitable Deviations," mimeo, University of Minnesota.

Rahman, D. (2009): "Dynamic Implementation," mimeo, University of Minnesota.

Rosenberg, D., E. Solan and N. Vieille (2007): "Social Learning in One-Armed Bandit Problems," *Econometrica*, 75, 1591–1611.

Rothschild, M. (1974): "A Two-Armed Bandit Theory of Market Pricing," *Journal of Economic Theory*, 9, 185–202.

Seierstad, A. and K. Sydsæter (1987): *Optimal Control Theory With Economic Applications*. Elsevier Science.

Shan, Y. (2011): "Repeated Moral Hazard in Multi-Stage R & D Projects," mimeo, University of Iowa.