On Risk and Time Pressure:
When to Think and When to Do[*]

Christoph Carnehl[†]   Johannes Schneider[‡]

November 14, 2021

Abstract

We study the tradeoff between fundamental risk and time. A time-constrained agent has to solve a problem. She dynamically allocates effort between implementing a risky initial idea and exploring alternatives. Discovering an alternative implies progress that has to be converted to a solution. As time runs out, the chances of converting it in time shrink. We show that the agent may return to the initial idea after having left it in the past to explore alternatives. Our model helps explain so-called false starts. The agent takes risks early on to quickly arrive at a solution, sacrificing the prospects of alternatives.

JEL Classification: D01, D83, O31
Keywords: dynamic problem solving, endogenous bandits, time pressure

[†]Bocconi University, Department of Economics and IGIER; E-mail: `christoph.wolf@unibocconi.it`
[‡]University of Mannheim and Carlos III de Madrid; E-mail: `jschneid@econ.uc3m.es`

The people that really create the things that change this industry are both the 'thinker-doer' in one person.

—*Steve Jobs. Machine That Changed The World (1990).*

1   Introduction

Problem solving under time pressure is central to many economic problems. For example, consider an entrepreneur who has received funding for her venture. While the initial funds cover her expenses to advance the project for some time, she requires additional financing rounds before the venture becomes profitable.[1] To raise new funds, the entrepreneur has to achieve a milestone: build a prototype, solve a technological problem, or prove that a promising market for the product exists. Often, the entrepreneur faces an explicit or implicit deadline to achieve the milestone.[2]

Suppose that the entrepreneur proposed a particular strategy to reach a milestone—for example, a successful launch in a particular market. While the strategy appears promising, there is a risk that it is fundamentally flawed. However, there may be alternative ways to prove the venture's worth to investors. Invoking such an alternative would imply that the entrepreneur *pivots* away from her initial idea to a new strategy.

A successful pivot requires preparation. First, stakeholders need to be convinced that the change in strategy is promising (McDonald and Bremner, 2020). To this end, the entrepreneur could, for example, conduct customer research to discover market needs, leading to a new business strategy. When contemplating the option to pivot, our entrepreneur faces the following tradeoff (Isenberg and DiFiore, 2020): While preparing a pivot requires scarce time and resources, continuing with the current strategy entails the risk of failing.

When should the entrepreneur prepare a pivot? Early on, when time pressure is still low? Or later, after she has experimented unsuccessfully with the initial idea? When—if ever—should she abort the preparations for the pivot and set focus on the initial idea again?

In this paper, we study the novel tradeoff between fundamental risk and time in a multiarmed, continuous-time bandit model. We consider an agent (e.g., our entrepreneur) that has to solve a problem (e.g., achieve a milestone) by a deadline. Following the Steve Jobs quote at the beginning, the agent can exert effort in two different ways: She can *do*, or she can *think*.[3] Doing corresponds to working on implementing her initial idea. Thinking corresponds to preparing a pivot. While doing, a *solution* arrives stochastically at an ex

---

[1]Gompers (1995) discusses the importance of staged financing for venture capital-backed startups. The empirical analysis highlights that financing rounds are short. On average, they last only slightly longer than a year.

[2]Running out of cash is the second most frequent reason for startup failure according to CB Insights (2021). The entrepreneur has to achieve the milestone before her funds run out: an implicit deadline. Moreover, Kaplan and Strömberg (2003) documents that venture capitalists use both ex ante staging—committing to milestones that have to be achieved by a deadline—and ex post staging—liquidating the venture if the entrepreneur's performance is not satisfactory when the new funding round is due.

[3]Bolton and Faure-Grimaud (2009) consider a related but different tradeoff between thinking and doing. Importantly, and different from us, their tradeoff does not vary over time.

ante unknown rate. Instead, while thinking, *progress* arrives stochastically at a known rate. While a solution results in a fixed payment (e.g., the next round of funding), progress needs to be converted first. Because converting progress is harder with little time remaining, the value of progress is lower the closer the deadline is.[4]

A large body of literature in economics studies problem solving as the choice between solution methods. Following Rothschild (1974) and Weitzman (1979), this literature focuses on an exogenous set of known methods chosen over an infinite horizon.[5] However, classical experimentation models with an infinite horizon and an exogenous set of methods do not capture the time-risk tradeoff the entrepreneur faces. Therefore, these models do not provide answers to the questions posed.

In reality, deadlines matter. Kirtley and O'Mahony (2020) document, for example, the factors driving entrepreneurs to prepare a pivot. They show that both time pressure and beliefs about the feasibility of the initial idea play a crucial role in an entrepreneur's decisions. Therefore, understanding entrepreneurial choices on pivots in light of time pressure and risk requires explicit modeling of both a deadline and the generation of new strategies.

Our first contribution is that we characterize the agent's optimal policy in such a model. The characterization is nontrivial. The finite horizon precludes the use of index policies à la Gittins and Jones (1974). Moreover, due to the shortening time window, the value of progress changes over time—independent of the agent's actions. Finally, the presence of a bandit arm with a time-varying value implies that the optimal myopic policy need not be dynamically optimal.

If the agent is optimistic about her initial idea and the time window is large, then the optimal policy is as follows. The agent starts by doing. If doing remains unsuccessful, then the agent switches to thinking. However, if progress does not arrive in due time, then the agent switches back to doing and in the remainder of the time aims for a solution via the initial idea.

The property that the agent returns to a previously discarded arm results from three model ingredients: the finite horizon, the fact that progress needs to be converted in a second step, and the positive cost of effort. When any of these ingredients are dropped, the optimal strategy becomes a classical one-time switching rule. If the arm that the agent started with is not successful in due time, then she switches to the other but never switches back. To develop an intuition for the incentives at play in the optimal policy, we separately study two benchmarks—no time pressure and zero cost of effort.

If we drop time pressure, then the problem becomes a standard, recursive infinite-horizon problem.[6] Thinking becomes equivalent to the safe option. Doing, the risky option, offers

---

[4]There are several ways to model this decreasing value. Our baseline model depicts the value of progress in an abstract reduced form. In Section 4.3, we provide several examples to microfound this reduced form. These include, for example, safe and risky exponential bandit arms or an Ornstein-Uhlenbeck payoff stream.

[5]In Rothschild (1974)'s language, a method is an arm. In Weitzman (1979)'s language, a method is a box.

[6]This benchmark is analogous to another potential benchmark. Consider a classical experimentation problem with only one arrival needed on both arms but with the thinking arm having a lower intensity rate.

a cost advantage. Therefore, the agent first approaches the problem through doing. If no solution arrives, then she becomes pessimistic about the quality of her initial idea. She switches to thinking. Because time pressure is absent and the time to convert progress never runs out, the value of thinking is constant. When the agent finds it optimal to switch to thinking at one point, the thinking arm dominates the doing arm for the remainder of the time.

If we drop the cost of effort but keep the time window finite, then the result reverses. Thinking early on has a higher value than thinking later. Early progress leaves ample time for conversion. The agent starts by thinking. If progress remains absent, then she becomes pessimistic about having sufficient time left to convert progress should it arrive. She switches to doing in the hope of an immediate solution.

Our second contribution is to use our model to explain some entrepreneurial decision-making peculiarities that traditional bandit models cannot explain. First, the structure of our optimal policy provides a theoretical rationale for *false starts*. A false start describes the entrepreneur's tendency to act on initial ideas (doing) rather than to invest in customer research to explore alternatives (thinking). As proposed by Eisenmann (2021), false starts are one of the main reasons startups fail. We show that, indeed, by doing early, entrepreneurs trade off some overall success probability against saving time and the resources needed for early thinking.[7]

Second, perhaps surprisingly, an increase in the ex ante belief about the initial idea can lower the overall success probability of the entrepreneur. A higher initial belief discourages early customer research, thereby amplifying the false-starts problem. If the deadline is not too short, then this effect may dominate the positive effect of a higher likelihood that the initial idea can deliver a solution.

Third, *increasing* time pressure can incentivize the agent to think early on. A venture capitalist can mitigate the false start problem through tighter deadlines that crowd out the entrepreneur's incentive to do early. In general, early doing is attractive because it can provide a quick solution. The entrepreneur may opt for this route if initially, the time pressure is not too high. By increasing the initial time pressure, the venture capitalist discourages early acting on the initial idea. Instead, it encourages customer research when it is most valuable.

We believe that our modeling framework can also be applied to other contexts. While entrepreneurial problem solving serves as our main application, we discuss other applications in Section 6, our Final Remarks.

_____

Both benchmarks remove the time-varying value of progress that derives from its multistage nature and the finite horizon.

[7]One case in point is the Triangulate venture, as discussed in Eisenmann (2021). In a post-mortem of his failure, the funder Sunil Nagaraj admitted that he rushed to launch the venture's platform Wings rather than spending time on customer research to verify the market need for an improved matching engine. He considers that behavior as one of the main mistakes leading to the eventual failure of Triangulate.

Related Literature. We contribute to a large body of literature that deals with the choice between approaches to innovation. One strand of the literature dating back to Weitzman (1979) has considered several variants of Pandora's box problem as a proxy for finding the right innovation strategy (e.g., Fershtman and Rubinstein, 1997; Olszewski and Weber, 2015; Doval, 2018). Other works have been concerned with how competition affects the search for the right approach (e.g., Aghion et al., 2001; Akcigit and Liu, 2015; Letina, 2016; Lemus and Temnyalov, 2019). Our contribution to this literature is that we endogenize the available approaches by giving the agent the option to explore an alternative route. Therefore, we also endogenize the cost of finding an alternative. We capture two aspects absent in the aforementioned literature. First, new ideas may be arrived at stochastically, and the cost the agent incurs to make progress on the alternative varies with the time it takes to find it. Second, and more importantly, the value of progress varies with the time window the agent has to convert progress into a solution. Thus, the *availability* of an alternative route depends on both luck and choices in the past, and the *value of discovering* an alternative depends on the time window left.

There is a strand of the management literature that addresses issues similar to ours. An example is Gans, Stern, and Wu (2019). However, their *Test Two, Choose One* result ignores the time dimension, which is the focus of our paper. The process of how to think about alternatives and the particularities of lean techniques are discussed in Felin et al. (2019). Our model provides a formal, economic method for these ideas.

Technically, our model falls into the class of multiarmed bandit problems (Rothschild, 1974). Bergemann and Välimäki (2008) provides an overview of the literature. The doing arm is a classical continuous-time exponential bandit, as used in most of the strategic experimentation literature (Keller, Rady, and Cripps, 2005). While most models feature infinite-horizon settings, we are interested in a time-constrained agent. Klein (2016) also considers a time-constrained agent. The crucial difference from our model is that in his case, both arms are exponential bandits that differ in their intensities, not in the number of arrivals needed.

In our case, the thinking arm is *restless*—the state of the arm evolves even when not pulled. The restless feature is essential for capturing the risk-time tradeoff but makes the model complicated (see, e.g., Fryer and Harms, 2019). The thinking arm concept is related to the few papers in the literature that study multistage bandits (Keller and Oldale, 2003; Hu, 2014; Green and Taylor, 2016; Wolf, 2018; Kim, 2021; Moroni, 2021).

The results in Kim (2021) appear reminiscent of ours, yet the environment and thus the mechanism differ substantially. We study an agent's decision whether to experiment with a risky project or take a step back and look for a safer alternative absent agency concerns. Kim (2021) instead abstracts from risk and focuses on the agency problem alone. In his model, multistage projects benefit an infinitely lived principal, as intermediate reports provide a monitoring tool. Absent agency concerns, multistage projects are inefficient and are never chosen. In our setting, it is a priori unknown which approach is efficient, and the

agent balances learning and managing time pressure.

Callander (2011), Garfagnini and Strulovici (2016), Francetich (2018), Nikandrova and Pancs (2018), and Che and Mierendorff (2019) study the problem of dynamically distributing effort across several projects. However, none of them addresses the risk-time tradeoff. Francetich (2018) studies the choice of allocating effort to two correlated bandits, Callander (2011) features myopic agents, and Garfagnini and Strulovici (2016) considers a two-period overlapping-generations model. By construction, neither reproduces the switching dynamics we obtain. Nikandrova and Pancs (2018) model an agent who irreversibly selects between two alternatives. The agent uses experimentation to learn about her options beforehand. In Che and Mierendorff (2019), an agent can try to find a solution or to show that no solution exists. At the optimum, the agent uses only one of the available routes until she reaches an absorbing state. All five models are related in spirit, but the research question, the modeling choices, and the results are different.

The two most closely related papers to ours are Bolton and Faure-Grimaud (2009) and Fershtman and Pavan (2021). Most notably, however, both of their models operate with an infinite time horizon and thus cannot produce the time-pressure dynamics of our finite-horizon problem. In their models, the tradeoff between arms is independent of calendar time. In our model, the available time window and associated time pressure constantly change, affecting the main tradeoff.

Bolton and Faure-Grimaud (2009) are also interested in the choice between thinking and doing. However, thinking plays a different role in their model. It is a tool to resolve uncertainties regarding future choices. In contrast, thinking in our model corresponds to the development of a new, previously unavailable route.

Similar to our work, Fershtman and Pavan (2021) study a model with endogenous arms. They consider an agent with an infinite horizon who decides whether to apply her initial idea or to search for alternatives. Their focus is on the search process itself. The agent can at a cost investigate several routes and learn about their quality. Instead, we focus on the risk-time tradeoff that the agent faces as the available time window closes. To gain tractability in a finite-horizon world, we abstract from some details of the search process. We collapse it into a unidimensional object. Because different approaches to the problem are taken, the results also differ. While Fershtman and Pavan (2021) show that in their setting, an index policy remains optimal, we show that the same does not hold when the risk-time tradeoff plays a role.

Roadmap. We set up our model in Section 2. We derive our main results in Section 3. Section 4 describes a set of economic implications derived from our findings. In Section 5, we discuss our modeling choices. Finally, Section 6 concludes and provides an outlook on other applications of our framework.

2 Model

We introduce a model capturing the risk-time tradeoff between solving a problem through *doing*—a fast but fundamentally risky approach—or through *thinking*—a slower but less risky approach. We provide a discussion of our assumptions and examples nested in the model after the analysis in Section 5.

Setting. Time is continuous and starts at $t = 0$. An agent has to solve a problem by a finite deadline $T < \infty$ and there is no discounting over time. At each instant of time, the agent can invest one divisible unit of effort into *doing*, a_t^d, and *thinking*, a_t^s, such that $a_t^d + a_t^s \leq 1$. Investing effort entails a flow cost of $(a_t^d + a_t^s)c$, with $c > 0$.

One arm, the *doing* arm, is risky. The arrival of a solution on that arm depends on the unobserved binary state $\theta \in \{0, 1\}$. The instantaneous arrival rate of the arm is $\theta\lambda$, where $\lambda > 0$, which implies the following: if $\theta = 1$, then the probability that a solution arrives when the agent invests a_t^d over a small time interval $[t, t + dt)$ is $\lambda a_t^d dt$; if $\theta = 0$, then a solution never arrives on the doing arm. The agent's belief that $\theta = 1$ at $t = 0$ is $\bar{p} \in (0, 1)$. A solution delivers a payoff of $B > 0$ to the agent and ends the game.

The other arm, the *thinking* arm, has a known instantaneous arrival rate $\mu > 0$. An arrival on the thinking arm at time t implies *progress*. Progress does not provide a solution directly. Instead, the agent has to convert progress into a solution that requires additional time and effort. We capture this second step in reduced form. The function $V(\tau)$ describes the agent's continuation payoff when progress occurs with time $\tau = T - t$ remaining to the deadline. We assume that $V(\cdot)$ is thrice continuously differentiable, increasing, i.e., that $V'(\cdot) > 0$, and sufficiently concave, i.e., that $-V''(\cdot)/V'(\cdot) \geq \bar{p}\lambda$. Moreover, an arrival on the thinking arm with no time remaining is worthless, $V(0) = 0$, and (abusing notation) $V(\infty) := \lim_{\tau \to \infty} V(\tau) > c/\mu$.[8]

To streamline the intuition, we interpret progress in our main analysis as the first step in a multistage problem. We imagine that after progress occurs, the agent works on converting progress into a solution. We provide detailed examples that microfound $V(\tau)$ for this interpretation and others in Section 4.3. However, formally, as with the doing arm, an arrival on the thinking arm ends the game.

Throughout this paper, we are interested in how the agent allocates effort *between* doing and thinking. We focus on cases in which the agent finds it optimal to exert full effort until the end of the game. We thus consider cases in which the solution yields a high reward; that is, we assume that B is sufficiently large. It is straightforward to show that such a reward B exists; Appendix D.4 provides the respective argument. A large B allows us to restrict attention to the agent's choice *between arms*, as it implies $a_t^d = 1 - a_t^s$. Dropping the

---

[8]The last condition states that the expected value of progress without time pressure is larger than the expected effort cost required to obtain progress. The following additional technical condition strengthens the concavity assumption: $\lim_{\tau \to \infty} V'(\tau)$ and $\lim_{\tau \to \infty} V''(\tau)$ exist with $\lim_{\tau \to \infty} -V''(\tau)/V'(\tau) \geq \bar{p}\lambda$. We need it for a few of our results.

superscript and denoting time variables in terms of the time remaining τ, we use $a_\tau \equiv a^d_{T-t}$.

3   Analysis

In this part, we characterize the agent's optimal policy. We begin by describing two benchmark results. After that, we provide an interpretable necessary condition for the agent's optimal policy. Finally, we derive an algorithm that characterizes the unique optimal solution under a mild technical assumption.

3.1   Benchmarks

The thinking arm has two essential features that distinguish it from the doing arm: (i) it becomes increasingly unattractive as the deadline approaches, and (ii) its expected payoff is different from that of the doing arm. The source of the first feature lies in the conversion of progress into a solution. Completing the additional steps in time becomes increasingly unlikely when less time remains. The source of the second feature is twofold. First, the expected effort cost until progress arrives, c/μ, may be different from the expected effort cost until a solution of a good ($\theta = 1$) doing arm arrives, c/λ. Second, the value of progress—even without time pressure—may be different from the value of a solution on the doing arm, $V(\infty) \neq B$. A reason for the latter is that upon progress, the agent has to exert additional time and effort to convert progress into a solution.

Our first benchmark ('no time pressure') shuts down the first channel, and our second benchmark ('no payoff difference') shuts down the second channel.

No time pressure.   Facing an infinite time horizon, $T = \infty$, the thinking arm is a safe alternative for the agent. Thinking long enough guarantees progress and thus *some* payoff. The following proposition describes our first benchmark. We relegate its proof along with all other proofs to the appendix.

Proposition 1. *Suppose that the time horizon is infinite, $T = \infty$. Then, the agent either works first on the doing arm and eventually switches to the thinking arm or works on the thinking arm throughout. The agent starts with doing if and only if*

$$\bar{p} \geq \hat{p} := \frac{c/\lambda}{B - V(\infty) + c/\mu}.$$

She switches when $p_t = \hat{p}$, that is, at time $t = \overline{\tau}_1 := \max\left\{\frac{1}{\lambda} \ln\left(\frac{\bar{p}(1-\hat{p})}{\hat{p}(1-\bar{p})}\right), 0\right\}$.

Proposition 1 shows that an agent who is sufficiently optimistic about the doing arm starts to work on it. An initial belief $\bar{p} \in (0,1)$ for such optimism exists if and only if the expected payoff from a good doing arm is higher than that from thinking, $B - c/\lambda > V(\infty) - c/\mu$.

Two effects push the agent towards an initial doing period: lower expected cost and larger expected benefits. The first effect is present if a solution arrives faster than progress

through thinking, $\lambda > \mu$. The second effect is present if a solution provides a higher payoff than progress even in the absence of time pressure $B > V(\infty)$. Higher payoffs from the doing arm may, for example, result from saving the additional effort required to convert progress.

No payoff difference. In this benchmark, we assume that in the absence of time pressure, there is no payoff difference between (productive) arms; that is, we assume

$$\lim_{\tau \to \infty} V(\tau) = B \qquad\qquad c = 0. \qquad\qquad \text{(C.1)}$$

Condition (C.1) ensures (i) that absent time pressure, successful thinking delivers no better or worse solution than doing, and (ii) that there is no difference in the expected cost of obtaining a solution and of obtaining and converting progress. This assumption allows us to focus exclusively on the role of time pressure. The following proposition describes the optimal policy in this case.

Proposition 2. *Suppose condition* (C.1) *holds. The agent either works first on the thinking arm and eventually switches to the doing arm or works on the doing arm throughout. The agent starts with thinking if and only if the deadline T is large enough such that a solution $\tau_3 \in (0, T]$ to*

$$\bar{p} = \frac{\mu V(\tau_3)}{B\left(\mu + (\lambda - \mu)e^{-\lambda \tau_3}\right)}$$

exists. In this case, the agent switches to doing when the time remaining τ_3 is equal to the smallest of the solutions.

The intuition behind Proposition 2 is the following: If the deadline is close, even if progress arrives momentarily, then the time left to convert it is short. The payoff $V(\tau)$ vanishes fast in τ. An arrival on the doing arm, instead, delivers a solution directly. As time runs out, the time pressure effect on the thinking arm trumps any fundamental uncertainty on the doing arm. The agent pulls the doing arm and throws a *Hail Mary*.[9]

Why and when to do? Proposition 2 shows that the payoff motive behind Proposition 1 is not the only reason for doing: if time pressure is high, then gambling on risk, i.e., having a good arm, is more promising than gambling on time, i.e., managing to convert progress.

Therefore, our benchmarks offer a first insight into why the agent opts for doing: (i) to materialize the payoff advantage of the doing arm[10] and (ii) to succumb to time pressure.

---

[9]The term originates from American Football. In 1975, Dallas Cowboys quarterback Roger Staubach threw a 50-yard pass in the final seconds of a game, desperately hoping to make the game-winning touchdown. Staubach commented that while throwing the ball, he "closed [his] eyes and said a Hail Mary". Since then, throwing a Hail Mary has become synonymous with taking a risky action in desperation, often because time is nearly expired.

[10]Recall that this payoff advantage can derive from effort-saving motives due to a faster arrival on the doing arm, from effort-saving motives due to not requiring an additional implementation stage, or from a higher payoff from a solution on the doing arm rather than from completed conversion of progress.

Our next step is to combine the two motives considering a setting with $T < \infty$ and $c > 0$ for which the optimal allocation of effort is yet to be determined.

3.2 Optimal Policy

We characterize the optimal policy in three steps. First, we state the agent's dynamic optimization problem. Second, we derive a set of necessary conditions for the optimal policy. These conditions have a straightforward economic interpretation that we discuss. Third, we state an algorithm which—under mild technical conditions—determines the uniquely optimal policy. The third step verifies the sufficiency of the necessary conditions.

The Agent's Problem

By construction, the agent exerts full effort until the game ends. However, she dynamically decides whether to invest in thinking or in doing. Consider a situation in which the remaining time is τ, and in which the agent holds a belief p_τ about the doing arm. Suppose the agent exerts effort a_τ on the doing arm for a small time interval of length dt. The instantaneous payoff of a solution is B. A solution arrives with probability $a_\tau p_\tau \lambda dt$. Suppose the agent exerts effort $1 - a_\tau$ into the thinking arm for a small time interval of length dt. The instantaneous payoff of progress is $V(\tau)$. Progress arrives with probability $(1 - a_\tau)\mu dt$.

The agent updates her belief about the doing arm according to Bayes' rule. Denote by $A_\tau := \int_\tau^T a_s ds$ the amount of effort the agent has invested in the doing arm in the past. Then, the belief about the doing arm with time τ remaining and past effort A_τ on the doing arm is

$$p_\tau = \frac{\bar{p}e^{-\lambda A_\tau}}{\bar{p}e^{-\lambda A_\tau} + (1 - \bar{p})}.$$

If no arrival, i.e., neither a solution nor progress, occurs during the interval $[\tau, \tau - dt)$, then the payoff of progress declines to $V(\tau - dt)$—regardless of the agent's choice a_τ. The belief, however, declines only when the doing arm was pulled with positive intensity, $a_\tau > 0$.[11]

The agent's objective is to dynamically maximize

$$\max_{(a_\tau)_{\tau=0}^T} \int_0^T \underbrace{e^{-\mu(T-\tau-A_\tau)}}_{P(\text{no progress yet})} \underbrace{\left(1 - \bar{p} + \bar{p}e^{-\lambda A_\tau}\right)}_{P(\text{no solution yet})} \underbrace{\left(\mu(1-a_\tau)V(\tau) + \lambda a_\tau p_\tau B\right)}_{\text{flow payoff}} dt$$

where the mapping $a_\tau : [0, T] \times [0, T - \tau] \to [0, 1]$ determines the strategy with time τ remaining and past effort A_τ on the doing arm. We use A_τ as the state variable to derive the necessary conditions for an optimal strategy via optimal control methods. The formal details can be found in Appendix A. We derive the following dynamic relative preference for the agent:

---

[11]In particular, the belief follows the standard ODE $dp_\tau/d\tau = p_\tau(1 - p_\tau)a_\tau\lambda$, where, again, the notation follows the time remaining rather than the calendar time.

$$\gamma_\tau = e^{-\mu(T-\tau-A_\tau)}\left(\left(1 - \bar{p} + \bar{p}e^{-\lambda A_\tau}\right)\mu V(\tau) - \bar{p}e^{-\lambda A_\tau}\lambda B\right) - \eta_\tau$$

where η_τ denotes the co-state of the optimal control problem. If $\gamma_\tau > 0$, then the agent pulls the thinking arm. If $\gamma_\tau < 0$, then the agent pulls the doing arm.

The co-state η_τ is determined by the boundary condition $\eta_0 = 0$ and its evolution

$$\frac{d\eta_\tau}{d\tau} = e^{-\mu(T-\tau-A_\tau)}\left(\mu(1-\bar{p})\left((1-a_\tau)\mu V(\tau) - c\right) \right.$$

$$\left. - (\lambda - \mu)e^{-\lambda A_\tau}\bar{p}\left((1-a_\tau)\mu V(\tau) + a_\tau \lambda B - c\right)\right). \tag{1}$$

Necessary Conditions

We derive the necessary conditions for the optimal policy from Pontryagin's principle. These necessary conditions substantially reduce the space of the candidate strategies.

Proposition 3 (Optimal Policy—Necessary Conditions)**.** *The optimal policy takes one of the following forms:*

1. *the agent exclusively uses the doing arm,*
2. *the agent starts by thinking and switches to the doing arm, or*
3. *the agent begins with the doing arm, switches to the thinking arm eventually, and switches back to the doing arm when little time remains.*

The critical insight leading to Proposition 3 is that if the agent leaves the thinking arm once, she does not return to it. At a high level, the intuition behind this insight is the following. If the agent decides to leave the thinking arm, then only because she considers the value of progress to be too low due to the deadline approaching. Importantly, the decline in the value of progress does not stop—even when pulling the doing arm.

While this observation is a substantial part of the story, it falls short in one aspect: Whenever the agent pulls the doing arm unsuccessfully, the payoff of that arm also declines because the belief about its state deteriorates.

The precise intuition behind the horse race of the two arms is subtle. A stepwise inspection of the effects at play is instructive. Consider the following equivalent formulation of the relative preference from above:

$$\gamma_\tau = e^{-\mu(T-\tau-A_\tau)}\left(1 - \bar{p} + \bar{p}e^{-\lambda A_\tau}\right)\overbrace{\underbrace{(\mu V(\tau) - p_\tau \lambda B)}_{\text{payoff difference}} - \underbrace{\eta_\tau}_{\substack{\text{effect of lower belief} \\ \text{on continuation value}}}}^{=:y_\tau}.$$

The agent thinks whenever $\gamma_\tau > 0$. We focus on the second part, y_τ, and consider an increase in the time remaining. Using the evolution of η_τ (equation (1)), we obtain

$$\frac{dy_\tau}{d\tau} = \underbrace{\mu V'(\tau)}_{\substack{\text{(i) deadline effect}}} + \underbrace{\mu p_\tau \lambda (V(\tau) - B)}_{\substack{\text{(ii) payoff-on-arrival} \\ \text{effect}}} + \underbrace{(\mu - \lambda p_\tau) c}_{\substack{\text{(iii) effort-to-arrival} \\ \text{effect}}} . \tag{2}$$

Observe that equation (2) is independent of the agent's action a_τ; that is, the agent's action has no first-order effect on the evolution of the relative preference. Any effect that the agent's current action has on instantaneous payoffs is compensated by a dynamic effect in the continuation value—a feature common in the bandit literature.[12]

As we increase the time to the deadline, equation equation (2) describes the three incentives that determine the change in the agent's relative preference: (i) the change in the value of progress due to the reduced time pressure (the deadline effect); (ii) the payoff differential between the two arms upon an arrival (payoff-on-arrival effect); and (iii) the change in the difference in the expected effort required to reach progress or a solution (effort-to-arrival effect).

The deadline effect is the only effect that can be signed unambiguously and is always positive—pushing the agent towards the thinking arm. As the deadline moves further away, obtaining progress has a higher value. The additional time makes it more likely to convert progress before the deadline. The other two effects can be either positive or negative. The sign of the payoff-on-arrival effect depends on the relative payoffs between arms. If $V(\tau) < B$, then the effect is negative—pushing the agent towards doing. As the deadline moves further away, a negative payoff-on-arrival effect pushes the agent to spend the additional time on the doing arm. The effort-to-arrival effect measures the relative expected cost difference of the arms. Thinking is expected to deliver an arrival faster than doing if $\mu > p_\tau \lambda$. In this case, there is a positive effort-to-arrival effect—pushing the agent towards thinking.

We derive the intuition for our critical insight—the agent pulls the thinking arm in at most one connected interval of time—from equation (2) and the illustrated effects. We use the following construction. Suppose that the agent splits her thinking effort into two disjoint intervals of time. Then, there has to be a doing interval that is both preceded and succeeded by a thinking period. This implies that (2) is positive at the beginning of this doing period, i.e., when τ is high. In this case, an expanding time window pushes the agent toward thinking. At the same time, (2) is negative toward the end of the doing period, i.e., when τ is low. An expanding time window pushes the agent toward doing.

To satisfy this property, (2) has to change signs during the doing interval from negative to positive as τ increases; thus, (2) has to cross zero from below. Formally, the agent's relative preference must attain an interior minimum during the doing interval. Economically, this implies that an increase in the time remaining—while the agent is pulling the doing

---

[12]In Appendix B we provide a derivation of (2) illustrating how the direct effect of the action drops out. Note, however, that there is a second-order effect of the agent's action through the belief p_τ.

arm—must change its effect on the relative preference: from pushing the agent further toward the doing arm to pushing her back toward the thinking arm. At such a minimum, the three effects must exactly balance each other. In particular, the positive deadline effect must be compensated for by the sum of the payoff-on-arrival and effort-to-arrival effects.

Observe that the deadline effect declines as the time remaining increases, $V''(\tau) < 0$. Thus, time pressure becomes less of a consideration in the agent's decision. One force pulling the agent to the thinking arm becomes weaker as τ increases. Recall that the supposed strategy requires the relative preference to attain an interior minimum. For this to occur, the payoff-on-arrival and effort-to-arrival effects combined (i) must pull the agent sufficiently toward the doing arm to compensate for the deadline effect and (ii) must evolve sufficiently in favor of the thinking arm to dominate the decline in the deadline effect. However, both properties can never be satisfied simultaneously if the decline in the deadline effect is sufficiently strong. In particular, they cannot be satisfied if $V(\tau)$ is sufficiently concave: $-V''(\tau) \geq p_\tau \lambda V'(\tau)$. Thus, any doing period is either preceded or succeeded by a thinking period but not both.[13]

Intuitively, the agent will not think twice because of the following observation: If the agent, at some point, finds the switch *doing → thinking* optimal, then she cannot have found the switch *thinking → doing* optimal when she had more time remaining. To see this, note that at the switch *doing → thinking*, the deadline effect must be weak because the agent found it optimal to do *before* that switch. However, as we move backward in time, i.e., increase the time remaining, the deadline effect weakens. If it weakens sufficiently fast, then the agent is pulled more toward doing, the further away the deadline is. Thus, it is never optimal for the agent to think twice. We provide and interpret examples that satisfy this property in Section 4.3.

Proposition 3 therefore follows from these two observations: (i) the agent never returns to the thinking arm and (ii) a Hail Mary is inevitable once the agent runs out of time, which delivers Proposition 3. Either the agent starts with the Hail Mary or has at most one thinking period that precedes the Hail Mary period.

The necessary conditions provide substantial structure as they limit the set of possible solutions. In the following subsection, we derive an intuitive algorithm to compute the optimal policy based on the necessary conditions. Under a mild additional assumption, the algorithm delivers the unique solution to the agent's problem. Moreover, it provides further intuition for the economics of the agent's problem and comparative statics.

---

[13]For the special case $\mu > \bar{p}\lambda$, observe that the effort-to-arrival effect is always positive. Moreover, the payoff-on-arrival effect has to be negative, $V(\tau) < B$, in any doing period that succeeds a thinking period. Thus, (2) increases in the time remaining only if $V'(\tau) + \lambda p_\tau V(\tau)$ increases. Given our concavity assumption, this is not the case. Hence, (2) is nonincreasing and y_τ concave.

3.3 Characterization of the Optimal Policy

The basic intuition for the optimal policy follows from combining the two benchmarks in the previous section: The agent may do when her belief about the doing arm is high, and the deadline is far as in Proposition 1. Moreover, the agent will do—independent of her belief about the doing arm—when the deadline is close as in Proposition 2.

We now construct a solution algorithm that, under the following assumption, delivers the unique solution to the agent's problem.

Assumption 1.

(i) $V(\infty) \leq B + \frac{c}{\mu}$.

(ii) If $\mu > \lambda$, then $\frac{\mu V''(\tau)}{(\mu-\lambda)U''(\tau)} - \hat{q}(\tau)$ is monotonic.

Assumption 1 is only a sufficient condition to ensure that a unique strategy satisfies the necessary conditions. It is a technical and by no means a necessary condition.[14] The first part of the assumption implies that the doing arm is sufficiently attractive to consider it a valuable arm beyond the Hail Mary period.

The interpretation of the second part of the assumption is somewhat more subtle. Note first that condition (ii) is only relevant if progress on the thinking arm is expected to arrive faster than a solution on the doing arm conditional on the doing arm being good. In such a case, the condition ensures that $q(\tau)$ is monotonic by requiring that the curvature of the arms' values is sufficiently regular.

Our algorithm constructs the optimal solution by working backwards from the Hail Mary period. As we show in the appendix, condition (ii) guarantees that the length of the Hail Mary period is continuous and monotonic in the belief that the agent holds about the doing arm at the beginning of that period. This observation allows us to use marginal arguments to show uniqueness. Invoking Proposition 3, we state the optimal policy in terms of three variables

1. the time spent in the Hail Mary period, τ_3,
2. the time spent in the thinking period, τ_2, and
3. the time spent in the initial doing period, τ_1.

We are looking for a solution to the equation $\tau_1 + \tau_2 + \tau_3 = T$. To construct it, we make use of the following expressions:

$$q(\tau) := \min \left\{ 1, \underbrace{\frac{\mu \left(V(\tau) + c\tau\right)}{\mu(B + c\tau) + (\lambda - \mu)\left(B - \left(1 - e^{-\lambda\tau}\right)\left(B - \frac{c}{\lambda}\right)\right)}}_{=:\hat{q}(\tau)} \right\}$$

---

[14]The crucial aspects for uniqueness are (i) that the belief at the beginning of the Hail Mary period is monotonic in the length of that period and (ii) that the maximum length of the thinking period is monotonic in the length of the Hail Mary period that follows. Both aspects must be true 'in the relevant regions.' Our assumptions on primitives ensure that they are universally true. Moreover, uniqueness only facilitates the computation. If it fails, then our algorithm can be straightforwardly extended to determine all candidate solutions which then have to be compared to determine the global solution. In particular, all solution candidates must satisfy Proposition 3.

$$\dot{y}(s;p,\xi) := \left.\frac{dy_{\xi+s}}{ds}\right|_{y_\xi=0} = \mu V'(\xi + s) + p\mu\lambda(V(s + \xi) - B) + (\mu - \lambda p)c, \text{ and}$$

$$\hat{y}(\tau;p,\xi) := \int_0^\tau e^{\mu s}\dot{y}(s;p,\xi)ds.$$

The first, $q(\tau)$, is the agent's belief when entering the Hail Mary period with time τ remaining. It originates from the agent's indifference between entering the Hail Mary period immediately and pulling the thinking arm for an infinitesimal measure of time before entering the Hail Mary period.

The second, $\dot{y}(s;p,\xi)$, describes the change in the relative preference due to a marginal increase in the deadline of an agent who enters a Hail Mary period with belief p and time remaining ξ and who pulls the thinking arm during the remaining times $[\xi + s, \xi)$ before switching to the doing arm with time remaining ξ.

The third, $\hat{y}(\tau;p,\xi)$, describes the value of the relative preference with deadline $T = \tau + \xi$ by integrating over the survival-weighted evolution of the relative preference assuming indifference at the start of the Hail Mary period $\hat{y}(0;p,\xi) = 0$.

We use these expressions to define the length of the initial doing period and the length of the thinking period as a function of the length of the final doing period:

$$\tau_1(\tau_3) := \frac{1}{\lambda}\ln\left(\frac{\bar{p}}{1-\bar{p}}\frac{1-q(\tau_3)}{q(\tau_3)}\right), \text{ and}$$

$$\tau_2(\tau_3) := \begin{cases} \min \tau > 0 \text{ s.t. } \hat{y}(\tau;q(\tau_3),\tau_3) = 0, & \text{if a root for } y \text{ given } \tau_3 \text{ exists,} \\ \infty & \text{otherwise.} \end{cases}$$

The first, $\tau_1(\tau_3)$, follows because the belief when entering the Hail Mary period is determined by the time spent doing in the initial doing period and by Bayes' rule.

The second, $\tau_2(\tau_3)$, follows because indifference is necessary when switching from doing to thinking for the first time and when switching back.

We are now ready to state our algorithm that solves the fixed-point problem $T = \tau_1(\tau_3) + \tau_2(\tau_3) + \tau_3$ and thereby characterizes the agent's optimal strategy. We provide a further discussion of the algorithm after stating the characterization result.[15]

Algorithm.

1. Set $\tau_1 = \tau_2 = \tau_3 = 0$.
2. Find the largest $\bar{\tau}_3$ such that

$$\forall t \in [0, \bar{\tau}_3] \qquad q(\bar{\tau}_3 - t) \leq \frac{\bar{p}e^{-\lambda t}}{(\bar{p}e^{-\lambda t} + 1 - \bar{p})}.$$

---

[15]A MATLAB program implementing the algorithm is available from the authors.
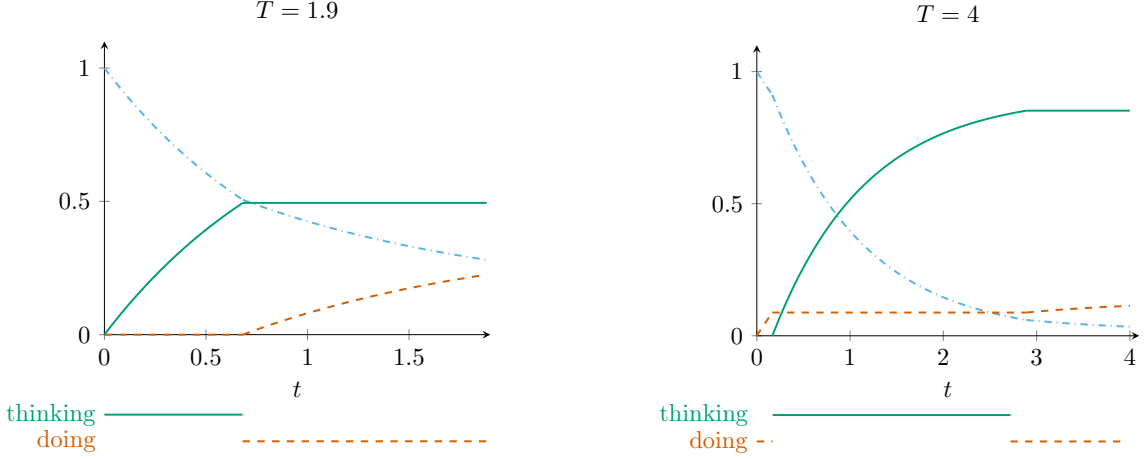
Figure 1: *Agent's optimal strategy and arrival probabilities.* The solid line plots the probability that the agent has made progress by time t, the dashed line the probability that the agent has found a solution through the doing arm by t, and the dash-dotted line the probability that the agent has neither made progress nor found a solution by t. Below we plot the time intervals in which the agent thinks or does absent any arrival. The left panel considers the optimal strategy given deadline $T = 1.9$, and the right panel considers the optimal strategy given deadline $T = 4$.
Parameters: $B = 5, \bar{p} = 3/4, c = 1/2, \lambda = 3/4, \mu = 1, V(\tau) = (1 - e^{-\tau})(B - c)$.

> If $\bar{\tau}_3 \geq T$, set $\tau_3 = T$, $\tau_2 = \tau_1 = 0$ and stop.

3. If $q(\bar{\tau}_3) \neq \bar{p}$ go to 5.

4. If $\tau_2(\bar{\tau}_3) \geq T - \bar{\tau}_3$, set $\tau_3 = \bar{\tau}_3$ and $\tau_2 = T - \bar{\tau}_3$ and stop.

5. Replace $\bar{\tau}_3$ by the largest $\bar{\tau}_3$ such that

$$\forall t \in [0, \bar{\tau}_3] \qquad q(\bar{\tau}_3 - t) \leq \frac{q(\bar{\tau}_3)e^{-\lambda t}}{q(\bar{\tau}_3)e^{-\lambda t} + 1 - q(\bar{\tau}_3)}.$$

6. Set $\tau_3 = \bar{\tau}_3$, $\tau_1 = \tau_1(\bar{\tau}_3)$ and $\tau_2 = \tau_2(\bar{\tau}_3)$. If $\tau_1(\bar{\tau}_3) + \tau_2(\bar{\tau}_3) + \bar{\tau}_3 = T$, stop. Otherwise, reduce $\bar{\tau}_3$ marginally and repeat 6.

Proposition 4 (Optimal Policy – Characterization). *Under Assumption 1, the above algorithm determines the unique optimal policy.*

To build an intuition for the algorithm, recall that Proposition 3 implies that it is without loss of generality to focus on three disjoint time intervals to characterize the solution: the Hail Mary period of length τ_3, the thinking period of length τ_2, and the initial doing period of length τ_1.

The algorithm constructs the solution via backward induction. It takes advantage of the property that the three time periods have to sum to the total time available to the deadline T—thereby constraining each other.

Given any length of the final doing period, τ_3, we can determine the agent's belief upon entering the Hail Mary period. There are two cases: either (i) the agent enters the Hail Mary period immediately with belief $\bar{p}$, or (ii) she enters the Hail Mary period after at least one period of thinking. If, in addition, she enters the Hail Mary period with a belief $\hat{p} < \bar{p}$,

then she must have pulled the doing arm before she started thinking.

First, in the case of (i), we need to ensure that the agent never finds it optimal to think for some positive measure of time until the deadline. To ensure this, the agent has to be sufficiently optimistic about the doing arm for any remaining time $\tau < T$. Item 2 finds the largest deadline $T = \bar{\tau}_3$ such that the agent is sufficiently optimistic about pulling the doing arm throughout. It takes both declining beliefs and declining time windows into account. If $T \le \bar{\tau}_3$, then the algorithm has found a solution.

Second, in the case of (ii), we need to ensure that the agent finds it optimal to switch to doing at the designated time τ_3 and not to switch back to thinking thereafter. Thus, in addition to satisfying $q(\tau) < p_\tau$ for all $\tau < \tau_3$, we require $q(\tau_3) = p_{\tau_3}$. The latter ensures that the agent is indifferent between the arms when time τ_3 remains. Optimality requires that the agent indeed prefers to think before the Hail Mary period. Using backward induction again—conditional on switching to doing at time τ_3—the agent finds it optimal to think with remaining time $\tau + \tau_3$ if and only if the expression $\hat{y}(\tau'; q(\tau_3), \tau_3) \ge 0$ holds for all $\tau' \in [0, \tau]$. Conditional optimality follows because $\dot{y}(s; q(\tau_3), \tau_3)$ describes the evolution of the agent's relative preference between the arms—derived from the necessary conditions of the optimal control problem—assuming that she switches to the doing arm with time τ_3 remaining. Item 4 of the algorithm stops if $\hat{y}(\tau; q(\tau_3), \tau_3) \ge 0$ for all $\tau \in [0, T - \tau_3]$: the agent finds it optimal to start by thinking for a period of time $\tau_2 = T - \tau_3$.

Third, if the agent engages in an initial doing period, then we must be in case (ii). We know that the initial doing period determines the agent's belief for the final doing period via Bayes' rule as a function of the length of the initial doing period, τ_1, and the agent's ex ante belief, $\bar{p}$. The expression $\tau_1(\tau_3)$ describes the time that the agent has to experiment without success on the doing arm such that her belief deteriorates to $q(\tau_3)$.

Fourth, the agent needs to be indifferent both after the initial doing period and when starting the final doing period. Whenever the agent finds it optimal to spend time $\tau_2 = T - \tau_1(\tau_3) - \tau_3$ in the thinking period, $y(T - \tau_1(\tau_3); q(\tau_3), \tau_3) = 0$. If the expression $\tau_2(\tau_3) > 0$, then that indifference is guaranteed with time remaining $\tau_2 + \tau_3$. If, instead, $\tau_2(\tau_3) = \infty$, then it is never optimal to leave the initial doing period with a belief $q(\tau_3)$. Item 6 of the algorithm stops only if all conditions are met and thus determines the fixed point $T = \tau_1(\tau_3) + \tau_2(\tau_3) + \tau_3$.

4 Application: Entrepreneurial Problem Solving

Our model highlights the tradeoff that an agent faces under time pressure: should she try to apply an uncertain method ready at hand, or should she take a step back and develop a different method that involves less fundamental uncertainty? The optimal strategy is a function of both the time horizon and her belief about the initial method. Our results emphasize how learning and time pressure interact.

We now apply our findings to our motivating application—an entrepreneur's decision

to meet a target to secure follow-up financing by a deadline. First, we relate the model to the specific context of entrepreneurs who need to achieve a milestone to obtain the next round of financing. Second, we derive implications from our main theoretical results in light of the application considered and relate it to empirical phenomena. Third, we discuss several microfoundations for the value of progress $V(\tau)$ in the context of entrepreneurial decision-making.

4.1 Entrepreneurial Problem Solving

Consider an entrepreneur who has raised funding for her venture. She has to prove the business's prospects by some deadline. Deadlines are ubiquitous in innovative entrepreneurship. Among many other reasons, they may come from (i) funders explicitly setting deadlines (e.g., via staged contracts, see Kaplan and Strömberg, 2003), (ii) the need to raise new funds before the startup runs out of cash,[16] or (iii) implicitly, according to the expectation that the market moves on after some time either by changing focus or by adopting a competitor's product.[17]

The flow cost of working on the problem, c, has two interpretations in the context of startups. The first interpretation is literal and derives from what is called the startup's *burn rate.* Each period, the startup has to pay its employees, rent an office or lab space, purchase equipment, etc. If the startup has access to initial funds C, then the burn rate implies an implicit deadline $T = C/c$ by which it has to have raised new funds.[18] Under this interpretation, the entrepreneur wants to complete the task with funds remaining in her pocket. She can invest these leftover funds in later stages. The second interpretation is to consider c as a (linear) time cost of the agent. Such an interpretation is proposed by Eisenmann (2021). He argues that entrepreneurs have a direct time cost in the form of an action bias and prefer to get things done as quickly as possible.

Our focus is on the entrepreneur's approach in trying to meet the requirements before a deadline, for example, successfully launching a product, developing a new product, improving an existing product, or meeting a revenue threshold. While the entrepreneur raises funding with an initial idea she may pursue—pulling the doing arm—, she can also attempt to pivot to a different approach to meet the target by the deadline—pulling the thinking arm. Investing time and effort into a change of the startup's strategy is commonly observed (see,

---

[16]See CB Insights (2021)—running out of cash is the most frequent reason startups fail based on CB Insights' analysis of startup failure post-mortems.

[17]The drone analytics provider, Airware, went out of business because, initially, they bundled their software with a self-engineered drone. However, once they were ready to launch the bundled product, cheaper alternative drones were already available. They pivoted to focusing on software development only but ran out of money and eventually ceased operations. Perhaps more famously, despite being a corporate favorite, Blackberry failed to innovate until it missed its deadline and the market had moved to iOS and Android. See also the discussion in Gans, Scott, and Stern (2018).

[18]The fact that we use continuous time with a bounded per-unit effort strengthens this interpretation. Instead of seeing the deadline as a clock ticking, we could interpret T as the total effort budget available. Investing cdt units of effort on doing or thinking from this budget implies an arrival with rates $p_\tau \lambda dt$ and μdt, respectively. The value of progress depends on the effort remaining within the budget. Having invested T without a solution makes the agent perish.

for example, Kirtley and O'Mahony, 2020) but has received little theoretical attention. While the doing arm resembles the standard experimentation approach, which is commonly used in modeling entrepreneurial strategy (for an overview, see Kerr, Nanda, and Rhodes-Kropf, 2014), the thinking arm captures that the value of new ideas depends on the resources and time available to convert them into solutions. In the following subsection, we provide several interpretations of the thinking arm in the spirit of our application.

4.2 Implications

Both beliefs and the time horizon matter for entrepreneurs when contemplating how to invest resources in their venture (see, for example, Kirtley and O'Mahony, 2020; Rahmani and Ramachandran, 2021). Moreover, entrepreneurs have a tendency to *do* early and "get things done" (see, for example, Gans, Scott, and Stern, 2018; Eisenmann, 2021). This observation is in line with our finding that the agent starts by doing if she is sufficiently optimistic about her initial approach and the time pressure is not too high initially (see Proposition 3).

However, this strategy comes at a cost: The entrepreneur reduces the expected time to solve the current problem by doing early. At the same time, this strategy reduces the probability of solving the problem in time. The reason is that hoping for an early solution produces a *false start* (Eisenmann, 2021): Doing early delays thinking about a pivot—e.g., shifting business to accommodate a different market—yet pivots occur with positive probability. If the entrepreneur ends up pivoting, then she suffers from the shrunk time window.

To formally address false starts, we need to put a structure on the thinking arm. We wish to compare how the agent trades off the expected effort cost against the probability of finding a solution. Therefore, we need to take a stance on *how* the agent converts progress into a solution. For clarity, here, we restrict attention to a straightforward example.

Example 0. An arrival on the doing arm delivers a new Poisson bandit with a known arrival rate $\nu \geq \lambda \overline{p}$. An arrival on this new arm implies a solution worth B to the agent, and the cost of pulling the arm is c.

Note that there is nothing inherently special about Example 0. For a more detailed discussion on this, including other examples in the context of startups, see Section 5. Example 1 in that section subsumes Example 0.

As a first result, we see that thinking early and backloading doing improves the ex ante probability of obtaining a solution in Example 0.

Proposition 5. *Consider Example 0. For any potential strategy inducing $\tau_1 > 0, \tau_2 > 0$, and $\tau_3 > 0$, backloading all effort on the doing arm, i.e., choosing $\tau_1' = 0, \tau_2' > \tau_2$, and $\tau_3' = \tau_1 + \tau_3$, increases the probability that a solution is found by the deadline.*
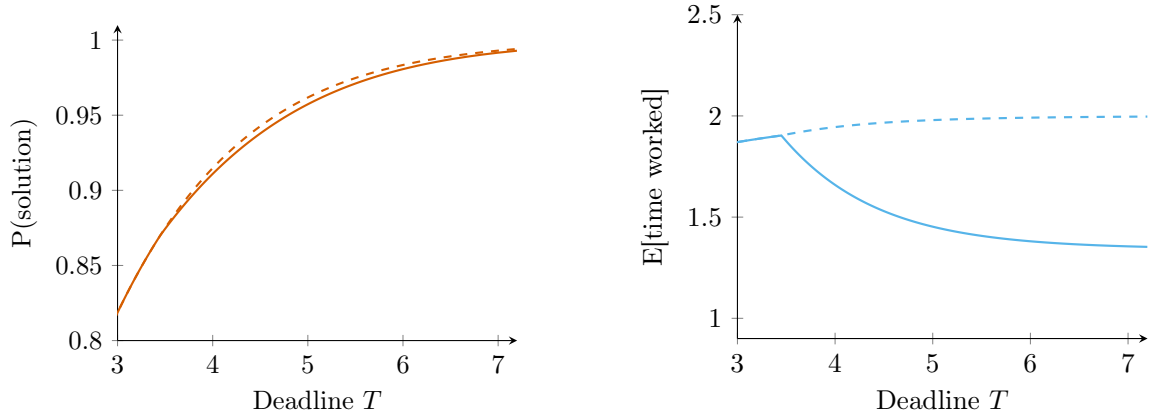
Figure 2: *Solution probability vs. cost reduction.* The left panel plots the ex ante probability of finding a solution against the ex ante deadline length.
The right panel plots the expected time the agent works.
Solid lines represent these under the agent-optimal strategy, and dashed lines represent these using the same total amounts of doing and thinking but backloading all doing. For deadlines shorter than the depicted range, the agent enters the Hail Mary period immediately, and the two curves coincide.
Parameters: $B = 5, \bar{p} = 3/4, c = 1/2, \lambda = 3/4, \mu = \nu = 1$.

Proposition 5 gives a theoretical foundation for the empirical phenomenon of *false starts*—the delay or entire absence of customer research before launching a minimum viable product. Whenever the entrepreneur starts with her initial idea right away, i.e., does early on, she sacrifices success probability—unless the initial deadline is very short. To see why this occurs, observe that it is straightforward to rewrite the agent's problem as

$$\max_{\mathbf{a}:=(a_t)_{t=0}^T} \mathbb{P}^{\mathbf{a}} \left[\text{solution before } T\right] B - \mathbb{E}^{\mathbf{a}} \left[\text{time worked}\right] c.$$

This rewriting makes it apparent that the agent wants to balance the probability of success against the expected time to solve the problem.

Within our model, the agent aims to reduce the expected effort because it is costly. In reality, there are multiple underpinnings for such effort cost: among these is a direct disutility of effort, the desire to save funds for the future, or a bias for moving forward fast with the venture. Figure 2 highlights the consequences. The larger the time horizon is, the more the agent saves on her expected effort. Perhaps surprisingly, the time the agent expects to work can decline in the deadline length. The reason is straightforward: the agent adjusts her strategy to do early. If doing is successful, then she finishes earlier, which, in turn, reduces the expected effort invested. However, the agent's investment choices come at the cost of reducing the expected probability of succeeding at all.

From a venture capitalist's perspective, the return of marginally expanding the entrepreneur's deadline may thus not fully translate into an increase in the probability of finding a solution—even though we expect the agent to work absent a solution. Instead, the entrepreneur may sacrifice some of the extra potential to arrive at a solution faster.

Proposition 6. *The length of the initial doing period, τ_1, and the length of the thinking*
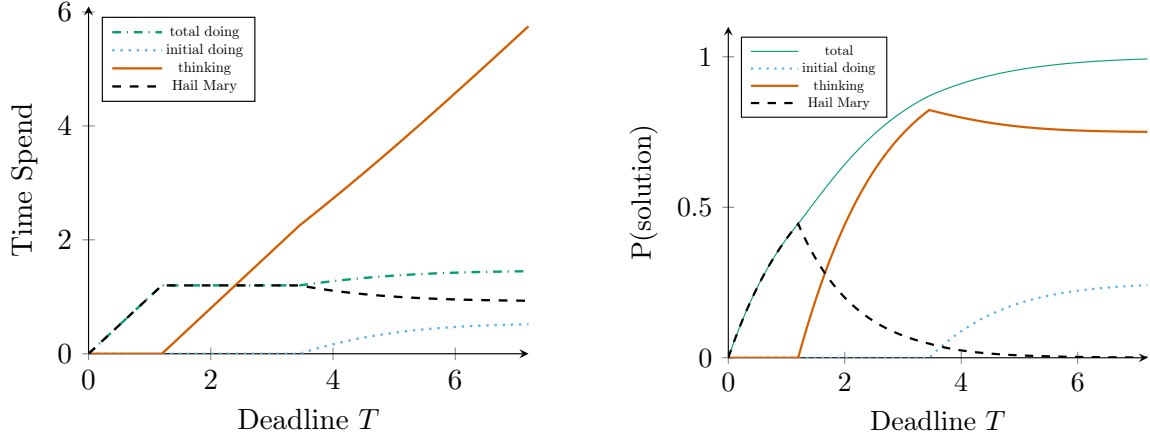
Figure 3: *Time spent in different periods (left), and probability of a solution by period (right) for different ex ante deadlines T.* The left panel plots the time spent in different phases under the optimal strategy against the initial time horizon: initial doing (dotted), thinking (solid), and Hail Mary (dashed). The dash-dotted line depicts the maximum time spent on the doing arm.

The right panel plots the ex ante probabilities of obtaining a solution by phase against the initial time horizon: initial doing (dotted), Hail Mary (dashed), and thinking—progress & conversion (thick solid). The thin solid line is the aggregate probability that a solution occurs (the sum of the other three curves).

Note: This figure compares different ex ante time horizons and must not be confused with the agent's decision over time. Parameters: $B = 5, \bar{p} = 3/4, c = 1/2, \lambda = 3/4, \mu = \nu = 1$.

period, τ_2, are nondecreasing in T. As $T \to \infty$, $\tau_2 \to \infty$.

Proposition 6 shows that in the beginning, the agent never decreases the time devoted to doing. As the left panel of Figure 3 suggests, the frontloading of doing and thus false starts become a larger problem with longer deadlines. An increase in the deadline can, therefore, not improve the induction of false starts. However, once the deadline offered is long enough, the problem becomes second-order: τ_2 becomes arbitrarily large, and the probability to obtain some solution converges to 1.

Depending on the product at hand, the venture capitalist may not only care about the entrepreneur finding a solution but also benefit from potential externalities depending on the solution method. For example, venture capitalists may benefit from customer research for other projects. In particular, if—as Gompers (1995) and Kaplan and Strömberg (2003) suggests—the entrepreneur's motivation within a financing stage comes mainly from meeting the explicit requirements, i.e., from solving the problem in time, then the venture capitalist's primary instrument is to expand or tighten the duration of the stage.

As we see in the right panel of Figure 3, tightening the entrepreneur's deadline may increase the probability that she finds her solution using customer research. The reason is that with low time pressure, the entrepreneur spends large portions of the extra time gambling on a quick and successful launch of her initial minimum viable product. Even if she fails initially, she remains confident that there is enough time to generate and convert insights from customer research. With a tighter deadline, the same entrepreneur engages in customer research earlier.

A direct consequence of this observation is that a venture capitalist may provide deadlines that are shorter than his actual time horizon to induce—perhaps surprisingly—a more thorough approach by the entrepreneur. The deadline discourages the entrepreneur from gambling on quick successes and incentivizes her to begin the project with customer research.[19]

Unsurprisingly, the agent's time spent on each approach is a function of the agent's initial belief. For example, suppose the entrepreneur is pessimistic that a launch will succeed without additional insights from customer research. In this case, she is not willing to launch it—unless the time window is small—and rather engages in customer research first. Instead, if the entrepreneur is optimistic, then she tries launching first to save her effort on customer research.

Indeed, as the following proposition shows, if the initial belief $\bar{p}$ is large, then—independent of the deadline—the agent never begins by thinking. At the same time, if the initial belief is low, then the agent only starts with doing when under immediate time pressure.

Recall $\hat{p}$ from Proposition 1,

$$\hat{p} = \frac{c/\lambda}{B - (V(\infty) - c/\mu)} = \frac{\mu\nu}{\lambda(\mu + \nu)}$$

which, in Example 0, is independent of B and c.[20]

Proposition 7. *Consider Example 0. Fix B, c, λ, μ, and V such that Assumption 1 holds. Independent of the time horizon T, the following statements hold:*

1. there is a $\tilde{p}$ such that if $\bar{p} > \tilde{p}$, then the agent begins with a doing period and $\tilde{p}$ solves

$$\tilde{p} = \frac{V'(q^{-1}(\tilde{p})) + c}{\lambda(B + c/\mu - V(q^{-1}(\tilde{p})))};$$

2. if $\bar{p} < \hat{p}$, then the agent switches arms at most once and only from thinking to doing;

3. if $\bar{p} \geq \hat{p}$, then the agent's belief never falls below

$$\check{p} := \frac{\hat{p}e^{-\lambda q^{-1}(\hat{p})}}{\hat{p}e^{-\lambda q^{-1}(\hat{p})} + 1 - \hat{p}} \geq \frac{\bar{p}e^{-\lambda T}}{\bar{p}e^{-\lambda T} + 1 - \bar{p}}.$$

Proposition 7 provides insights into the potential for the venture capitalist who receives a payoff $\Pi > 0$ if the entrepreneur successfully launches the product in some way before a deadline T^{VC}.

Note that the venture capitalist's first best is identical to that derived in Proposition 2. The venture capitalist does not incur the entrepreneur's cost. Whether he can implement his

---

[19]Interestingly, external risk, described as risks equally uncertain to both the venture capitalist and the entrepreneur—e.g., future demand for an undeveloped product—significantly lowers the time until the subsequent financing round in Kaplan and Strömberg (2003). Our results provide one mechanism that can rationalize this observation: when external risk is high, the value of customer research is high. Venture capitalists can encourage early customer research with intermediate deadlines: long enough to prevent an immediate Hail Mary but short enough to discourage a false start.

[20]The proof of Proposition 7 also makes the proposition applicable outside of Example 0.

first best depends on whether he can design a contract (B, T) such that the agent switches once and at the right time.

It is trivial that the venture capitalist can induce his first best if his preferred strategy is to throw the Hail Mary throughout, i. e., if his own deadline T^{VC} is relatively short. However, for larger deadlines, he has to find a payment B such that the entrepreneur's optimal switching time $q^{-1}(\bar{p})$ coincides with his optimal switching time τ^{VC}. It turns out that such a payment may not exist. For example, for the case $\mu \geq \lambda = \nu = 1$, no such B exists for any $\bar{p}, B > 0$, and $c > 0$.[21]

If the venture capitalist cannot control B, for example, because the entrepreneur is motivated by success rather than by payments from the venture capitalist and does not face immense time pressure, then obtaining the venture capitalist's first best reduces at most to a nongeneric coincidence. Whenever the belief about the entrepreneur's initial idea is too high, $\bar{p} > \tilde{p}$, then achieving it with any deadline is impossible.

Proposition 7 admits the following corollary which has further implications on how venture capitalists can use time pressure to induce entrepreneurs to exert desired actions.

Corollary 1. *The agent's belief during the thinking phase is larger than* $\min\{\hat{p}, \bar{p}\}$. *If* $\bar{p} \leq \hat{p}$, *then the probability that the agent solves the problem through the doing arm is maximized with deadline* $T_1 = q^{-1}(\bar{p})$.

The corollary states that the agent is constrained in the amount of experimentation she is willing to exert by her option to instead think.

The corollary is, for example, relevant in the following setting. Suppose that doing corresponds to the entrepreneur trying to launch a product in a particular business-to-business (B2B) context. In contrast, thinking corresponds to exploring potential direct-to-consumer (D2C) markets to which the entrepreneur can pivot with her product. Suppose further that it is known that B2B is not the ideal market but that it is uncertain *which* D2C market is the right market. To resolve this uncertainty, the entrepreneur needs to carry out customer research. However, the venture capitalist may be interested in entering the B2B market to establish his reputation and therefore may have the preference (subject to success) that the entrepreneur launches in the B2B market.

Because the entrepreneur can pivot, she is not exploring all the options to launch in the B2B market. By imposing time pressure on the entrepreneur, the venture capitalist can maximize the chances of entering the B2B market.[22]

---

[21]The venture capitalist's switching time, τ, solves $\bar{p} = \frac{\mu V^{VC}(\tau)}{\Pi(\mu + (\lambda - \mu)e^{-\nu\tau})} = \frac{\mu(1 - e^{-\nu\tau})}{(\mu + (\lambda - \mu)e^{-\lambda\tau})}$;

the entrepreneur's switching time, τ, solves $\bar{p} = \frac{\mu\left((1 - e^{-\nu\tau})(B - c/\nu) + c\tau\right)}{\mu(B + c\tau) + (\lambda - \mu)\left(B - (1 - e^{-\lambda\tau})(B - c/\lambda)\right)}$. If $c > 0$, for example, then both equations cannot hold for any B if $\mu \geq \nu = \lambda = 1$.

[22]Although the possibility result stated in Corollary 1 relies on the fact that a pessimistic entrepreneur is not going to have an initial doing phase, it is often the case that even if an infinite deadline involves an initial doing phase, the likelihood of obtaining a solution through the thinking arm is maximized with $T = T_1$; see, e.g., the configuration in the left panel of Figure 3: Even as $T \to \infty$, the dash-dotted line will not be higher than at the interior maximum.
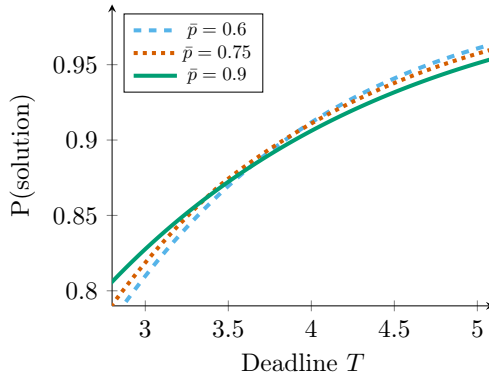
Figure 4: *Solution probability for different initial beliefs $\bar{p}$ against the initial time horizon.* Parameters: $B = 5, \bar{p} = 3/4, c = 1/2, \lambda = 3/4, \mu = \nu = 1$.

Although one may conjecture that higher initial beliefs $\bar{p}$ always increase the ex ante probability of obtaining a solution, this need not be the case. A higher initial belief may push the entrepreneur toward increasing her initial efforts to obtain a solution through the doing arm and to save on the cost of effort. This, in turn, may lower the probability of obtaining a solution in the given time frame. See Figure 4 for an example.[23]

We want to stress that our findings in this section are in stark contrast to those from a classical infinite horizon two-armed bandit model in which both arms have time-independent payoffs and one arm has a lower arrival rate. For both the deadline and the initial belief, such a model predicts nondecreasing probabilities of solving the task with a particular arm. Therefore, our model highlights a significant economic incentive that would be absent in a model that does not feature a time dependence of the thinking arm's payoff: the deadline regulates not only the overall success probability but also how success is achieved.[24]

4.3 Examples

The crucial feature of the thinking arm is that the value of progress depends on the time remaining until the deadline. To illustrate the model's flexibility in the context of startups, we provide a set of examples nested in our model. We provide a formal verification that each example meets our assumptions in Appendix D.1. We provide examples *not* nested within our framework in the subsequent discussion in Section 5 and show whether and how the result extends.

Implementation Phase or Delay upon Progress

In our first set of examples, progress triggers a new bandit arm that can be pulled following progress. Here, pulling the thinking arm can, e.g., correspond to customer research as in the

---

[23]We cannot derive meaningful conditions for when that nonmonotonicity occurs, but numerically it appears relatively robust.

[24]Indeed, it may also happen that an increase in the deadline reduces the probability of a solution through the doing arm—see, e.g., the left panel of Figure 3. Thus, an increase in the deadline—holding the initial belief about the doing arm fixed—increases the probability of a *pivot*.

application in Section 4.2. In contrast, pulling the doing arm corresponds to immediately launching the minimum viable product without customer research. These two approaches resemble the discussion in Eisenmann (2021), in particular, the discussion of the Triangulate venture and the choices of its founder. For the new doing arm, Examples 1-4 provide particular models that fall into this category.

Example 1. The simplest case is when successful customer research, i.e., the arrival of progress on the thinking arm, delivers a new product design, i.e., a new approach corresponding to a Poisson bandit arm, that can be marketed with certainty—given enough time—with arrival rate $\nu \geq \lambda \bar{p}$. The payoff of a solution with this new arm is B_ν, and the cost of pulling it is c_ν. This new arm may either be an improved version of the original idea or an entirely new product inducing a *pivot* of the entrepreneur.

Example 2. This example is in a similar vein. However, successful customer research may not remove all uncertainty about the feasibility of a new product idea. Instead, the new product may deliver better prospects than the original product, but some uncertainty remains. The new arm is then characterized by $\bar{p}^\nu$, the ex ante belief about the new arm's feasibility, and by ν, the new arm's arrival rate. An arrival on this arm provides value B_ν, and pulling it has a flow cost of c_ν.

Example 3. Our model also nests cases in which successful thinking triggers a new arm that has a time-varying intensity rate $\nu(t)$. Successful customer research suggests a product that is quite different from the original idea and the entrepreneur's experience. The time-varying intensity can go both ways. On the one hand, there can be an increasing intensity: initially, the agent may not be very knowledgeable about the new approach, but her knowledge improves over time. On the other hand, there can be a decreasing intensity: the customer research may demonstrate some immediate implications expected to work with high probability. However, if these initial attempts fail, then it becomes harder to succeed. For concreteness, suppose the intensity rate follows an exponential function $\nu(t) = \nu e^{\alpha + \beta t}$, with $\alpha \geq 0$ and $\nu B > c$.[25]

Example 4. Finally, consider a model in which customer research provides enough evidence for the entrepreneur to raise an additional round of financing. However, it takes an unknown amount of time to analyze and polish the data, prepare a convincing pitch, or find the right venture capitalist. In such a model, the problem is solved by progress on the thinking arm, but the payoff is realized only after some random delay, which follows a Poisson process with intensity rate $\nu \geq \bar{p}\lambda$. Suppose that the entrepreneur has to incur a flow cost of c_ν until its arrival (for example, to pay her employees). Moreover, the project's

---

[25]While we do not restrict the sign of β to allow for both an increasing ($\beta > 0$) and a decreasing ($\beta < 0$) intensity rate, we impose a lower bound on $\beta \geq -\frac{1}{\tau}$, which implies $\nu(t)B \geq c$ given $\alpha \geq 0$. In particular, $\alpha \geq \ln\left(\frac{c}{\nu B}\right) - \beta\tau$, where the right-hand side is strictly negative for all $\beta \geq 1/\tau$, as $\nu B > c$ by assumption.

payoff is B_ν if the success arrives before the deadline and 0 if not. This example is formally equivalent to Example 1.

Abandoning the Doing Arm upon Progress.   In the above examples, the entrepreneur abandons the doing arm upon progress on the thinking arm. While this is an endogenous outcome in Example 1 and Example 4, we assume it for Example 2 and Example 3. We want to highlight that this assumption can be justified in several ways. We comment on each version separately below.

1. The new arm replaces the initial doing arm.
2. The belief about the new arm is high.
3. There is a maintenance cost of keeping the old doing arm idle.
4. There is a high switching cost to return to the old doing arm.

The first interpretation of such an arm is that its arrival *replaces* the old arm. For example, customer research verifies that the initial hypothesis of the entrepreneur is indeed true and that it is possible to implement it successfully.

The second interpretation leaves the initial doing arm unaffected, but the entrepreneur endogenously will not return to it. For example, the entrepreneur has found a new, ex ante better-suited market for her product. Suppose her belief is sufficiently high that a successful launch in that market is feasible. In this case, she will not switch to the old market before the deadline arrives, even if the attempts on the new market have been unsuccessful throughout.

The third interpretation assumes that the entrepreneur can either pay a maintenance cost to keep the first doing arm idle or abandon it entirely at no additional cost. For example, the entrepreneur detects another way to launch her product. To do so, she has to hire new expertise. Keeping the old expertise within the company is costly. Forand (2015) shows that for the case $\nu = \lambda$, *any* such maintenance cost would imply that the entrepreneur abandons the old arm for good.[26] Even if $\nu \neq \lambda$, the intuition continues to hold at least for a sufficiently large maintenance cost.

The fourth interpretation is similar to the third. Instead of a cost that keeps one arm idle, it is costly for the entrepreneur to return to the old arm. Pivoting away from the initial plan requires consent from stakeholders, is thus public, and leads to a deteriorated reputation of the initial idea; see, e.g., McDonald and Bremner (2020). Indeed, we are not aware of a single case in which a startup pivots back to the initial strategy after having pivoted to a new strategy.

---

[26]The result follows from Propositions 2 and 3 in Forand (2015) taking the limit of $r \to 0$ and noting that at the time of arrival, we must have $\bar{p}^\nu > p_\tau$. Otherwise, the entrepreneur had never pulled the thinking arm for a positive measure immediately before the arrival of progress.

Payoff Stream upon Progress

An alternative model is one in which successful thinking triggers a payoff stream. For example, instead of attempting to build a new prototype with an advanced technology (i.e., continuing to pull the doing arm), the entrepreneur finds a way to market the product as is without further improvements. This market opportunity generates a revenue stream until the deadline. We impose that once the entrepreneur uses this market opportunity, she cannot return to the initial doing arm. For example, it requires specific skills to advance the current technology, and rehiring that expertise is nearly impossible. An entrepreneur in this framework resembles one that raises funding from investors applying the scorecard method.[27] Under the scorecard method, an entrepreneur can invest in either improving its product (pulling the doing arm) or bringing a new product to the market to increase sales (pulling the thinking arm to find the right way to launch the product and collecting the revenue stream upon progress).

Example 5. We model the launch of a product as paying out $db(t)$ in the interval $[t, t + dt)$ and assume that the payments follow an Ornstein-Uhlenbeck process:[28] $db(t) = \nu(B_\nu - b(t))dt + \sigma dW_t$, where W_t is a standard Brownian motion and the initial value is $b(0) = 0$. The payoff $db(t)$ can be thought of as a flow profit that has B_ν as its long-run expectation. $\nu \geq \bar{p}\lambda$ is the rate of mean reversion of the profit process.

5 Discussion of Modeling Choices

In this section, we discuss the motivation behind our modeling choices and their direct consequences on a more abstract level. We specifically want to emphasize the role of three model ingredients: (i) the model includes a deadline but excludes exponential discounting; (ii) the value of successful thinking diminishes and does so at a sufficiently increasing rate; and (iii) the value of successful thinking is independent of the belief about the doing arm.

Time Cost. Our choice not to include standard exponential discounting is motivated by our focus on the changing time pressure. The implicit assumption in (infinite-horizon) exponential discounting models is that time pressure is constant at any point in time—e.g., because the risk of an exogenous termination of the game is constant. However, with a deadline in mind, this form of time pressure becomes less relevant—at least close to the deadline. Instead, the agent fears that she has insufficient time to finish her task before the game ends with certainty.

---

[27]See, for example, https://www.forbes.com/sites/mariannehudson/2016/01/27/scorecard-helps-angels-value-early-stage-companies/?sh=4e8eb9c96874. When the scorecard method is applied, a startup is valued according to its performance in several dimensions. These include its product/technology and its revenues/sales but also an evaluation of the entrepreneur and the team.

[28]To keep the notation simple, here, we reset time to 0 once the market opportunity is used.

In our applications, deadlines are foreseeable dates on the time horizon, and time pressure increases as agents move closer to the deadline. Therefore, to ensure a transparent and tractable discussion of the effect of changing time pressure, we abstract from additional exponential discounting.

However, it should become clear from the analysis that including exponential discounting would not alter the economic effects of our model but would come at a substantive loss of tractability.[29] We focus on a world in which the agent has no incentive to shirk. As we see, e.g., in Figure 2, the agent has an incentive to obtain results early, even absent exponential discounting: the agent incurs the cost of experimentation.

Diminishing Value of Progress. We assume that the returns to thinking diminish at an increasing rate as the deadline approaches. This assumption captures the idea that successful thinking implies progress but not a solution. When progress arrives, the closer the deadline is, the less time remains to convert the progress made—the return shrinks.

In light of our application, the simplest interpretation of this assumption is to think of progress as triggering a random process determining ex post payoffs. Thus, the greater the time remaining is, the more likely the agent can succeed in time despite a sequence of adverse shocks. That notion of progress differs from one in which a solution arrives deterministically with some delay. In the latter world, the value of progress is a positive constant until it drops to zero once the remaining time falls below a threshold.

In our model, *any* decrease in the time remaining upon the arrival of progress reduces its value: there is less time to convert progress into a solution. Hence, the agent faces a crucial time tradeoff: delaying thinking reduces the expected time remaining when progress arrives and thus makes it less valuable.

Relative Concavity of the Value of Progress. The primary assumption that leads to Proposition 3 is that the relative concavity of the value of progress is sufficiently high. This assumption implies that the evolution of the deadline effect dominates the other effects on the agent's preference as the time remaining increases. We believe that this assumption is helpful for focusing on the main tradeoff between risk and time pressure. However, we could weaken the assumption without losing tractability. The main difference is that we may obtain two disjoint intervals in which the agent pulls the thinking arm and two disjoint intervals in which the agent pulls the doing arm. The underlying economic reason is that once the time pressure has become relatively weak, the thinking arm may have a payoff advantage over the doing arm.

Example 6. Consider a variant of Example 1 in which $\nu < \bar{p}\lambda$; that is, the arrival of progress triggers a new bandit arm that has a relatively low arrival rate. It follows

---

[29]Indeed, it is straightforward yet cumbersome to adjust our key lemmata to include exponential discounting and to verify that Proposition 3 continues to hold. The same holds when deadlines arrive stochastically but become more and more likely as time progresses.

that the change in the deadline effect, $V''(\tau)$, will be dominated by the change in the payoff-on-arrival effect, $p_\tau \lambda V'(\tau)$, when p_τ—and thus the weight on the payoff on the arrival effect—is relatively high.

In particular, to further simplify this example for illustrative purposes, assume that $B_\nu = B + \frac{c}{\mu}$ and $c_\nu = 0$. Once the belief has sufficiently deteriorated such that $p_\tau \lambda = \nu$, the continuation game is nested by our model. Proposition 3 applies. We show numerically that the optimal policy may have the following structure: the agent (i) starts thinking, (ii) switches to doing, (iii) switches back to thinking, and (iv) and returns to the doing arm for a Hail Mary. For an intuition of the resulting changes and additional details, see Appendix D.2.

Value of Progress Independent of the Belief about the Doing Arm. We assume that the belief about the doing arm has no direct consequences on the payoff of the thinking arm. With time τ remaining, the agent attaches the same value to the thinking arm if she is almost certain that the state of the world is $\theta = 0$ or if she is almost certain that it is $\theta = 1$. While we show a set of examples in Section 4.3 that satisfy this assumption, it is nevertheless restrictive. We impose it, as it tremendously simplifies the analysis.

However, this assumption is not crucial for our results. The following example—a version of Example 2—provides a setting in which the value of progress depends on the belief yet our main result remains unchanged.

Example 7. As in Example 2, progress implies a new risky arm with intensity ν, but suppose now that switching back to the initial doing arm is costless. Suppose further that $\nu = \lambda = 1$, $\bar{p} \leq 2/3$, and $B > 2c$. In this case, the entrepreneur will split her effort equally between the two arms once $p_\tau^\nu = p_\tau$.[30]

In Appendix D.3 we provide the formal analysis for a model in which the value of progress depends on the agent's belief about the doing arm that nests Example 7. We derive and discuss a condition that ensures that Proposition 3 remains valid.

The only change in the generalization in Appendix D.3 is that we formulate the value of progress as a function of both the time remaining and the belief that the agent holds at the time of progress. The condition we derive is directly on this value of progress. Therefore, the model outlined in Appendix D.3 captures also other extensions to the baseline case. One such extension is a world in which the value of progress correlates directly with the underlying state θ. This is relevant, for example, if the absence of success is informative about the *problem's* difficulty rather than about the quality of the doing arm.

Feedback about Progress. We assume that if thinking delivers progress in our model, the agent directly observes it. It turns out that this assumption is crucial. To illustrate the

---

[30]Note that upon arrival of a new arm with $\bar{p}^\nu > p_\tau$, the entrepreneur will first only pull the new arm. Thus, its belief will decline while the other arm's belief remains constant until the beliefs on both arms are equal.

role of the assumption, we analyze a version of the model in which the agent understands that thinking delivers progress at a given rate. However, she is not aware whether progress has already arrived. We imagine a setting in which thinking delivers ideas, but some turn out to be dead ends. Continuing to think, the agent either finds out—at a Poisson rate—that the idea was not a dead end, which—for simplicity—momentarily converts it into a solution, or the agent learns that the previous idea was a dead end and obtains a new one that may be a dead end again.

Formally, we model the setting as follows.

Example 8. The thinking arm is a Poisson bandit with the following properties. It delivers progress with known intensity μ if no progress has arrived earlier. It delivers a solution with known intensity ν if progress has arrived earlier. Moreover, we assume that a solution delivers payoff B and that pulling this thinking arm costs c throughout.

The agent observes only the arrival of a solution but *not* that of progress. Example 8 is essentially equivalent to Example 1 with the additional feature that the agent does not observe progress.

First, observe that the logic of the Hail Mary period in our benchmark model fails in this example. Whenever the deadline is close, and the agent has worked on the thinking arm in the past, she attaches some positive probability to unobservable progress having arrived. Thus, the instantaneous probability of solving the problem via progress on the thinking arm does not vanish as it does in the observable progress case. In particular, the agent will pull the doing arm at the deadline whenever the instantaneous probability of a solution via the thinking arm is higher than that on the doing arm.

Therefore, we need to construct the instantaneous solution probability of the thinking arm that takes the unobservability of progress into account. Denote this instantaneous solution probability by $\hat{f}(\tau, A_\tau^{th})$ with A_τ^{th} being the time the agent spent thinking in the past.[31] It turns out that the longer the time the agent has thought in the past is, the higher the probability that unobservable progress has arrived.

This observation is crucial to understand the implications of the absence of direct feedback. If the agent finds it optimal to pull the thinking arm at any point in time, then she becomes more optimistic about that arm in the future. In response, she continues pulling it until either a solution is found or the deadline arrives.

Proposition 8. *Consider Example 8. The agent will never return to the doing arm if she ever optimally leaves it to pull the thinking arm for some positive measure of time.*

The intuition behind Proposition 8 derives directly from the evolution of the arms' expected value. The thinking arm becomes increasingly attractive the more it has been used. The doing arm, instead, becomes increasingly unattractive the more it has been used. More time spent on the thinking arm without a solution corresponds to a higher belief that

---

[31]We can explicitly solve for this probability to obtain $\hat{f}(\tau, A_\tau^{th}) = \left(e^{-\nu A^{th}(\tau)} - e^{-\mu A^{th}(\tau)} \right) \frac{\mu\nu}{\mu-\nu}$.

unobserved intermediate progress did occur. More time spent on the doing arm without a solution corresponds to a lower belief that the doing arm will ever yield a solution. Once the agent considers the doing arm sufficiently unattractive, her relative preference will never revert to favoring it.

Comparing Proposition 8 with the structure from Proposition 4 indicates that whether the agent recognizes progress on the thinking arm qualitatively matters for the optimal strategy. In particular, it is not the progress itself but the observation of the lack thereof that incentivizes the agent to abandon the thinking arm and to return to the doing arm under time pressure. A Bayesian agent who does not observe progress will infer whether her past investment in thinking generated a good idea. Over time she becomes more confident that it indeed has. Absent feedback, she holds on to thinking once she starts.

6 Final Remarks

We address a time-constrained agent's dynamic decision when to *do*—address a problem using an initial idea—and when to *think* about an alternative, less risky method. We show that she should think neither too early nor too late. Overall, the agent never thinks twice. Once she stops thinking and moves to her initial idea, she abandons thinking for good.

We conclude by putting our findings in a broader context by arguing that several insights carry over to settings beyond startup innovation.

One insight is that the agent desires to split her doing time into an initial doing period and a Hail Mary period. She starts working on her initial idea hoping for a quick solution before moving to explore alternative roads. She only returns to the initial idea once the time pressure becomes high. Such behavior resembles the pattern observed in the investment in developing nuclear fusion for power generation. In the 1970s, the US aimed to develop fusion power. The goal was to create a safe, green, and reliable energy source. The Fusion Development Plan of 1976 (see Dean, 1998, for a description) considered several scenarios of funding to develop fusion power plants at different time scales. However, progress was slow initially, and uncertainty about practical feasibility remained high. Consequently, the attention shifted to exploring other options, such as developing nuclear fission, improving the efficiency of coal and gas, or setting up wind and solar energy parks. Fusion power was put on hold because of funding concerns: Funding levels dropped *below* the lowest level discussed in the Fusion Development Plan—a level that the plan considered insufficient to ever build a fusion power system. As the pressure increased to provide large amounts of clean and reliable power, interest in fusion power revived. Government funding has increased since 2018, and private investment has increased in the last few years.[32]

Alternatively, we can use the specification of our model from Section 4 to predict the time an agent needs to find a solution as a function of her initial deadline. Depending on

---

[32]See, e.g., Lierop (2019) and Kramer (2020) for more on the renewed interest in fusion power. For data on inflation-adjusted government funding, see, e.g., `http://qedfusion.org/FPA/OFESbudget.html`.

the expected speed of the various arms, perhaps counterintuitively, we may see that the average time an agent needs to find a solution decreases in her initial deadline. Such a prediction is—in some settings—directly testable. For example, suppose that the thinking and conversion process is sufficiently fast in a field of research. In this case, we would expect that researchers who undergo evaluation after an intermediate tenure clock would take on average *longer* to fulfill the tenure requirements than those with longer or shorter tenure clocks. The mechanism is the following: those with shorter tenure clocks fail more often: the late bloomers drop out and the average time conditional on making tenure is shorter. Those with longer tenure clocks work first on converting their job market papers to influential publications and begin to branch only upon failure. Since some succeed, they need considerably less time to fulfill the requirements—they outpace their counterparts with intermediate clocks who branch from the beginning.[33]

Finally, our result from Proposition 5 can serve as a cautionary tale regarding the efforts by politicians to meet specific precommitted goals. For example, consider the Paris agreement. Many countries aim to target their goals through small-scale efforts such as incentivizing the use of electric cars. However, potentially, there is an action bias. Governments prefer implementing measures that are directly at hand—even at the risk of these being insufficient—instead of investing in "transformational change" that is warranted, e.g., by IPCC (2018). Only if it becomes imminent that small-scale effort will not be sufficient will governments pivot to thinking about transformation. The late pivot is optimal from the government's perspective. Nevertheless, it occurs too late on a larger scale: the chances of meeting the goals in time decline.

---

[33]We want to emphasize that this result does not hold under arbitrary arrival rates. Therefore, to test these predictions, one must understand the arrival rates in place.

Appendix

A   The Optimal Control Problem

A.1   Notation

Variable	Description
$\bar{p}$	Ex ante probability that $\theta = 1$.
B	Benefit of arrival on the doing arm.
$V(\tau)$	Value of progress with time τ remaining.
c	Flow cost of effort.
$\tau := T - t$	Time until the deadline.
a_τ	Relative intensity of pulling the doing arm with τ periods remaining.
$A_\tau := \int_\tau^T a_s ds$	Total amount of time spent pulling the doing arm in the past. *The state.*
$p_\tau := \frac{\bar{p}e^{-\lambda A_\tau}}{\bar{p}e^{-\lambda A_\tau} + 1 - \bar{p}}$	Likelihood that the doing arm is suitable with τ periods remaining.
$U(\tau) := (B - c/\lambda)(1 - e^{-\lambda\tau})$	Value of pulling an arm with known intensity λ for time τ.
$\gamma_\tau := -\frac{dH_\tau(a_\tau; A_\tau)}{da_\tau}$	Relative preference for thinking with τ periods remaining.
η_τ	Co-state with τ periods remaining.
$Z^d(p, \tau) := pU(\tau) - (1 - p)c\tau$	Expected value from pulling the doing arm throughout with time τ remaining when the belief is p.
$\begin{aligned} Z^t(\varepsilon; p, \tau) := &(\mu V(\tau) - c)\varepsilon \\ &+ (1 - \mu\varepsilon)Z^d(p, \tau - \varepsilon) \\ &+ o(\varepsilon) \end{aligned}$	Expected value from pulling the thinking arm for a small measure of time ε and pulling the doing arm for the remaining time with time τ remaining when the belief is p.

A.2   Necessary Conditions for Optimality

Much of our arguments rely on the necessary conditions from Pontryagin's maximum principle. The existence of an optimal control follows from standard arguments.[34] We verify the sufficiency of the necessary conditions in the proof of Proposition 4 by showing that there is a unique solution to the necessary conditions.

---

[34]See, for example, Clarke (2013). The evolution of the state is continuous and bounded, the control is bounded, the agent's value is finite, the running cost is convex in the control, and the set of admissible effort paths is nonempty.

Using the notation from Appendix A.1 the agent's objective is to maximize

$$\int_0^T e^{-\mu(T-\tau-A_\tau)}(1-\bar{p}+\bar{p}e^{-\lambda A_\tau})J(p_\tau,\tau,a_\tau)d\tau$$

with $J(a_\tau,A_\tau,\tau) := \mu(1-a_\tau)V(\tau) + \lambda a_\tau p_\tau B - c$ the expected flow payoff at time $\tau := T - t$.

Using A_τ as the state, the Hamiltonian at time τ is thus

$$\begin{aligned}
H_\tau(a_\tau; A_\tau) :=&e^{-\mu(T-\tau-A_\tau)}(1-\bar{p}+\bar{p}e^{-\lambda A_\tau})J(p_\tau,\tau,a_\tau) + a_\tau\eta_\tau, \\
=&e^{-\mu(T-\tau-A_\tau)}(1-\bar{p})\left((1-a_\tau)\mu V(\tau) - c\right) \\
&+ e^{-\mu(T-\tau-\mu A_\tau)}\bar{p}e^{-\lambda A_\tau}\left((1-a_\tau)\mu V(\tau) + a_\tau\lambda B - c\right) + a_\tau\eta_\tau
\end{aligned} \tag{3}$$

where η is the co-state. It has terminal condition $\eta_0 = 0$ and evolves according to[35]

$$\begin{aligned}
\frac{d\eta_\tau}{d\tau} :=& \frac{dH_\tau(a_\tau; A_\tau)}{dA_\tau} \\
=&e^{-\mu(T-\tau-A_\tau)}\left(\mu(1-\bar{p})\left((1-a_\tau)\mu V(\tau) - c\right)\right. \\
&\left. - (\lambda-\mu)e^{-\lambda A_\tau}\bar{p}\left((1-a_\tau)\mu V(\tau) + a_\tau\lambda B - c\right)\right).
\end{aligned} \tag{4}$$

Conditional on no arrival, the relative preference of thinking is determined by

$$\begin{aligned}
\gamma_\tau =& -\frac{dH_\tau(a_\tau; A_\tau)}{da_\tau} \\
=& e^{-\mu(T-\tau-A_\tau)}\left(\left(1-\bar{p}+\bar{p}e^{-\lambda A_\tau}\right)\mu V(\tau) - \bar{p}e^{-\lambda A_\tau}\lambda B\right) - \eta_\tau.
\end{aligned}$$

If $\gamma_\tau < 0$, then the agent strictly prefers to do, $a_\tau = 1$; if $\gamma_\tau > 0$, then she strictly prefers to think, $a_\tau = 0$.

B   Key Lemmata

We state and prove four key lemmata. Combining these delivers most of our results up to and including parts of the characterization in Proposition 4. The first, Lemma 1, states that the agent prefers to pull the doing arm close to the deadline independent of her belief about its quality. The second, Lemma 2, states the evolution of the agent's relative preference between the arms over time. Because the Hamiltonian is linear in the agent's action, the evolution of the relative preference is independent of the agent's action. The third, Lemma 3, states that the agent's relative preference has no interior minimum. This implies that the agent never *returns* to the thinking arm if she stopped thinking without progress. The fourth, Lemma 4, determines a condition such that the strategy "doing throughout" is dominated by the strategy "think for a measure of time dt before doing for the remainder of the time."

Lemma 1. *Suppose that the agent has not observed an arrival on either arm and holds belief $p \in (0,1)$ on the doing arm at the deadline. There is a remaining time $\hat{\tau}$ such that for*

---

[35]Note that because we take the derivative with respect to time remaining, the sign on the partial of the Hamiltonian is positive instead of negative.

the entire interval $\tau \in [0, \hat{\tau})$, the agent strictly prefers to pull the doing arm over pulling the thinking arm.

Proof. γ_τ is continuous in τ, and $\eta_0 = 0$. The value of a success on the thinking arm is continuous, and when $\tau = 0$ it is $V(0) = 0$. Thus, the terminal value of γ is

$$\gamma_0 = -e^{-\mu(T-A_\tau)}pe^{-\lambda A_\tau}\lambda B < 0.$$

By the continuity of γ_τ, there exists for any strategy $(a_\tau)_{\tau=0}^T$ a remaining time $\hat{\tau} > 0$ such that $\gamma_\tau < 0$ for $\tau < \hat{\tau}$, which proves the claim.[36] $\qquad\square$

Lemma 2. *A marginal increase in the time to the deadline τ changes the agent's policy function γ_τ by*

$$\frac{d\gamma_\tau}{d\tau} = e^{-\mu(T-\tau-A_\tau)}(1 - \bar{p} + \bar{p}e^{-\lambda A_\tau})\left(\mu V'(\tau) + p_\tau\mu\lambda(V(\tau) - B) + (\mu - \lambda p_\tau)c\right). \qquad (5)$$

The change is independent of the agent's decision a_τ.

Proof. First, recall that $H_\tau(a_\tau; A_\tau)$ is affine in a_τ, and note that for a function affine in x, $f(x;\theta) = t(\theta) + m(\theta)x$, it holds that

$$\frac{d^2 f(x;\theta)}{dxd\theta}x = m'(\theta)x = \frac{df}{d\theta} - \frac{df}{d\theta}\Big|_{x=0}.$$

Second, recall that $\frac{d\eta_\tau}{d\tau} = \frac{dH_\tau(a_\tau;A_\tau)}{dA_\tau}$ and that $-\frac{dA(\tau)}{d\tau} = a_\tau$, which yields (using $\gamma_\tau = -\frac{dH}{da_\tau}$)

$$\frac{d\gamma}{d\tau} = -\frac{d^2 H_\tau(a_\tau; A_\tau)}{da_\tau d\tau} = -\frac{d^2 J(a_\tau; A_\tau, \tau)}{da_\tau d\tau} - \frac{d\eta}{d\tau} = -\frac{\partial}{\partial\tau}\frac{dJ(a_\tau; A_\tau, \tau)}{da_\tau} - \frac{d\eta}{d\tau}\Big|_{a_\tau=0} \qquad (6)$$

where we used that the law of motion of the state is independent of the state itself, $\frac{d^2 J}{da_\tau d\tau} = \frac{\partial}{\partial\tau}\frac{dJ}{da_\tau} - \frac{d^2 H}{dA_\tau da_\tau}a_\tau$ and $\frac{d\eta}{d\tau} - \frac{d^2 H}{dA_\tau da_\tau}a_\tau = \frac{d\eta}{d\tau}\Big|_{a_\tau=0}$ based on the Hamiltonian being affine in a_τ.

Third, from (4) we obtain

$$\frac{d\eta}{d\tau}\Big|_{a_\tau=0} = e^{-\mu(T-\tau-A_\tau)}(1 - \bar{p} + \bar{p}e^{-\lambda A_\tau})\left(\mu V(\tau) - c\right)(\mu - \lambda p_\tau). \qquad (7)$$

Moreover,

$$-\frac{\partial}{\partial\tau}\frac{dJ(a_\tau; A_\tau, \tau)}{da_\tau} = e^{-\mu(T-\tau-A_\tau)}(1 - \bar{p} + \bar{p}e^{-\lambda A_\tau})\left(\mu V'(\tau) + \mu\left(\mu V(\tau) - p_\tau\lambda B\right)\right)$$

which implies

$$\begin{aligned}
\frac{d\gamma}{d\tau} = \quad & e^{-\mu(T-\tau-A_\tau)}(1 - \bar{p} + \bar{p}e^{-\lambda A_\tau})\left(\mu V'(\tau) + \mu\left(\mu V(\tau) - p_\tau\lambda B\right)\right) \\
& - e^{-\mu(T-\tau-A_\tau)}(1 - \bar{p} + \bar{p}e^{-\lambda A_\tau})\left(\mu V(\tau) - c\right)(\mu - \lambda p_\tau) \\
= \quad & e^{-\mu(T-\tau-A_\tau)}(1 - \bar{p} + \bar{p}e^{-\lambda A_\tau})\underbrace{\left(\mu V'(\tau) + \mu p_\tau\lambda\left(V(\tau) - B\right) + (\mu - \lambda p_\tau)c\right)}_{=:dy_\tau/d\tau}.
\end{aligned} \qquad (8)$$

---

[36]For any initial belief $\bar{p} \in (0, 1)$ and any strategy $(a_\tau)_{\tau=0}^T$, the agent's terminal belief is in $(0, 1)$.

Note that the sign of $d\gamma_\tau/d\tau$ is determined by the sign of $dy_\tau/d\tau$ only. □

The last line of (8) implies that the agent's action has no first-order effect on the evolution of the switching function y_τ. Any direct effect of the action on the instantaneous payoffs is counteracted by an effect on the continuation value. It is important, however, to keep in mind that the action has a second-order effect on y_τ through its effect on the evolution of the belief.

Lemma 3. *Any minimum of γ_τ is either at $\tau = T$ or at $\tau = 0$. Moreover, y_τ is strictly concave whenever $a_\tau = 0$.*

Proof. Since η_τ is continuously differentiable in τ and defined for any $\tau \in \mathbb{R}_+$, so is y_τ. To prove Lemma 3, we use that any interior (local) minimum has to be a critical point. From equation (8) we know that

$$\frac{d\gamma}{d\tau} = \underbrace{e^{-\mu(T-\tau-A_\tau)}(1-\bar{p}+\bar{p}e^{-\lambda A_\tau})}_{f(\tau)}\underbrace{(\mu V'(\tau) + p_\tau\mu\lambda(V(\tau)-B) + (\mu-\lambda p_\tau)c)}_{\frac{dy_\tau}{d\tau}=:g(\tau)}.$$

If a critical point constitutes a local minimum, then it satisfies

$$f(\tau)g(\tau) = 0 \qquad \text{and} \qquad f'(\tau)g(\tau) + f(\tau)g'(\tau) > 0$$

because $f(\tau) > 0$ for all $\tau < \infty$, which implies $g(\tau) = 0$, and any local minimum also requires $g'(\tau) = \frac{d^2 y_\tau}{d\tau d\tau} > 0$. To show that such a local minimum cannot exist, we show that $g(\tau) = 0$ and $g'(\tau) > 0$ cannot be satisfied at the same time. It follows from $g(\tau) = 0$ that

$$\mu\lambda(V(\tau)-B) = -\frac{\mu V'(\tau) + (\mu-\lambda p_\tau)c}{p_\tau}.$$

Differentiating $g(\tau)$ yields

$$\frac{d^2 y_\tau}{d\tau d\tau} = \frac{dp_\tau}{d\tau}\left(\mu\lambda(V(\tau)-B) - \lambda c\right) + \mu\left(V''(\tau) + \lambda p_\tau V'(\tau)\right)$$

and plugging in for $g(\tau) = 0$ yields

$$\frac{d^2 y_\tau}{d\tau d\tau}\Big|_{\frac{dy_\tau}{d\tau}=0} = -\frac{dp_\tau}{d\tau}\frac{\mu}{p_\tau}\left(V'(\tau) + c\right) + \mu\left(V''(\tau) + \lambda p_\tau V'(\tau)\right) \leq 0$$

where the inequality follows as $\frac{dp_\tau}{d\tau} \geq 0$, $V'(\tau) > 0$ and $-\frac{V''(\tau)}{V'(\tau)} \geq p_\tau\lambda$ as $p_\tau \leq p^0$.

The concavity of y_τ while the agent thinks follows straightforwardly by observing that $\frac{dp_\tau}{d\tau} = 0$ in this case.

□

To state Lemma 4, we define:

$$Z^d(p,\tau) := p\int_0^\tau e^{-\lambda t}(\lambda B - c)dt - (1-p)\int_0^\tau cdt = pU(\tau) - (1-p)c\tau,$$

and

$$Z^t(\varepsilon; p,\tau) := \int_0^\varepsilon e^{-\mu t}(\mu V(\tau-t) - c)dt + e^{-\mu\varepsilon}Z^d(p,\tau-\varepsilon)$$
$$= (\mu V(\tau) - c)\varepsilon + (1-\mu\varepsilon)Z^d(p,\tau-\varepsilon) + o(\varepsilon)$$

where the expression follows from a Taylor expansion around 0.

The first, Z^d, describes the value absent a success of the strategy "pull the doing arm from now until $\tau = 0$", given the belief p and the time remaining τ. The second, Z^t, describes the value absent a success of the strategy "pull the thinking arm for a small measure of time $\varepsilon > 0$, then pull the doing arm until $\tau = 0$", given the belief p and the time remaining τ.

Lemma 4.
$$\lim_{\varepsilon \to 0} Z^t(\varepsilon; p, \tau) - Z^d(p, \tau) > 0$$

if and only if
$$\hat{q}(\tau) := \frac{\mu \left(V(\tau) + c\tau \right)}{\mu \left(B + c\tau \right) + \left(\lambda - \mu \right) \left(B - U(\tau) \right)} > p.$$

Moreover for any $q \in (0, 1)$ there is a τ such that $\hat{q}(\tau) = q$.

Proof.

$$\lim_{\varepsilon \to 0} Z^t(\varepsilon; p, \tau) - Z^d(p, \tau) > 0$$

$$\Leftrightarrow \quad \lim_{\varepsilon \to 0} \left(\mu V(\tau) - c \right) \varepsilon + (1 - \mu\varepsilon) Z^d \left((p, \tau - \varepsilon) - Z^d(p, \tau) \right) + \mu\varepsilon Z^d(p, \tau) > 0$$

$$\Leftrightarrow \quad \mu V(\tau) - c - \mu Z^d(p, \tau) - \lim_{\varepsilon \to 0} \frac{(Z^d(p, \tau) - Z^d(p, \tau - \varepsilon))}{\varepsilon} > 0$$

$$\Leftrightarrow \quad \mu V(\tau) - c - \mu Z^d(p, \tau) - \frac{\partial Z^d(p, \tau)}{\partial \tau} > 0$$

$$\Leftrightarrow \quad \mu V(\tau) - p U'(\tau) - pc - \mu p U(\tau) + (1 - p)\mu c\tau > 0$$

which is equivalent to

$$\hat{q}(\tau) := \frac{\mu \left(V(\tau) + c\tau \right)}{\mu(U(\tau) + c\tau) + U'(\tau) + c} = \frac{\mu \left(V(\tau) + c\tau \right)}{\mu(B + c\tau) + (\lambda - \mu)(B - U(\tau))} > p. \tag{9}$$

The last claim follows because $\hat{q}(0) = 0$, $\lim_{\tau \to \infty} \hat{q}(\tau) = 1$ and $\hat{q}$ is continuous.

The limit $\tau \to \infty$ follows using L'Hôpital's rule,

$$\lim_{\tau \to \infty} \hat{q}(\tau) = \lim_{\tau \to \infty} \frac{\mu(V'(\tau) + c)}{\mu(U'(\tau) + c) + U''(\tau)} = 1$$

where the last equality follows from $\lim_{\tau \to \infty} U''(\tau) = \lim_{\tau \to \infty} U'(\tau) = 0$ and $\lim_{\tau \to \infty} V'(\tau) = 0$ because on an unbounded support any strictly concave, increasing, yet bounded function has to have a slope converging to zero if the limit of its derivative exists, which holds by assumption. $\square$

C Proofs of Statements in the Main Text

C.1 Proof of Proposition 1

Proof. When $T = \infty$, the value of an arrival on the thinking arm, $V(\infty)$, is constant over time. We can apply the standard dynamic programming approach for exponential bandits.

The value function $u(p)$ given belief p prior to termination satisfies[37]

$$u(p) = \max_{a \in [0,1]} (ap\lambda B + (1-a)\mu V(\infty)) - c)dt + (1 - (ap\lambda + (1-a)\mu)dt)(u(p) - p(1-p)\lambda a dt u'(p)).$$

Letting dt go to zero, dropping second-order terms and rearranging, we obtain the Bellman equation

$$0 = \max_{a \in [0,1]} (ap\lambda B + (1-a)\mu V(\infty)) - c) - (ap\lambda + (1-a)\mu)u(p) - p(1-p)\lambda a u'(p)$$

where the maximand is linear in a. Whenever pulling the thinking arm is optimal at some time t, the agent chooses $a_t = 0$, and it will remain optimal to pull $a_{t'} = 0$ for all $t' > t$ as the belief remains constant. Thus, $u(p) = \int_0^\infty e^{-\mu t}(\mu V(\infty) - c)dt = V(\infty) - \frac{c}{\mu}$ whenever $a = 0$. Whenever, pulling the doing arm is optimal, we can rewrite the Bellman equation as

$$0 = p\lambda B - c - p\lambda u(p) - p(1-p)\lambda u'(p)$$

and solving this differential equation yields

$$u(p) = B - \frac{c}{\lambda} - (1-p)\left(\mathbb{C} + c\ln\left(\frac{p}{1-p}\right)\right)$$

where $\mathbb{C}$ is a constant of integration. Using the value matching condition that the agent is indifferent between thinking and doing at $p = \hat{p}$, $u(\hat{p}) = V(\infty) - \frac{c}{\mu}$, and the smooth pasting condition, $u'(\hat{p}) = 0$, we can obtain the constant of integration as well as $\hat{p}$, which are

$$\hat{p} = \frac{\mu}{\lambda}\frac{c}{c + \mu(B - V(\infty))}$$

$$\mathbb{C} = B - \frac{c}{\lambda}\ln\left(\frac{\hat{p}}{1-\hat{p}}\right) - \left(V(\infty) - \frac{c}{\mu}\right).$$

$\square$

C.2 Proof of Proposition 2

Proof. We make use of $dy_\tau/d\tau$ as defined in the proof of Lemma 2. By Lemma 3, the policy function γ_τ is twice continuously differentiable and has no interior minimum in τ. This implies that if $dy_\tau/d\tau \geq 0$ for some τ, then $dy_\tau/d\tau \geq 0$ for all $\tau' \in [0, \tau]$. Because condition (C.1) holds, we obtain that for all $\tau \geq 0$

$$\frac{dy_\tau}{d\tau} = \mu\left(V'(\tau) - \bar{p}\lambda(B - V(\tau))\right) \geq 0$$

as $V(\tau) \leq B$ and $V'(\tau) \geq 0$. Thus, y_τ is increasing in the time remaining τ throughout. The agent pulls the doing arm close to the deadline by Lemma 1. What remains to be shown is if and when the agent switches from thinking to doing. Invoking Lemma 4 assuming $c = 0$ yields that the agent never thinks (i.e., $\gamma_\tau < 0$) if an only if $\forall \tau \leq T$

$$\hat{q}(\tau|c = 0) := \frac{\mu V(\tau)}{U'(\tau) + \mu U(\tau)} < \bar{p}.$$

---

[37]We obtain the expression using Taylor approximations for the success probabilities when dt is small and using the ODE $dp/dt = p(1-p)a_\tau\lambda$. We do not require discounting as a single breakthrough on any arm ends the problem generating a finite payoff. Hence, the value is bounded even without discounting.

Otherwise, she starts thinking and switches to doing with time τ_3 remaining, where τ_3 is the smallest solution to $\hat{q}(\tau_3|c=0) = \bar{p}$. $\qquad\square$

C.3 Proof of Proposition 3

Proof. The result follows from Lemmata 1 to 3.

By Lemma 1, the agent pulls the doing arm shortly before the deadline whenever the game has not yet terminated. Lemma 2 shows that y_τ is strictly concave at any critical point. Thus, $y_\tau \neq 0$ almost everywhere; i.e., the agent is generically not indifferent. Finally, by Lemma 3, y_τ has no interior minimum: once the agent abandons the thinking arm, she does not return to it. Only the three strategies in Proposition 3 remain possible. $\qquad\square$

C.4 Proof of Proposition 4

Proof. First, we show that any solution provided by the algorithm satisfies the necessary condition of the agent's optimal control problem. Second, we show that the algorithm provides a solution. Third, we show that there is a unique solution to the necessary conditions of the optimal control problem under Assumption 1.

Step 1. The algorithm's solution is a candidate. Here, we show that any solution to the algorithm satisfies the necessary conditions of the optimal control problem. We consider the different termination cases of the algorithm.

1a. The algorithm stops in item 2. In this case, the algorithm's solution implies that the agent pulls the doing arm throughout. By Lemma 4, it is optimal for the agent to follow this strategy as the agent will never be indifferent between thinking and doing.

1b. The algorithm stops in item 4. Lemma 4 implies that it is optimal for the agent to pull the doing arm for the final remaining time $\tau < \overline{\tau}_3$ when she holds belief $\bar{p}$ at τ_3. The function $\dot{y}(\tau; q(\overline{\tau}_3), \overline{\tau}_3)$ has the same sign as the slope of the agent's policy function from Lemma 2 when pulling the thinking arm conditional on the agent pulling the doing arm for any remaining time $\tau < \overline{\tau}_3$. The next Lemma shows that $\dot{y}(\tau; q(\overline{\tau}_3), \overline{\tau}_3) \geq 0$.

Lemma 5. *Suppose that it is optimal for the agent to switch from pulling the thinking arm for a positive measure of time to pulling the doing arm with time τ_3 remaining. Then*

$$\dot{y}(\tau; q(\tau_3), \tau_3) > 0.$$

Proof. To the contrary, assume that $\dot{y}(\tau; q(\tau_3), \tau_3) \leq 0$. Because switching to the doing arm is optimal with time τ_3 remaining, by Proposition 3 and γ_τ being a continuously differentiable function, we must have that $\gamma_{\tau_3} = 0$. With $\dot{y}(\tau; q(\tau_3), \tau_3) < 0$, this implies that there is an $\overline{\varepsilon}$ such that $\gamma_{\tau_3 - \varepsilon} > 0$ for all $\varepsilon \in (0, \overline{\varepsilon})$, which implies (strict) optimality of thinking with time $\tau_3 - \varepsilon$ remaining, a contradiction to switching to the doing arm with time τ_3 remaining. With $\dot{y}(\tau; q(\tau_3), \tau_3) = 0$, the agent would pull the doing arm immediately again for $\tau > \tau_3$, as any critical point corresponds to a strict local maximum, which is a contradiction to thinking for a positive measure of time before τ_3. $\qquad\square$

By Lemma 5, there is a $\Delta > 0$ such that for all $\tau \in (\overline{\tau}_3, \overline{\tau}_3 + \Delta]$, $\hat{y}(\tau; \bar{p}, \overline{\tau}_3) > 0$, as $\hat{y}$ is continuous in τ. Thus, $\tau_2(\overline{\tau}_3)$ is defined, and for any time horizon $T < \tau_2(\overline{\tau}_3) + \overline{\tau}_3$, it is optimal for the agent to start thinking before switching to the doing arm for the remaining time.

1c. The algorithm stops in item 6. The continuation game with time $\tau_2(\overline{\tau_3}) + \overline{\tau_3}$ is identical to one in which the agent starts out with belief $\bar{p} = q(\overline{\tau_3})$ and deadline $T = \tau_2(\overline{\tau_3}) + \overline{\tau_3}$. Moreover, $\tau_1(\tau_3)$ describes the length of the initial doing period to arrive at this continuation game. By construction, at $\tau_2(\tau_3) + \tau_3$, the agent's policy function γ must be increasing, as it coincides with $\hat{y}$ multiplied by a positive constant. By Lemma 3, γ_τ cannot have a minimum and, therefore, must be negative for all $\tau \in (T - \tau_2(\tau_3) - \tau_3, T]$.

Step 2. The algorithm finds a solution. Here, we show that the algorithm always provides a solution. First, we state a lemma that will be useful for the remainder.

Lemma 6. *Under Assumption 1, the following monotonicity statements hold.*
  (i) $q(\tau_3)$ is monotonically increasing in τ_3.
  (ii) $\tau_1(\tau_3)$ is monotonically decreasing in τ_3.
  (iii) $\tau_2(\tau_3)$ is monotonically decreasing in τ_3.

Proof. We prove each statement separately. Recall that $q(\tau) = \min\{1, \hat{q}(\tau)\}$.

Statement (i). To simplify exposition, we use the notation $\hat{q}(\tau) = \frac{x}{z}$, with $x = \mu(V(\tau) + c\tau)$ and $z = \mu(U(\tau) + c\tau) + U'(\tau) + c$. Note that $q(\tau)$ is continuous and that $\hat{q}'(\tau = 0) > 0$. Thus, if $\hat{q}(\tau)$ ever decreases a local maximum must exist. Moreover, $\lim_{\tau \to \infty} \hat{q}(\tau) = 1$. Thus, if $q(\tau)$ ever decreases, $\hat{q}(\tau)$ must have a local minimum with $\hat{q}(\tau) < 1$. We will show that $q(\tau)$ cannot be decreasing by showing that $\hat{q}(\tau)$ has no local minimum with $\hat{q}(\tau) < 1$.

Case 1: $\mu \le \lambda$. Consider $\mu \le \lambda$. In this case, a local maximum implies that $\hat{q}'(\tau) = 0$ and $\hat{q}''(\tau) < 0$. Thus,

$$\frac{V'(\tau) + c}{V(\tau) + c\tau} = \frac{(\mu - \lambda)U'(\tau) + \mu c}{\mu(U(\tau) + c\tau) + U'(\tau) + c} \tag{10}$$

and

$$\frac{V''(\tau)}{V'(\tau) + c} < \frac{(\mu - \lambda)U''}{(\mu - \lambda)U'(\tau) + \mu c}, \tag{11}$$

where we have used $U'' = -\lambda U'$ and that $\hat{q}'(\tau) = 0$ implies $\frac{x'}{x} = \frac{z'}{z}$. A critical point requires $\frac{x'}{x} = \frac{z'}{z}$, and since x', x and z are trivially greater than zero, so is z'. It follows that whenever $\mu < \lambda$, the right-hand side of (11) is positive, implying that any critical point must be a local maximum. Because $\lim_{\tau \to \infty} \hat{q}(\tau) = 1$, any local maximum at $\hat{\tau}$ with $\hat{q}(\hat{\tau}) < 1$ would imply the existence of a local minimum for some $\check{\tau} > \hat{\tau}$. Hence, any local maximum that may exist can only occur at some $\hat{\tau}$ such that $\hat{q}(\tau) > 1$ for all $\tau \ge \hat{\tau}$, implying that $q(\tau) = 1$ for all $\tau \ge \hat{\tau}$ and thus that $q(\tau)$ is monotonic for all $\tau \ge 0$.

Case 2: $\mu > \lambda$. Consider $\mu > \lambda$. Recall that whenever $\hat{q}'(\check{\tau}) < 0$ and $\hat{q}(\check{\tau}) < 1$, there must be some local minimum; denote the time remaining at the local minimum by $\check{\tau}_2$. As $\hat{q}'(0) > 0$, there must be a local maximum of $\hat{q}(\tau)$ first; denote the time remaining at the local maximum by $\check{\tau}_1$, with $\check{\tau}_1 < \check{\tau}_2$. Moreover, as $\lim_{\tau \to \infty} \hat{q}(\tau) = 1$, there must be either another local maximum, the time remaining of which is denoted by $\check{\tau}_3$, or there is some $\tilde{\tau}$ such that for all $\tau > \tilde{\tau}$, $\hat{q}''(\tau) < 0$. Define $\varphi(\tau) := x''z - z''x$.

These three observations imply the following: $\varphi(\check{\tau}_1) < 0$; $\varphi(\check{\tau}_2) > 0$; if $\check{\tau}_3$ exists, then $\varphi(\check{\tau}_3) < 0$; and if $\check{\tau}_3$ does not exist, then there must be some $\tilde{\tilde{\tau}}$ such that $\varphi(\tilde{\tilde{\tau}}) < \varphi(\check{\tau}_2)$. The latter conclusion follows by the observation that $\hat{q}''(\tau) < 0$ whenever $\hat{q}(\tau)$ converges

from below to 1. $\hat{q}''(\tau) < 0$ implies that $\varphi(\tau) < 2\frac{z'}{z}(x'z - xz')$, where the right-hand side converges to 0 as $\tau \to 1$.

Thus, we know that as τ moves from $\tau \le \check{\tau}_1$ to ∞, $\phi(\tau)$ is strictly negative (at $\check{\tau}_1$), strictly positive (at $\check{\tau}_2$) and arbitrarily small as we approach $\tau = \infty$. Thus, $\phi(\tau)$ has to be nonmonotonic. Part (ii) of Assumption 1 rules nonmontonicity out. Thus, no local minimum of $\hat{q}(\tau)$ with $\hat{q}(\tau) < 1$ exists, and $q(\tau)$ is monotonic.

Statement (ii). The monotonicity of τ_1 follows by the monotonicity of $q(\tau_3)$ and the observation that $\tau_1(\tau_3)$ decreases in $q(\tau_3)$.

Statement (iii). To see that $\tau_2(\tau_3)$ decreases, recall that τ_2 is determined via the root of $\hat{y}(\tau; q(\tau_3), \tau_3)$ whenever this root exists for some $\tau > 0$. In this case, we require by definition of $\tau_2(\tau_3)$[38]

$$\frac{\mathrm{d}\hat{y}(\tau_2(\tau_3); q(\tau_3), \tau_3)}{\mathrm{d}\tau_3} = \frac{\partial\hat{y}(\tau_2(\tau_3); q(\tau_3), \tau_3)}{\partial\tau_3} + \frac{\partial\hat{y}(\tau_2(\tau_3); q(\tau_3), \tau_3)}{\partial q(\tau_3)}\frac{\partial q(\tau_3)}{\partial\tau_3}$$
$$+ \frac{\partial\hat{y}(\tau_2(\tau_3); q(\tau_3), \tau_3)}{\partial\tau_2}\frac{\partial\tau_2(\tau_3)}{\partial\tau_3} = 0. \tag{12}$$

Note that under Assumption 1, $\frac{d}{d\tau_3}\dot{y}(s; q(\tau_3), \tau_3) < 0$ because

$$\frac{d\dot{y}(s; q(\tau_3), \tau_3)}{d\tau_3} = \mu V''(s + \tau_3) + \mu\lambda\left(q(\tau_3)V'(s + \tau_3) + \frac{dq(\tau_3)}{d\tau_3}\left(V(s + \tau_3) - B - \frac{c}{\mu}\right)\right)$$

which is negative for all s, as $-V''(\tau_3 + s)/V'(\tau_3 + s) \ge \bar{p}\lambda \ge q(\tau_3)\lambda$ and $V(\tau_3 + s) \le B + \frac{c}{\mu}$ by assumption while $q(\tau_3)$ is increasing in τ_3.

Hence, we know that

$$\frac{\partial\hat{y}(\tau_2(\tau_3); q(\tau_3), \tau_3)}{\partial\tau_3} + \frac{\partial\hat{y}(\tau_2(\tau_3); q(\tau_3), \tau_3)}{\partial q(\tau_3)}\frac{\partial q(\tau_3)}{\partial\tau_3} < 0$$

and, moreover, that $\partial\hat{y}(\tau = \tau_2; p, \tau_3)/\partial\tau < 0$ because τ_2 is the root and $\hat{y}(\tau; p, \tau_3) > 0$ if $\tau < \tau_2$ by construction. Thus, to satisfy (12), we need $\partial\tau_2(\tau_3)/\partial\tau_3 < 0$. $\quad\square$

Second, we show that if there is a solution without an initial doing period, then the algorithm always returns such a solution.

If such a solution exists, then there exists a $\bar{\tau}_3$ such that $\bar{p} = q(\bar{\tau}_3)$, as $\lim_{\tau\to\infty} q(\tau) = 1 > \bar{p}$, which item 2 of the algorithm will detect because q is monotonic by Lemma 6. If the solution is such that only a Hail Mary period is possible, then item 2 ensures that $\tau_3 = T$, as the solution detected $\bar{\tau}_3 > T$. If a solution in which the agent starts by thinking is possible, then item 4 detects one such solution, i.e., if $\bar{\tau}_3 < T$ from item 2 and $q(\bar{\tau}_3) = \bar{p}$. Neither item 2 nor item 4 returns a solution only if any policy that involves only a single doing period does not satisfy the necessary conditions of the optimal control problem.

Third, we show that if all solutions involve an initial doing period, then item 6 of the algorithm finds such a solution.

If an initial doing period exists, then the belief held at the beginning of the Hail Mary

---

[38]Recall that because $\bar{p} < 1$, $q(\tau)$ is differentiable in the relevant part.

period must satisfy $q(\tau_3) < \bar{p}$. At the same time, by Lemma 1, $q(\tau_3) > 0$, and

$$q(\tau_3 - t) \leq \frac{q(\tau_3)e^{-\lambda(\tau_3-t)}}{(q(\tau_3)e^{-\lambda(\tau_3-t)} + 1 - q(\tau_3)}$$

for all $t \in [0, \tau_3]$ such that Lemma 4 does not imply any additional switches for any time remaining $\tau < \tau_3$.

By Lemmata 2 and 3, any solution with an initial doing period implies the existence of two roots of the policy function γ_τ. Because $\hat{y}(\tau; q(\tau_3), \tau_3) = 0$ determines the smallest root $\tau > 0$ of γ_τ conditional on a Hail Mary period of length τ_3, the algorithm detects that root if it exists. Finally, because $q(\tau_3) < \bar{p}$ and beliefs are constant while thinking, the length of the initial doing period is determined by Bayes' rule and the belief conditional on reaching the Hail Mary period $q(\tau_3) = \bar{p}e^{-\lambda t}/(\bar{p}e^{-\lambda t} + 1 - \bar{p})$, which results in $\tau_1(\tau_3)$.

Item 6 of the algorithm considers all possible combinations of $\tau_1(\tau_3), \tau_2(\tau_3)$ and τ_3 until a solution is found that satisfies the necessary conditions. If a solution exists, the algorithm converges.

Fourth and finally, a solution to the optimal control problem exists because the evolution of the state is continuous and bounded, the control is bounded, the agent's value is finite, the running cost is convex in the control and the set of admissible effort paths is nonempty (see, e.g., Clarke (2013) for details). By Proposition 3, any solution is of one of the three types the algorithm considers. Thus, the algorithm determines a candidate solution.

Step 3. The algorithm's solution is the unique candidate. Finally, we show that the algorithm identifies the uniquely optimal policy. To show uniqueness, we have to show that given a solution τ_3, there is no other $\tau_3' \neq \tau_3$ that solves the fixed point problem.

Any two solutions in which the agent pulls the doing arm for only one time interval are identical on any positive measure of time. This is immediate because if the agent pulls the doing arm on only one interval, then she has to pull it in the end. Either the agent pulls only the doing arm in which the strategy is trivially unique and $q(\tau) < p_\tau$ for any $\tau \in [0, T]$ or she begins by pulling the thinking arm. In the latter case, she switches when the time remaining is $q^{-1}(\bar{p})$, which has a unique solution by Lemma 6, as $\bar{p} < 1$.

Thus, if there are two candidate strategies satisfying the necessary conditions, then the agent needs to split the time spent on the doing arm between two disjoint intervals in at least one of those strategies. If the agent splits her time doing in at least one solution and the two solutions differ on a positive measure of time, then there must be two different lengths of the Hail Mary period, τ_3' and τ_3, both of which satisfy the necessary conditions. Assume without loss of generality that $\tau_3' > \tau_3$. Both τ_3' and τ_3 have associated terminal beliefs, $\underline{p}'$ and $\underline{p}$. The terminal belief is the agent's belief at the deadline conditional on failing to find any solution. Note that for the case of τ_3, the agent's strategy must involve two distinct doing periods. We proceed by cases and derive a contradiction for each of them.

Assume $\underline{p} > \underline{p}'$. Consider the agent's belief with τ_3 periods remaining, and assume that she pulled the doing arm in the interval $[\tau_3', \tau_3)$ with initial belief $q(\tau_3')$. Since $\underline{p} > \underline{p}'$, the agent has to hold a belief $\tilde{p}(\tau_3) < q(\tau_3)$ with τ_3 periods remaining. However, then the agent prefers to pull the thinking arm with τ_3 periods remaining by Lemma 4, which is a contradiction.

Assume $\underline{p} = \underline{p}'$. Consider the agent's policy function under the strategy that implies the last switch to occur at τ_3': γ_τ'. The necessary conditions imply that $\gamma_{\tau_3'}' = 0$. Because the terminal beliefs coincide, the policy function and hence the strategy in the continuation

game for $\tau < \tau_3$ coincide with the policy function and the strategy corresponding to a Hail Mary period of length τ_3 only—as the terminal condition γ_0 depends only on the terminal belief. In turn, this observation implies that $\gamma'_{\tau_3} = \gamma_{\tau_3} = 0$. However, by construction, the agent pulls the doing arm with time remaining $\tau = [\tau'_3, \tau_3]$, implying that $\gamma'_\tau \leq 0$ on this interval. As a consequence, γ'_τ has to have a critical point at τ_3. The arguments in the proof of Lemma 3 imply that γ'_τ is strictly concave at any critical point, and thus, γ'_τ attains a maximum at τ_3. As the beliefs coincide at τ_3, the policy functions under both strategies attain a maximum at τ_3. By Lemma 3, none of the policy functions will attain a maximum, and thus, the agent pulls the doing arm throughout under both policy functions, contradicting the assumption that the strategies differ and that there is a switching time $\tau'_3 > \tau_3$.

Assume $\underline{p} < \underline{p}'$. In this case, the agent's overall time spent on the doing arm must be smaller with switching time τ'_3 than with τ_3. This implies that both strategies involve two distinct doing periods. Moreover, $\tau_2(\tau'_3) > \tau_2(\tau_3)$ for both τ_3 and τ'_3 to be a solution to the fixed point problem. By Lemma 6, τ_2 decreases in τ_3, which is a contradiction. $\qquad\square$

C.5   Proof of Proposition 5

Proof. The probability that the agent obtains a success before the deadline for any τ_1, τ_2, and τ_3 is

$$P(\tau_1, \tau_2, \tau_3; T) = \bar{p}(1 - e^{-\lambda\tau_1}) + (\bar{p}e^{-\lambda\tau_1} + 1 - \bar{p}) \tag{13}$$
$$\cdot \left(1 - e^{-\mu\tau_2} - \frac{e^{-\nu T} - e^{-\mu\tau_2 - \nu(T - \tau_2)}}{\mu - \nu} + e^{-\mu\tau_2} \frac{\bar{p}e^{-\lambda\tau_1}}{\bar{p}e^{-\lambda\tau_1} + 1 - \bar{p}} (1 - e^{-\lambda\tau_3}) \right) \tag{14}$$

where $\tau_3 = T - \tau_1 - \tau_2$. Consider the derivative of $P(\tau_1, \tau_2, \tau_3; T)$ with respect to τ_1, which is

$$e^{-\nu(T - \tau_1) - \mu\tau_2 - \lambda\tau_1} \frac{e^{\mu\tau - 2} - e^{\nu\tau_2}}{\mu - \nu} \mu(\bar{p}(\lambda - \nu) - (1 - \bar{p})e^{\lambda\tau_1}\nu). \tag{15}$$

Its sign is determined by the sign of the last term, which is negative whenever $\nu \geq p_{\tau_1}\lambda$. This condition is satisfied by our assumptions; in particular, it is a consequence of the relative concavity assumption. $\qquad\square$

C.6   Proof of Proposition 6

Proof. Suppose towards a contradiction that τ_1 decreases in T. In particular, consider two scenarios: (i) deadline T and (ii) deadline $T' > T$. Moreover, suppose that the associated initial doing periods are such that $\tau'_1 < \tau_1$.

The belief $q(\tau'_3)$ that the agent holds during the thinking period in scenario (ii) is larger than the belief $q(\tau_3)$ the agent holds in scenario (i). By Lemmata 4 and 6, $q(\cdot)$ is monotonic and increasing in τ_3, which implies $\tau'_3 > \tau_3$.

Consider both scenarios with time τ_3 remaining. In scenario (i), the agent is indifferent between thinking and doing by construction and prefers doing for the remainder of the time. In scenario (ii), she prefers doing at τ_3 and for the remainder of time because $\tau'_3 > \tau_3$. This implies for the corresponding beliefs at remaining time τ_3 that $p'_{\tau_3} \geq p_{\tau_3} = q(\tau_3)$. For any subsequent period, the agent pulls the doing arm in both scenarios. It follows that the terminal beliefs are $\underline{p}' \geq \underline{p}$.

Because in both scenarios the agent starts with a belief $\bar{p}$, a larger terminal belief in scenario (ii) implies that the maximum time the agent pulls the doing arm decreases in this scenario compared to scenario (i); i.e., $\tau_1 + \tau_3 \geq \tau_1' + \tau_3'$. Because $T' > T$, it follows that $\tau_2' > \tau_2$. By Lemma 6, $\tau_3' \geq \tau_3$ implies $\tau_2' \leq \tau_2$. In addition, $\frac{d\dot{y}}{dp} < 0$, implying that τ_2 decreases in the belief as well. Thus, $\tau_3' \geq \tau_3$ and $q(\tau_3') \geq q(\tau_3)$ imply $\tau_2' \leq \tau_2$, which is a contradiction.

The length of the Hail Mary period increases in T if the agent immediately enters this period and is constant whenever the agent starts by thinking. Finally, because τ_1 is nondecreasing, it follows that $q(\tau_3)$ is nonincreasing, and thus, by Lemma 6, τ_3 nonincreasing.

Whenever the agent has no initial doing period, τ_2 is trivially nondecreasing in T. Because τ_3 is nonincreasing when $\tau_1 > 0$, it follows from Lemma 6 that τ_2 weakly increases. Finally, $\tau_1 + \tau_3$ is bounded because $q^{-1}(\tau_3)$ is bounded by $\hat{p} > 0$ defined in Proposition 1, which in turn implies that both τ_1 and τ_3 are bounded. However, then, because a solution exists for every T, we must have that $\tau_2 \to \infty$ as $T \to \infty$. $\qquad\square$

C.7 Proof of Proposition 7

Proof. We prove each item separately. At several points in the proof, we invoke Lemmata 5 and 6, which can be found in Appendix C.4 Proof of Proposition 4.

Proof of item 1. $\tilde{p}$ is constructed such that $\dot{y}(0, \tilde{p}, q(\tilde{p})) = 0$. Thus, if $\bar{p} > \tilde{p}$, then

$$\underbrace{\mu V'(q^{-1}(\bar{p})) - \bar{p}\lambda\mu(B - V(q^{-1}(\bar{p}))) + (\mu - \lambda\bar{p})c}_{=\dot{y}(0,\bar{p},q(\bar{p}))} < 0,$$

because $q^{-1}(p)$ is monotonic for $p \in (0,1)$ and the LHS is decreasing since $-V''(\tau)/V'(\tau) > \lambda\bar{p}$ by the relative concavity assumption. Using Lemma 5, this implies that it cannot be optimal to switch from thinking to doing with $q^{-1}(\bar{p})$ remaining or, equivalently, with a belief $\bar{p} > \tilde{p}$.

Proof of item 2. Observe that the sign of function $\dot{y}(s; p, \xi)$ is the same as the sign of function $\frac{d\gamma_\tau}{d\tau}$ conditional on $\gamma_\xi = 0$ and an agent that thinks with time remaining $\tau \in [\xi, s+\xi]$. A necessary condition for an initial doing period is that there exists a $\tau_2 > 0$ such that $y(\tau_2; q(\tau_3), \tau_3) = 0$ with $\tau_3 = q^{-1}(p)$ and $p < \bar{p}$ the belief held during the thinking period. Because the agent (i) is indifferent with time $\tau_2 + \tau_3$ remaining, (ii) pulls the doing arm for a positive measure of time before that and (iii) expects to pull the thinking arm for a positive measure of time thereafter, $\dot{y}(\tau_2; p, \tau_3) < 0$ by the same arguments that proved Lemma 5. Because $-V''(\tau)/V'(\tau) > p\lambda V'$ by assumption, $\dot{y}(\cdot)$ decreases in τ_2. Thus, if $\lim_{\tau_2 \to \infty} \dot{y}(\tau_2; p, \tau_3) \geq 0$, then there is no second root of $y(\tau_2; p, \tau_3)$ for any $\tau_2 > 0$. It follows that if $\bar{p} < \hat{p}$, then

$$\lim_{\tau_2 \to \infty} \dot{y}(\tau_2; p, \tau_3) \geq 0 \Leftrightarrow p < \frac{V'(\tau_2 + \tau_3) + c}{\lambda\left(B + \frac{c}{\mu} - V(\tau_2 + \tau_3)\right)} = \frac{c}{\lambda\left(B + \frac{c}{\mu} - V(\infty)\right)} = \hat{p}.$$

Proof of item 3. First, we show that if $\bar{p} \geq \hat{p}$, then the agent enters the Hail Mary period with a belief $p_{\tau_3} \geq \hat{p}$. Suppose towards a contradiction that the agent switches to the Hail Mary period with time τ_3 remaining and a belief $p_{\tau_3} < \hat{p}$. Because $\bar{p} \geq \hat{p}$, there has to be a continuation game with time τ remaining at which the agent is in the initial doing phase and holds a belief $p_\tau < \hat{p}$. By item 2 such a continuation game cannot exist.

Second, observe that if the agent enters the Hail Mary period with a belief $\hat{p}$, then her terminal belief at the end of an unsuccessful Hail Mary period is $\check{p}$.

Third, suppose that the terminal belief was smaller than $\check{p}$. Then, it must be true that with time $q^{-1}(\hat{p})$ remaining, the agent is in the Hail Mary period because the agent enters the final doing period with a belief weakly greater than $\hat{p}$ and because $q^{-1}(\hat{p})$ is the time length of pulling the doing arm required to deteriorate a belief of $\hat{p}$ to $\check{p}$. However, that the terminal belief lies below $\check{p}$ implies that with time $q^{-1}(\hat{p})$ remaining, the agent's belief is less than $\hat{p}$. However, this contradicts the necessary conditions for an optimal strategy, as $q(q^{-1}(\hat{p})) > p_{q^{-1}(\hat{p})}$, which implies that the agent is not in the Hail Mary period. $\qquad\square$

C.8 Proof of Proposition 8

Proof. For this proof only, it is instructive to think of time as moving forward. We start by deriving the probabilities for different events in this setting. Consider an agent at calendar time t. Let $A^{th}(t)$ be the amount of time the agent has pulled the thinking arm in the past. Moreover, observe that the probability that the agent observes *no solution* when pulling the thinking arm for a measure of time A is

$$F(A) := 1 - \frac{\mu e^{-\nu A} - \nu e^{-\mu A}}{\mu - \nu}.$$

The probability of having obtained progress *without having obtained a solution*, Ψ^{pro}, when pulling the thinking arm for a measure of time A is

$$\Psi^{pro}(A) := \frac{e^{-\mu A} - e^{-\nu A}}{\nu e^{-\mu A} - \mu e^{-\nu A}} \mu.$$

Thus, we obtain for the probability of an arrival of a solution at time t, $\hat{f}(A^{th}(t))$

$$\hat{f}(A^{th}(t)) := \left(e^{-\nu A^{th}(t)} - e^{-\mu A^{th}(t)} \right) \frac{\mu \nu}{\mu - \nu}.$$

Note that in our setting $A^{th}(t) = t - A_t$, and we can replace the state variable of the thinking arm $A^{th}(t)$ simply by the state of the doing arm A_t together with calendar time t. Let $1 - G(A_t) := \bar{p}e^{-\lambda A_t} + 1 - \bar{p}$ and $\hat{g}(A_t) := \lambda \bar{p}e^{-\lambda A_t}$. Thus, the Hamiltonian—*in calendar time!*—corresponding to Example 8 is

$$\mathcal{H} = (1 - F(t, A_t))(1 - G(A_t))\left(((1 - a_t)\hat{f}(t, A_t) + a_t\hat{g}(A_t))B - c \right) + a_t\eta_t \qquad (16)$$

with co-state evolution

$$-\dot{\eta}_t = -(F_A(t, A_t)(1 - G(A_t)) + G_A(A_t)(1 - F(t, A_t)))\left(((1 - a_t)\hat{f}(t, A_t) + a_t\hat{g}(A_t))B - c \right) \qquad (17)$$

$$+ (1 - F(t, A_t))(1 - G(A_t))(\hat{f}_A(t, A_t)(1 - a_t) + \hat{g}_A(A_t)a_t)B. \qquad (18)$$

The switching function $\gamma_t = -\frac{d\mathcal{H}}{da_t}$ is

$$\gamma_t = (1 - F(t, A_t))(1 - G(A_t))(\hat{f}(t, A_t) - \hat{g}(A_t))B - \eta_t \qquad (19)$$

with boundary condition $\gamma_T = (1 - F(T, A_T))(1 - G(A_T))(\hat{f}(T, A_T) - \hat{g}(A_T))B$, implying that the agent works on the arm with the higher success probability at the deadline. In particular, she works on the thinking arm at the deadline if $\hat{f}(T, A_T) > \hat{g}(A_T)$.

Next, observe that whenever $\hat{f}(t, A_t) > \hat{g}(A_t)$ at some t, $\hat{f}(t', A_{t'}) > \hat{g}(A_{t'})$ for all $t' > t$. The success probabilities of the arms, $\hat{g}(A_t)$ and $\hat{f}(t, A_t)$, are constant whenever the other arm is pulled. Moreover, $\hat{g}(A_t)$ decreases whenever the doing arm is pulled, while $\hat{f}(t, A_t)$ increases whenever the thinking arm is pulled. Thus, $\hat{f}(t, A_t) > \hat{g}(A_t) \Leftrightarrow \hat{f}(t', A_{t'}) > \hat{g}(A_{t'})$ for all $t' > t$.

The switching function has evolution

$$\dot{\gamma}_t = (1 - F(t, A_t))(1 - G(A_t))(\hat{g}(A_t) - \hat{f}(t, A_t))c. \tag{20}$$

Thus, γ_t increases whenever the probability of solving the problem with the doing arm exceeds the probability of solving it with the thinking arm and decreases otherwise. At the beginning, $\dot{\gamma}_0 > 0$ as $\hat{f}(0, 0) = 0$.

It follows that whenever $\dot{\gamma}_t < 0$ for some t, $\hat{f}(t, A_t) > \hat{g}(A_t)$, and therefore, $\hat{f}(t', A_{t'}) > \hat{g}(A_{t'})$ for all $t' > t$, which implies that (i) if the agent leaves the thinking arm, then she does not return to it, (ii) $\hat{f}(T, A_T) > \hat{g}(A_T)$, and therefore, the agent thinks at the deadline. The continuity of γ_t then implies that the agent never switches from thinking to doing, which proves the claim. $\qquad\square$

References

Aghion, P., C. Harris, P. Howitt, and J. Vickers (2001). "Competition, Imitation and Growth with Step-by-Step Innovation". *Review of Economic Studies* 68 (3), pp. 467–492.

Akcigit, U. and Q. Liu (2015). "The role of information in innovation and competition". *Journal of the European Economic Association* 14 (4), pp. 828–870.

Bergemann, D. and J. Välimäki (2008). "Bandit problems". *The New Palgrave Dictionary of Economics: Volume 1–8*, pp. 336–340.

Bolton, P. and A. Faure-Grimaud (2009). "Thinking Ahead: The Decision Problem". *The Review of Economic Studies* 76 (4), pp. 1205–1238.

Callander, S. (2011). "Searching and Learning by Trial and Error". *American Economic Review* 101 (6), pp. 2277–2308.

CB Insights (2021). *The Top 12 Reasons Startups Fail.* available at: `https://www.cbinsights.com/research/startup-failure-reasons-top/`.

Che, Y.-K. and K. Mierendorff (2019). "Optimal Dynamic Allocation of Attention". *American Economic Review* 109 (8), pp. 2993–3029.

Clarke, F. (2013). *Functional analysis, calculus of variations and optimal control.* Vol. 264. Springer Science & Business Media.

Dean, S. O. (1998). "Fusion power by magnetic confinement program plan". *Journal of fusion energy* 17 (4), pp. 263–287.

Doval, L. (2018). "Whether or not to open Pandora's box". *Journal of Economic Theory* 175, pp. 127–158.

Eisenmann, T. (2021). "Why Start-Ups Fail". *Harvard Business Review.*

Felin, T., A. Gambardella, S. Stern, and T. Zenger (2019). "Lean startup and the business model: Experimentation revisited". *Long Range Planning*, p. 101889.

Fershtman, C. and A. Rubinstein (1997). "A simple model of equilibrium in search procedures". *Journal of Economic Theory* 72 (2), pp. 432–441.

Fershtman, D. and A. Pavan (2021). "Searching for Arms: Experimentation with Endogenous Consideration Sets". *mimeo.*

Forand, J. G. (2015). "Keeping your options open". *Journal of Economic Dynamics and Control* 53, pp. 47–68.

Francetich, A. (2018). "Efficient multi-agent experimentation and multi-choice bandits". *Economics Bulletin* 38 (4), pp. 1757–1761.

Fryer, R. and P. Harms (2019). "Two-Armed Restless Bandits with Imperfect Information: Stochastic Control and Indexability". *Mathematics of Operations Research* 43 (2), pp. 399–427.

Gans, J., E. L. Scott, and S. Stern (2018). "Strategy for start-ups". *Harvard Business Review* 96 (3), pp. 44–51.

Gans, J. S., S. Stern, and J. Wu (2019). "Foundations of entrepreneurial strategy". *Strategic Management Journal* 40 (5), pp. 736–756.

Garfagnini, U. and B. Strulovici (2016). "Social Experimentation with Interdependent and Expanding Technologies". *Review of Economic Studies* 83 (4), pp. 1579–1613.

Gittins, J. and D. Jones (1974). "A dynamic allocation index for the sequential allocation of experiments". *Progress in statistics*, pp. 241–266.

Gompers, P. A. (1995). "Optimal Investment, Monitoring, and the Staging of Venture Capital". *The Journal of Finance* 50 (5), pp. 1461–1489.

Green, B. and C. R. Taylor (2016). "Breakthroughs, Deadlines, and Self-Reported Progress: Contracting for Multistage Projects". *American Economic Review* 106 (12), pp. 3660–99.

Hu, Z. (2014). "Financing Innovation with Unobserved Progress". *mimeo.*

IPCC (2018). *Global Warming of 1.5°C.An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty.* available at: `https://www.ipcc.ch/site/assets/uploads/sites/2/2019/06/SR15_Full_Report_Low_Res.pdf`.

Isenberg, D. and A. DiFiore (2020). "You Don't Have to Pivot in a Crisis". *Harvard Business Review.*

Kaplan, S. N. and P. Strömberg (2003). "Financial Contracting Theory Meets the Real World: An Empirical Analysis of Venture Capital Contracts". *The Review of Economic Studies* 70 (2), pp. 281–315.

Keller, G. and A. Oldale (2003). "Branching bandits: a sequential search process with correlated pay-offs". *Journal of Economic Theory* 113 (2), pp. 302–315.

Keller, G., S. Rady, and M. Cripps (2005). "Strategic Experimentation with Exponential Bandits". *Econometrica* 73 (1), pp. 39–68.

Kerr, W. R., R. Nanda, and M. Rhodes-Kropf (2014). "Entrepreneurship as Experimentation". *Journal of Economic Perspectives* 28 (3), pp. 25–48.

Kim, Y. (2021). "The Direct vs. the Sequential Approach in Project Management". *mimeo.*

Kirtley, J. and S. O'Mahony (2020). "What is a pivot? Explaining when and how entrepreneurial firms decide to make strategic change and pivot". *Strategic Management Journal*, pp. 1–34.

Klein, N. (2016). "The importance of being honest". *Theoretical Economics* 11 (3), pp. 773–811.

Kramer, D. (2020). "Investments in privately funded fusion ventures grow". *Physics Today* 2, 1013a.

Lemus, J. and E. Temnyalov (2019). "Diversification and Information in Contests". *mimeo.*

Letina, I. (2016). "The road not taken: competition and the R&D portfolio". *RAND Journal of Economics* 47 (2), pp. 433–460.

Lierop, W. van (2019). *Fusion Energy: Who Has The Courage To Take It To Market.* available at: `https://www.forbes.com/sites/walvanlierop/2019/08/21/fusion-energy-who-has-the-courage-to-take-it-to-market/`.

Machine That Changed The World, T. (1990). *The Paperback Computer.* Interview with Steve Jobs. Available at: `http://openvault.wgbh.org/catalog/V_AD9E0BC353BF435E83F28DEF165D4F40`.

McDonald, R. and R. Bremner (2020). "When It's Time to Pivot, What's Your Story?" *Harvard Business Review.*

Moroni, S. (2021). "Experimentation in Organizations". *Theoretical Economics* forthcoming.

Nikandrova, A. and R. Pancs (2018). "Dynamic project selection". *Theoretical Economics* 13 (1), pp. 115–143.

Olszewski, W. and R. Weber (2015). "A more general Pandora rule?" *Journal of Economic Theory* 160, pp. 429–437.

Rahmani, M. and K. Ramachandran (2021). "Delegating Innovation Projects with Deadline: Committed vs. Flexible Stopping". *Management Science* 67 (10), pp. 5969–6627.

Rothschild, M. (1974). "A two-armed bandit theory of market pricing". *Journal of Economic Theory* 9 (2), pp. 185–202.

Weitzman, M. L. (1979). "Optimal search for the best alternative". *Econometrica* 47 (3), pp. 641–654.

Wolf, C. (2018). "Informative Milestones in Experimentation". *mimeo.*

D   Supplementary Material

In this part, we verify that the examples considered in the main text satisfy our assumptions, consider the extension to the value of progress being dependent on the state A_τ, consider an example in which the relative concavity assumption does not hold, and provide a sufficient condition such that the agent never wants to shirk.

D.1   Verification of Assumptions for Examples

In this part, we verify that the examples discussed in Section 4.3 satisfy the assumptions of our model.

Verification of Example 1 & Example 4

The value of such an arm given remaining time τ is

$$V(\tau) = (1 - e^{-\nu\tau}) \left(B_\nu - \frac{c_\nu}{\nu} \right),$$

with

$$V(0) = 0, \ V'(\tau) = e^{-\nu\tau} \left(\nu B_\nu - c_\nu \right), \ V''(\tau) = -\nu e^{-\nu\tau} \left(\nu B_\nu - c_\nu \right), \ -\frac{V''(\tau)}{V'(\tau)} = \nu.$$

Verification of Example 2

As with any risky bandit, the agent would have an incentive to eventually stop pulling the arm if her time were unlimited. Assume that this occurs after pulling the arm for $\hat{t}$ periods.

$$V(\tau) = \begin{cases} \bar{p}^\nu (1 - e^{-\nu\tau}) \left(B_\nu - \frac{c_\nu}{\nu} \right) - (1 - \bar{p}^\nu) c_\nu \tau & , \text{ if } \tau \le \hat{t} \\ \bar{p}^\nu B_\nu - \frac{c_\nu}{\nu} (1 + (1 - \bar{p}^\nu) \hat{t}) & , \text{ if } \tau \ge \hat{t}, \end{cases}$$

where $\hat{t}$ is the time at which an agent with initial belief $\bar{p}^\nu$ would stop experimenting on the new arm. Such an arm satisfies our desired conditions when the initial belief is sufficiently high given the deadline, i.e., whenever $\tau \le \hat{t}$:[39]

$$V(0) = 0$$
$$V'(\tau) = \bar{p}^\nu e^{-\nu\tau} \left(\nu B_\nu - c_\nu \right) - (1 - \bar{p}^\nu) c_\nu$$
$$V''(\tau) = -\nu \bar{p}^\nu (e^{-\nu\tau}) \left(\nu B_\nu - c_\nu \right)$$
$$-\frac{V''(\tau)}{V'(\tau)} = \nu \frac{\bar{p}^\nu (e^{-\nu\tau}) \left(\nu B_\nu - c_\nu \right)}{\bar{p}^\nu e^{-\nu\tau} \left(\nu B_\nu - c_\nu \right) - (1 - \bar{p}^\nu) c_\nu} > \nu.$$

Verification of Example 3

In this version of the model, the value does not have a closed form solution, but we can verify that our assumptions are satisfied whenever $\beta < \hat{\beta} := e^\alpha \nu + \frac{\bar{p}\lambda}{\nu} \frac{ce^{-\alpha} - \nu B}{\nu B} - \frac{c}{B}$.[40]

$$V(\tau) = \int_0^\tau e^{-\int_0^t \nu e^{\alpha + \beta s} ds} \left(\nu e^{\alpha + \beta t} B - c \right) dt < B$$
$$V(0) = 0$$
$$V'(\tau) = e^{-\nu e^{\alpha + \beta\tau} \tau} \left(\nu e^{\alpha + \beta\tau} B - c \right) > 0$$
$$V''(\tau) = -\nu e^{\alpha + \beta\tau - e^{\alpha + \beta\tau}} \left((1 + \beta\tau)(\nu e^{\alpha + \beta\tau} B - c - \beta B) \right) < 0$$
$$-\frac{V''(\tau)}{V'(\tau)} = e^{\alpha + \beta\tau}(1 + \beta\tau) - \beta \frac{\nu B e^{\alpha + \beta\tau}}{\nu B e^{\alpha + \beta\tau} - c} > \bar{p}\lambda.$$

To see the sign of $V''(\tau)$, note that the term in parentheses is increasing in τ and positive for $\tau = 0$, which determines the sign of the second derivative of the value. To see the sign

---

[39]Note that $\hat{t}$ increases in $\bar{p}^\nu$ and converges to ∞ as $\bar{p}^\nu \to 1$, the case in which Example 2 converges to Example 1.

[40]Depending on parameters, this bound can be either positive or negative. In particular, it is strictly positive whenever ν is sufficiently high.

of $-\frac{V''(\tau)}{V'(\tau)}$, note that this expression is also increasing in τ. Evaluating $-\frac{V''(0)}{V'(0)}$ delivers the desired expression for $\hat{\beta}$, which ensures that our conditions are satisfied.

Verification of Example 5

The expected value at time τ of $b(t)$, where $b(t)$ follows an Ornstein-Uhlenbeck process, yields the benefit of triggering this payoff stream

$$V(\tau) = b(0)e^{-\nu\tau} + B_\nu(1 - e^{-\nu\tau})$$
$$= B_\nu(1 - e^{-\nu\tau}).$$

Thus, we obtain

$$V(0) = 0, \ V'(\tau) = \nu B_\nu e^{-\nu\tau}, \ V''(\tau) = -\nu^2 B_\nu e^{-\nu\tau} \ -\frac{V''(\tau)}{V'(\tau)} = \nu.$$

D.2 Discussion of Example 6

Consider the model of Example 6. Note that whenever p_τ is such that $\nu = p_\tau \lambda$, the continuation game satisfies our assumptions, and therefore, Proposition 3 applies to the continuation game.

Assume that for some time remaining $\tau < T$, the belief is indeed such that $p_\tau \lambda = \nu$ and that $p_\tau < \bar{p}$. If in addition $\gamma_\tau < 0$, then we know by the continuity of γ_τ that there is a neighborhood of remaining time $\tau + \varepsilon > \tau$ such that Proposition 3 continues to hold in this neighborhood too.

However, in this neighborhood, $p_{\tau+\varepsilon}\lambda > \nu$, which violates the assumption on relative concavity. Thus, in particular, Lemma 3 may be violated, which, in turn, implies that—once said neighborhood becomes large—eventually $\gamma_{\tau+\varepsilon}$ may be increasing and may become positive. As a consequence, the agent may engage in an initial thinking period before returning to the path described by Proposition 3. A numerical solution of Example 6 for various deadline lengths is provided in Figure 5. As we see, once T is large enough, the optimal policy adds an initial thinking period.

It is apparent from Figure 5 that if the time horizon is long enough, then the agent will indeed start by thinking before reverting to our doing-thinking-doing pattern. The reason is as follows: once $\nu < \bar{p}\lambda$ and there is plenty of time remaining, the deadline effect and hence the time pressure is not the agent's primary concern. Instead, in this example, the payoff of successful thinking with a sufficiently long deadline is *higher* than the payoff of successful doing. Thus, with low time pressure, thinking has a payoff advantage over doing and is preferred until the time pressure deteriorates the value of progress on the thinking arm. Once the value of progress is low enough, the intuition and, eventually, the formal analysis of our main model apply again.

D.3 Belief-Dependent Continuation Value

In this part, we show how the key lemmata (Lemmata 1 to 3) that lead to Proposition 3 extend to the case in which the continuation value of an arrival on the thinking arm also depends on the belief about the doing arm. Moreover, we verify that Proposition 3 also holds for the example of a risky new arm without a switching cost.

Modified optimal control problem. When the continuation value also depends on A_τ, denote by $V(\tau, A_\tau)$ the value of an arrival on the thinking arm. The Hamiltonian
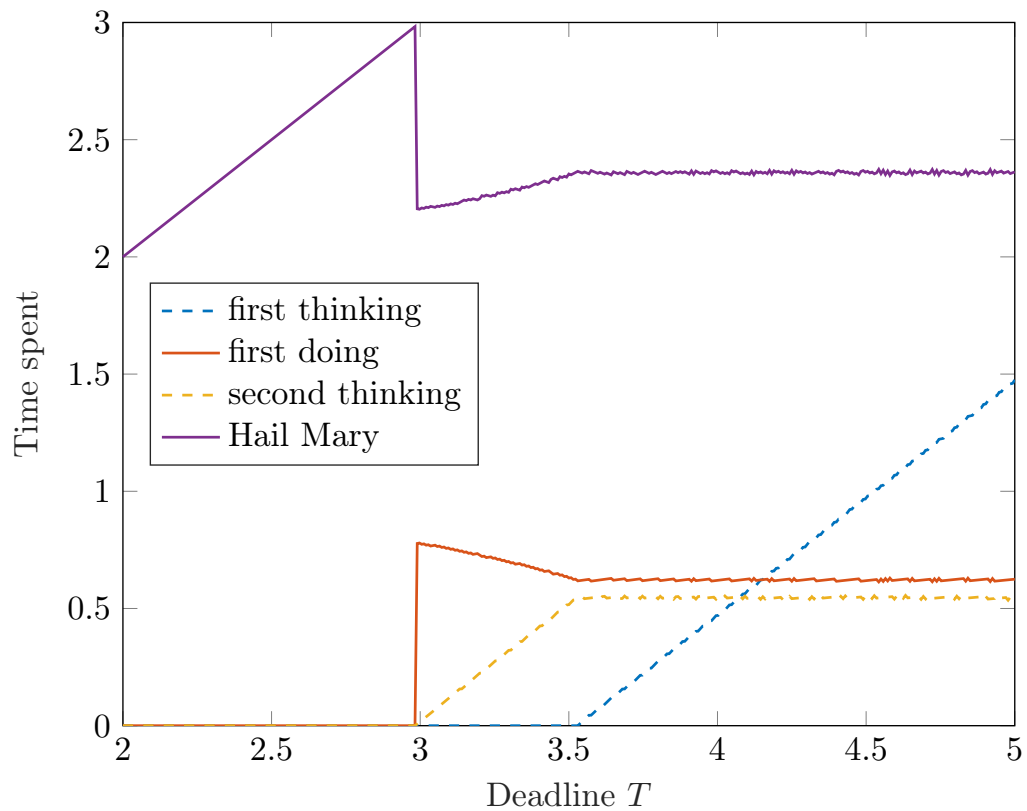
Figure 5: *Interval length by deadlines without relative concavity assumption.* The figure shows a numerical solution of the time spent in each interval using a particular approach as a function of the deadline.

Parameters: $B = 9, \bar{p} = 0.8, c = 0.5, \lambda = 1, \mu = 0.4, \nu = 0.5.$

corresponding to the modified optimal control problem becomes

$$\begin{aligned}
H_\tau(a_\tau; A_\tau) :=& e^{-\mu(T-\tau-A_\tau)}(1 - \bar{p} + \bar{p}e^{-\lambda A_\tau})J(p_\tau, \tau, a_\tau) + a_\tau \eta_\tau, \\
=& e^{-\mu(T-\tau-A_\tau)}(1 - \bar{p})\left((1 - a_\tau)\mu V(\tau, A_\tau) - c\right) \\
& + e^{-\mu(T-\tau-\mu A_\tau)}\bar{p}e^{-\lambda A_\tau}\left((1 - a_\tau)\mu V(\tau, A_\tau) + a_\tau \lambda B - c\right) + a_\tau \eta_\tau
\end{aligned}$$

and the co-state evolution becomes

$$\begin{aligned}
\dot{\eta}_\tau =\ & e^{-\mu(T-\tau-A_\tau)}(\bar{p}e^{-\lambda A_\tau} + 1 - \bar{p}) \\
& \cdot \left(p_\tau \lambda(\mu - \lambda)a_\tau B + \mu(1 - a_\tau)\left((\mu - \lambda p_\tau)V(\tau, A_\tau) + V_A(\tau, A_\tau)\right) - (\mu - p_\tau \lambda)c\right).
\end{aligned}$$

Corresponding Lemma 1. Lemma 1 holds trivially because $\lim_{\tau \to 0} V(\tau, A_\tau) = 0$ in this case as well and the boundary condition of the optimal control problem is unchanged, $\eta_{\tau=0} = 0$.

Corresponding Lemma 2. The resulting switching function is

$$\gamma_\tau = e^{-\mu(T-\tau-A_\tau)}(\bar{p}e^{-\lambda A_\tau} + 1 - \bar{p})(\mu V(\tau, A_\tau) - p_\tau \lambda B) - \eta_\tau$$

with evolution

$$\dot{\gamma}_\tau = e^{-\mu(T-\tau-A_\tau)}(\bar{p}e^{-\lambda A_\tau} + 1 - \bar{p})\left(p_\tau \lambda \mu(V(\tau, A_\tau) - B) + \mu(V_\tau(\tau, A_\tau) - V_A(\tau, A_\tau)) + (\mu - p_\tau \lambda)c\right).$$

Corresponding Lemma 3. The derivative with respect to τ of the analogue of $g(\tau)$ in the switching function, as in Lemma 3, is

$$\begin{aligned}
\ddot{\gamma}_\tau = \mu\Bigg(& \frac{dp_\tau}{d\tau}\lambda\left(B - V(\tau, A_\tau) + \frac{c}{\mu}\right) \\
& + \lambda p_\tau(V_\tau(\tau, A_\tau) + a_\tau V_A(\tau, A_\tau)) + a_\tau(a_\tau V_{A,A}(\tau, A_\tau) + 2V_{\tau,A}(\tau, A_\tau) - V_{\tau,\tau}(\tau, A_\tau))\Bigg).
\end{aligned}$$

Observe that $\ddot{\gamma}_\tau$ is negative for all a_τ if $V(\tau, A_\tau) \leq B + \frac{c}{\mu}$ and[41]

$$-\frac{\frac{d^2}{d\tau^2}V(\tau, A_\tau)}{\frac{d}{d\tau}V(\tau, A_\tau)} \geq p_\tau \lambda$$

as $\frac{dp_\tau}{d\tau} \geq 0$.

Thus, no interior local minimum can exist under these assumptions, which are analogous to those in the main text.[42]

Corresponding Proposition 3. The proof of the analogous result as in Proposition 3 follows directly by combining the corresponding Lemmata 1 to 3.

---

[41]Where the total derivative with respect to time is $\partial_\tau V(\tau, A_\tau) - \partial_A V(\tau, A_\tau)a_\tau$.

[42]Note that we could dispense with the assumption $V(\tau, A_\tau) \leq B + \frac{c}{\mu}$ by plugging $\dot{\gamma}_\tau = 0$ into $\ddot{\gamma}_\tau$, as in the proof Lemma 3.

Verification of Example 7

We next verify that a simple risky new arm satisfies our assumptions when there is no cost of switching between doing arms and the agent may mix continuously between the two arms. To save on notation and case distinctions, we assume that $\bar{p}^\nu > \bar{p}$ and $\bar{p} \leq 2/3$ and set $\lambda = 1$ for both arms.

Lemma 7 constructs the reduced form $V(\tau; A_\tau)$ for this example. Lemma 8 shows that the constructed $V(\tau; A_\tau)$ satisfies the condition required for the Corresponding Proposition 3, $\left(-\frac{d^2}{d\tau^2} V(\tau, A_\tau)\right) / \left(\frac{d}{d\tau} V(\tau, A_\tau)\right) \geq p_\tau \lambda$.

Lemma 7. *While holding a belief $p_\tau = \bar{p} e^{-\lambda A_\tau} / (\bar{p} e^{-\lambda A_\tau} + 1 - \bar{p})$ on the initial doing arm with time remaining τ, the value of having access to a new risky doing arm with initial belief $\bar{p}^\nu$ is*

$$V(\tau, A_\tau) = \bar{p}^\nu (1 - e^{-\min\{\hat{t}(A_\tau), \tau\}}) (B - c) - (1 - \bar{p}^\nu) c \min\{\hat{t}(A_\tau), \tau\}$$
$$+ (1 - \bar{p}^\nu) \frac{p_\tau}{1 - p_\tau} \left((B - c) \left(p_\tau^2 (1 - e^{-\hat{\tau}}) + 2p_\tau(1 - p_\tau)(1 - e^{-\frac{1}{2}\hat{\tau}})\right)\right)$$
$$- c \left(2p_\tau(1 - p_\tau)(1 - e^{-\frac{1}{2}\hat{\tau}}) + (1 - p_\tau)^2 \hat{\tau}\right),$$

where $\hat{t}(A_\tau) = \ln\left(\frac{\bar{p}^\nu}{1 - \bar{p}^\nu} \frac{1 - p_\tau}{p_\tau}\right)$, which is the time at which after discovering the new arm, the agent switches from pulling the new arm exclusively to mixing between both arms whenever $\tau > \hat{t}(A_\tau)$.

Proof. During the time in which the agent exclusively uses the new arm, i.e., for the periods $t \in [0, \min\{\hat{t}(A_\tau), \tau\}]$, the agent's payoff is

$$V^{new}(\tau; A_\tau) := \bar{p}^\nu (1 - e^{-\min\{\hat{t}(A_\tau), \tau\}}) (B - c) - (1 - \bar{p}^\nu) c \min\{\hat{t}(A_\tau), \tau\}.$$

With probability $\frac{1 - \bar{p}^\nu}{1 - p_\tau}$, the agent does not obtain a success before $\min\{\tau, \hat{t}\}$. In this event, the agent will mix for the remaining time $\hat{\tau} := \min\{0, \tau - \hat{t}\}$. Because the agent mixes between the two arms instead of using a single arm at a full rate, the beliefs decline at a lower rate on each arm, while the instantaneous success rate at time τ is still p_τ. In particular, both arms are pulled with the same intensity $a_\tau = \frac{1}{2}$, which implies that

$$\dot{p}_\tau^\nu = \dot{p}_\tau = \frac{1}{2} p_\tau (1 - p_\tau).$$

Solving for the agent's value upon mixing for the remaining time $\hat{\tau}$, we obtain

$$V^{mix}(\hat{\tau}; A_\tau) := \int_0^{\hat{\tau}} e^{-p_\tau} (p_\tau B - c) \, d\tau$$
$$= (B - c) \left(p_\tau^2 (1 - e^{-\hat{\tau}}) + 2p_\tau(1 - p_\tau)(1 - e^{-\frac{1}{2}\hat{\tau}})\right)$$
$$- c \left(2p_\tau(1 - p_\tau)(1 - e^{-\frac{1}{2}\hat{\tau}}) + (1 - p_\tau)^2 \hat{\tau}\right).$$

Putting the pieces together delivers the result. $\square$

Finally, we have to ensure that the relative concavity assumption is satisfied. Note that the maximal time remaining such that the agent uses any of the two arms is bounded from above, as eventually the beliefs would become too low to generate a positive expected payoff. Denote this upper bound, which we explicitly define below, by $\bar{\tau}$.

Lemma 8. *Under the assumptions that $\bar{p}^\nu > \bar{p}$, $\bar{p} \leq 2/3$, and $\lambda = 1$, $V(\tau, A_\tau)$ satisfies*

$$-\frac{\frac{d^2}{d\tau^2}V(\tau, A_\tau)}{\frac{d}{d\tau}V(\tau, A_\tau)} \geq p_\tau.$$

Proof. Note that varying the time remaining has different effects on $V(\tau, A_\tau)$ depending on whether the agent thinks or does. If she does, A_τ and τ vary both, affecting the value of a thinking success. If she thinks, only the change in τ affects the value of a thinking success.

To verify our assumptions, note that[43]

$$\frac{d}{d\tau}V(\tau, p_\tau) = \frac{\partial V}{\partial \tau} + \dot{p}_\tau \frac{\partial V}{\partial p}$$

$$\frac{d^2}{d\tau^2}V(\tau, p_\tau) = \frac{\partial^2 V}{(\partial \tau)^2} + 2\frac{\partial^2 V}{\partial \tau \partial p}\dot{p}_\tau + \dot{p}_\tau^2 \frac{\partial^2 V}{(\partial p)^2} + \ddot{p}_\tau \frac{\partial V}{\partial p}.$$

Moreover, $V(\tau, p_\tau) = V^{new} + \frac{1-\bar{p}^\nu}{1-p_\tau}V^{mix}$.

If the deadline is too close when thinking is successful, then we are in the case of Example 2, and our assumptions are satisfied. The domain of time remaining under which mixing is relevant is $\tau \in [\underline{\tau}, \overline{\tau}]$, where $\overline{\tau} := \ln\left(\frac{\bar{p}^\nu}{1-\bar{p}^\nu}\frac{1-p_\tau}{p_\tau}\right) + 2\ln\left(\frac{p_\tau}{1-p_\tau}\frac{B-c}{c}\right)$ is the time at which the agent would prefer shirking over working on either, as both beliefs have declined too much, and where $\underline{\tau} := \ln\left(\frac{\bar{p}^\nu}{1-\bar{p}^\nu}\frac{1-p_\tau}{p_\tau}\right)$ is the time at which the belief about the new arm has declined to the current belief of the doing arm.

To simplify the notation, define $V^{p-mix} := \frac{1-\bar{p}^\nu}{1-p_\tau}V^{mix}$, and note that for $b, d > 0$, if $\frac{a}{b} \geq x$ and $\frac{c}{d} \geq x$, then $\frac{a+c}{b+d} \geq x$.[44] Thus, it is sufficient to show that (a) $\frac{-\frac{d^2}{d\tau^2}V^{new}}{\frac{d}{d\tau}V^{new}} > p_\tau$ and (b) $\frac{-\frac{d^2}{d\tau^2}V^{p-mix}}{\frac{d}{d\tau}V^{p-mix}} \geq p_\tau$.

It is straightforward to see that (a) is satisfied as

$$\frac{-d^2 V^{new}}{d\tau^2}\bigg/\frac{dV^{new}}{d\tau} = \frac{p_\tau(B-c)\frac{1-\bar{p}^\nu}{1-p_\tau}}{(p_\tau B - c)\frac{1-\bar{p}^\nu}{1-p_\tau}} > p_\tau.$$

Next, we show (b), i.e., that $\frac{-\frac{d^2}{d\tau^2}V^{p-mix}}{\frac{d}{d\tau}V^{p-mix}} \geq p_\tau$. Note that both the numerator and denominator are positive. Hence, a lower bound for the fraction is given by dividing the lower bound of $-\frac{d^2}{d\tau^2}V^{p-mix}$ by the upper bound of $\frac{d}{d\tau}V^{p-mix}$. We obtain the bounds by showing that for any feasible parameter constellation, (i) the smallest numerator is attained for $\overline{\tau}$.[45] and (ii) the greatest numerator is attained for $\underline{\tau}$.[46]

To see (i), observe that

$$\frac{d^2}{d\tau^2}\left(-\frac{d^2}{d\tau^2}V^{p-mix}\right) = -\frac{p_\tau \bar{p}^\nu}{\sqrt{\frac{\bar{p}^\nu}{1-\bar{p}^\nu}\frac{1-p_\tau}{p_\tau}}}\left((B-2c)e^{\frac{\tau}{2}}(1-p_\tau)^2 + (B-2c)p_\tau(3-2p_\tau)\sqrt{\frac{\bar{p}^\nu}{1-\bar{p}^\nu}\frac{1-p_\tau}{p_\tau}}\right)$$

$$< 0.$$

---

[43]We simplify notation by supressing arguments whenever it should not cause confusion. Moreover, we use the belief as the state variable, which is equivalent to using A_τ.

[44]Observe that $\frac{a}{b} \geq x$ and that $\frac{c}{d} \geq x$ imply $a > bx$ and $c > dx$; thus, $a + c > (b + d)x$.

[45]This is the upper bound on the deadline such that the agent is willing to exert effort for any parameters and time remaining.

[46]This is the lower bound on the deadline such that the mixing phase is reached.

Hence, the derivative of $\left(-\frac{d^2}{d\tau^2}V^{p-mix}\right)$ is decreasing. Next, note that

$$\lim_{\tau\to\bar\tau}\frac{d}{d\tau}\left(-\frac{d^2}{d\tau^2}V^{p-mix}\right) = (1-\bar p^\nu)(1-p_\tau)p_\tau\frac{p_\tau Bc}{B-c} > 0$$

and therefore that $-\frac{d^2}{d\tau^2}V^{p-mix}$ is increasing for all τ on the relevant domain. Hence, its lower bound is attained at $\tau = \underline\tau$ and is given by

$$\lim_{\tau\to\underline\tau}\left(-\frac{d^2}{d\tau^2}V^{p-mix}\right) = \frac{1-\bar p^\nu}{1-p_\tau}\frac{p_\tau}{B-c}\left(B^2(p_\tau(p_\tau(2p_\tau-5)+4)-2) + 2Bc(3-2p_\tau)p_\tau^2\right.$$
$$\left. -2c(1-p_\tau)^2(1-2p_\tau)(B-c)\ln\left(\frac{p_\tau(B-c)}{c(1-p_\tau)}\right) - c^2(4(2-p_\tau)p_\tau-3)\right).$$

To see (ii), observe that

$$\frac{d^2}{d\tau^2}\left(\frac{d}{d\tau}V^{p-mix}\right) = -\frac{p_\tau\bar p^\nu}{2}e^{-\tau}\left(\frac{(p_\tau-1)e^{\frac{\tau}{2}}(B-2c)}{\sqrt{\frac{(p_\tau-1)\bar p^\nu}{p_\tau(\bar p^\nu-1)}}} - 2Bp_\tau + 2cp_\tau\right) > 0$$

which implies that $\frac{d}{d\tau}\left(\frac{d}{d\tau}V^{p-mix}\right)$ is increasing. At the upper bound, $\lim_{\tau\to\bar\tau}\frac{d^2}{d\tau^2}\left(\frac{d}{d\tau}V^{p-mix}\right) = 0$. Hence, the first derivative of $\frac{d}{d\tau}V^{p-mix}$ with respect to τ is negative throughout, and the upper bound is attained at $\tau \to \underline\tau$, with

$$\lim_{\tau\to\underline\tau}\frac{d}{d\tau}V^{p-mix} = \frac{2(1-\bar p^\nu)(Bp_\tau-c)}{1-p_\tau} > 0.$$

We therefore obtain that

$$\frac{-\frac{d^2}{d\tau^2}V^{p-mix}}{\frac{d}{d\tau}V^{p-mix}}$$
$$> \frac{B^2(p_\tau(p_\tau(2p_\tau-5)+4)-2) + 2Bc(3-2p_\tau)p_\tau^2}{Bc(1-p_\tau)^2}$$
$$- \frac{2(1-p_\tau)^2(1-2p_\tau)(B-c)\ln\left(\frac{p_\tau(B-c)}{c(1-p_\tau)}\right) - c(4(2-p_\tau-2)p_\tau-3)}{Bc(1-p_\tau)^2}.$$

Finally, we need to verify that the right-hand side of the last expression is greater than p_τ. To see this, we compute a lower bound of it using the fact that it is decreasing in c (see below). As mixing requires that $p_\tau B \geq c$, a lower bound is attained for $c = p_\tau B$.

To see that the term is decreasing in c, observe that it is convex in c, as its second derivative is

$$\frac{2B(B(p_\tau(p_\tau(5-2p_\tau)-4)+2)-c)}{c^3(1-p_\tau)^2(B-c)}$$

which is positive, as $B > 2c$ and $p_\tau < 2/3$. Thus, the first derivative is increasing, and at the upper bound of c, it reduces to

$$-\frac{2}{B(1-p_\tau)p_\tau^2} < 0.$$

Hence, a lower bound of the fraction under consideration is attained for $c = p_\tau B$, which is $2/p_\tau$, which is strictly larger than p_τ. □

D.4 No-Shirking Condition

In the text, we assume that B is high enough such that the agent never shirks if she has not yet found a solution. Here, we show that such a B always exists and is finite. Moreover, we provide an (implicit) construction. It is sufficient to show that the agent has an incentive to pull an arm at the deadline. The agent does not shirk if her terminal belief $\underline{p} \geq \frac{c}{\lambda B}$. For any $\underline{p} > 0$ and $T < \infty$, there is a $\underline{B} < \infty$ such that the above condition holds for any $B \geq \underline{B}$. A corollary to Proposition 7 is that for any $T < \infty$, the terminal belief is weakly larger than

$$p^{\min} = \frac{\bar{p}e^{-\lambda T}}{\bar{p}e^{-\lambda T} + 1 - \bar{p}} > 0.$$

By Lemma 4, for any B, there is a $\bar{T} < \infty$ such that $\underline{p} > p^{min}$ for $T > \bar{T}$. Thus, there is a $\underline{B} < \infty$ such that the agent never shirks for *any* deadline, including the limit $T \to \infty$.