The Risk of Failure: Trial and Error Learning and Long-Run Performance*

Steven Callander† Stanford University

Niko Matouschek‡ Northwestern University

September 13, 2017

Abstract

Innovation is often the key to sustained progress, yet innovation itself is difficult and highly risky. Success is not guaranteed as breakthroughs are mixed with setbacks and the path of learning is typically far from smooth. How decision makers learn by trial and error and the efficacy of the process are inextricably linked to the incentives of the decision makers themselves and, in particular, to their tolerance for risk. In this paper we develop a model of trial and error learning with risk averse agents who learn by observing the choices of earlier agents and the outcomes that are realized. We identify sufficient conditions for the existence of optimal actions. We show that behavior within each period varies in risk and performance and that a performance trap develops, such that low performing agents opt to not experiment and thus fail to gain the knowledge necessary to improve performance. We also show that the impact of risk reverberates across periods, leading, on average, to divergence in long-run performance across agents.

Keywords: innovation, learning, risk, long-run performance.

JEL classifications: D81, D83, L25

*We are very grateful for comments from Wouter Dessein, Umberto Garfagnini, Luis Garicano, Bob Gibbons, Johannes Gierlinger, Marina Halac, Richard Holden, Emir Kamenica, Nicolas Lambert, Gilat Levy, John Matsusaka, Andrea Prat, Debraj Ray, Luis Rayo, John Roberts, Mike Ryall, Bruno Strulovici, Balazs Szentes, and participants of numerous conferences and seminars. We thank Can Urgun for excellent research assistance. All remaining errors are our own.

†Graduate School of Business, Stanford University, sjc@stanford.edu.

‡Kellogg School of Management, Northwestern University, n-matouschek@kellogg.northwestern.edu.

1 Introduction

In many endeavors innovation is the key to sustained progress. The challenge, of course, is that innovation itself is difficult and highly risky. A scientist does not know whether a new research agenda will lead to a breakthrough or a dead-end; a firm does not know when introducing a new product—think 'new Coke'—whether there is market demand, or whether a reorganization will produce hoped-for efficiency gains. Indeed, it is possible that efforts to improve performance may actually make things worse, a possibility famously captured by Merton (1936) in his Law of Unintended Consequences.

The uncertainty that is inherent to innovation leads to rich dynamics in experimentation and in long-run performance as the mix of breakthroughs and setbacks generates staccato-like learning paths. This is evident in academic research as fields rise and fall. It is evident in markets as upstart firms rise to challenge for market supremacy and today's leader can become tomorrow's also-ran. This richness does not, however, imply that progress is entirely random. Despite the possibility of displacement, both evidence and intuition indicate that performance is durable, at least in part, and that today's market leader is more likely than not to also be tomorrow's leader (Klepper 2015; Gibbons and Henderson 2013).

The centrality of risk to innovation points to a tight connection between the willingness of agents to experiment and their willingness to engage with risk. To explore this connection, we develop a model of trial and error learning with risk averse agents. We show how risk aversion drives both the decision whether to experiment and the boldness of experimentation of individual agents. We characterize the effects of risk over time, showing how the *within* period effects of risk aversion reverberate *across* time and drive a long-run divergence in performance. This implies a strong path dependence for trial and error learning. Early leaders are more likely than not to hold onto and grow their lead than fall behind. Yet, at the same time, a set-back does not wash out over time, such that a single stroke of bad luck can shape long run performance.

1.1 Modeling Approach

The model we develop builds on an emerging literature of trial-and-error learning that represents the mapping from actions to outcomes as the realized path of a Brownian motion. This setting captures an informationally challenging decision problem. Agents face a continuum of actions from which to choose, yet they possess only a tenuous understanding of how actions map into outcomes. Formally, agents begin with knowledge of only a single *status quo* point. Over time, agents experiment with other, risky actions, and each time an experiment is undertaken a new

point in the mapping is revealed and knowledge accumulates.

An appealing property of the Brownian motion representation is that it captures the reality that learning is not random. In addition to the *practical* knowledge of points in the mapping, agents possess the *theoretical* knowledge of the underlying environment, represented by the drift and variance parameters. Agents combine their theoretical and practical knowledge to form beliefs about *all* actions and use these beliefs to guide their choice. These beliefs are updated and refined as more practical knowledge comes from experience. The agents never fully learn the underlying environment, however, regardless of how many experiments have been undertaken. In this way, the Brownian motion representation allows us to understand not only the decision to experiment or not, but also the direction and the boldness of experimentation, and how experience shapes the trajectory of innovation over time.

In this informationally rich environment, we analyze decision making by a countable sequence of agents who each choose a single action. This represents a long-term perspective on innovation and trial and error learning, capturing the idea that research builds upon the work of those who have come before. As Newton famously remarked about his own breakthroughs, "If I have seen further, it is by standing on the shoulders of giants." Later in the paper, we extend each agent's planning horizon to two periods and show that our main findings are substantively unchanged by the longer planning horizon and gain insight into how experimentation varies within the life-cycle of individual agents.

Although the Brownian motion permits a particularly tractable representation, many of its theoretical virtues generalize to other stochastic processes. The appeal of the Brownian motion is that it enables a tight focus on risk aversion as a driver of experimentation and long-run dynamics. The Brownian motion presumes that drift and variance are constant (i.e., increments are independent and identically distributed and drift is linear). This implies that the marginal benefit as well as the marginal risk of experimentation in unexplored areas of the action space are independent of the history of past actions and performance.[1] High performing agents face exactly the same opportunities from further experimentation as do lower performing agents. Thus, any differences in behavior cannot be attributed to the environment and must be attributed to the relative risk aversion of agents at different performance levels.

This neutrality is not present in other stochastic processes. For instance, a geometric Brownian motion, or even a Brownian motion with a non-linear drift, imply that the marginal opportunity for experimentation varies in the performance level and/or the action undertaken. Thus, it becomes

[1] This same property is also present, although typically implicit, in models of R&D and growth, such as in the influential contribution of Harris and Vickers (1987).

necessary to disentangle the effects of time and performance on experimentation from the effect of risk aversion. These factors do, of course, in practice interact in an agent's decision to experiment. Our approach is intended to highlight most clearly the role of risk aversion in this decision. After presenting the formal results, we sketch the behavioral possibilities that might emerge for other stochastic processes.

1.2 Behavior in the Short and the Long Run

The Brownian motion captures the trade-off between risk and return intuitively and concisely. For a positive drift term, experimenting to the right of all known actions—what we refer to as the *knowledge frontier*—offers higher expected performance along with higher risk, with both increasing linearly in the size of the experimental step. This, combined with the unbounded action space, gives rise to a technical question: Does an optimal action exist? We first show that risk aversion is insufficient to guarantee existence. It is possible that a risk averse agent always prefers the incremental return of a larger experiment over the additional risk. Nevertheless, we prove that this is not the case for all risk averse agents and we identify a pair of sufficient conditions on risk aversion that ensure the existence of an optimal action for all performance histories. The first of these conditions, standard risk aversion (Kimball 1993), reveals an unexpected connection to the literature on precautionary savings. We show that the same condition that ensures precautionary saving in an intertemporal budgeting problem also ensures the existence of an optimal action in our setting when combined with a simple crossing condition on the level of absolute risk aversion.

The behavior generated by these conditions is intuitive. The lower is an agent's risk aversion the more he is willing to experiment and the more boldly he experiments. Thus, the higher is past performance—and the lower is the agent's level of absolute risk aversion—the greater is the experimental step.[2] This variation extends to one of type and not just degree. We show that for low enough performance, the risk of experimentation overwhelms the potential gain and an agent does not experiment at all, sticking to a known action. In this case learning stagnates and becomes caught in a *performance trap*. Critical to the existence of the performance trap is the crossing condition on the level of absolute risk aversion. Without it, all agents would prefer to experiment, regardless of performance level, and the performance trap would not exist. However, the absence of the crossing condition also implies that an optimal action fails to exist for any agent as they desire ever more risk. Thus, the two phenomena are codetermined: There cannot exist an optimal action without the creation of a performance trap.

[2] Standard risk aversion implies declining absolute risk aversion (DARA) in our setting.

Behavior becomes richer and more subtle after the first period and is no longer solely determined by the trade-off of risk and return at the knowledge frontier. As time passes and knowledge accumulates, agents must also contend with the history of past actions. We show how history can matter and identify the mechanism through which past performance *away* from the knowledge frontier can impact experimentation *at* the frontier. In the mechanism we identify, history may only constrain experimentation rather than encourage it and an effect only emerges when performance suffers a set-back at the knowledge frontier. The conditionality of this effect highlights the importance of including set-backs as well as breakthroughs in a model of innovation.

It is easiest to see the role of history when the outcome of an experiment falls below the threshold for a performance trap. This implies an agent does not want to experiment further to the right, but instead of remaining at the frontier he backtracks to the best performing past action. History matters in this case not so much whether to experiment *per se* but rather on which known action behavior settles. Yet this is only part of the mechanism. To see the full effect of history, consider instead an experimental outcome that falls slightly above the original performance trap. In this case an agent prefers to experiment rather than remain at the frontier. Yet the gains from experimentation are small, and if they are small enough, the agent will be tempted to abandon experimentation altogether and return to an earlier higher performing action. In this way the performance trap expands beyond its initial threshold, not trapping performance itself but rather driving agents back to the safety of past choices. The better is past performance relative to the frontier, the more tempting it is, and the more history acts as a constraint on experimentation.

This leads us to questions about experimentation and performance in the long-run. Our main result is to show that performance diverges in expectation over time. Therefore, a higher starting performance translates into higher expected performance in the future. Arrow (1962) observed many years ago that lower risk aversion leads to greater experimentation. Our result closes the causal loop on Arrow by showing that greater experimentation, in turn, leads to better performance, which further lowers risk aversion. This causal loop then feeds on itself. Good performance begets good performance, whereas bad performance leads to more bad performance. Initial differences grow and performance diverges over time.

Divergence holds only in expectation, however, and an early advantage may disappear with a single stroke of bad luck. In this case the implications of divergence are more pernicious. It implies that every turn of luck, be it good or bad, has a lasting impact on performance. Should a leader fall behind, divergence implies that it should expect to remain behind thereafter. In this way our model is simultaneously consistent with both the persistence of performance differences that is evident in

many applications, and the possibility that the order of performance can nevertheless be shaken up.

Divergence follows from behavior at the knowledge frontier and from the weight of history behind it. At the knowledge frontier the feedback loop that drives divergence is clear. At the frontier a complementarity develops between experimentation and performance. Better performance lowers risk aversion, which leads to bolder experimentation and better performance on average, which, in turn, further relaxes risk aversion and begins the cycle anew.

Away from the knowledge frontier the mechanism of divergence is more subtle. For learning to stop agents must experience a setback and be ensnared by the performance trap. As the performance trap increases in past performance, and because it is the better performers who experiment most boldly and engage the risk of every larger setbacks, one might think that it is the better performers who are more likely to be weighed down by history. We find, however, that this is exactly backwards. Although the performance trap grows with performance, we show that it grows at a slower rate, and that this rate is slow enough that the probability of getting ensnared in the performance trap is strictly decreasing in frontier performance. Thus, the weight of history bears greatest on those who perform poorly, exacerbating the gap in performance between them and higher performing agents.

The decreasing rate at which learning is caught in the performance trap leads to a final question: Does learning inevitably stop or may it continue forever? We show that the rate at which learning is caught by the performance trap does converge to zero. Moreover, we show that it converges sufficiently fast that, with strictly positive probability, learning escapes the performance trap and continues indefinitely. We explore this "escape probability" numerically and find that it exhibits a sharp two-phase pattern. In early periods the chance of being ensnared by the performance trap is potentially high, but, if learning survives, the probability of getting caught drops precipitously and remains at effectively zero thereafter. The path of learning, therefore, exhibits, a somewhat tumultuous early phase followed by a later relatively smooth but nonetheless stochastic growth path.

1.3 Relation to the Literature

Our use of the Brownian motion follows Callander (2011). That paper imposes quadratic utility and presumes the existence of an internal optimum (i.e., an ideal point). In contrast, we analyze the standard economic environment of unbounded and unsatiated utility and allow for a broad

class of preferences, yet show that the equilibrium is simpler.[3] Moreover, these differences matter for how agents experiment and how they perform. The presence of an ideal outcome means agents know not only how they are performing but whether they are above or below the ideal and they use this directional information in guiding experimentation. The ideal outcome also implies that learning inevitably stops when an outcome is realized near enough to zero and, by implication, that performance converges, the opposite of what we find here.

Garfagnini and Strulovici (2016) allow for unbounded utility but they set aside the question of risk aversion altogether, focusing exclusively on risk-neutral agents. Their interest is in characterizing behavior for agents with two period planning horizons. The longer horizon leads to an existence problem similar to the one we encounter here, even when the drift of the Brownian motion is zero. To work around this problem, Garfagnini and Strulovici impose an exogenous cost of experimentation and assume that the cost increases in the novelty of the experiment. We show how such a constraint emerges endogenously through risk aversion.[4] This difference in source drives a difference in long-run behavior. In Garfagnini and Strulovici (2016) learning ends with probability one, whereas here, with risk aversion constraining learning, the constraint naturally relaxes as performance grows and this allows for a strictly positive escape probability and for learning to continue forever. A further difference is in the generality of preferences considered. Whilst Garfagnini and Strulovici (2016) permit only a single utility function (risk neutrality), we allow for the broad class of standard risk averse preferences in our one-period horizon results. We restrict this class when we extend agents' planning horizon to two periods, although even then we allow for a family of linear-exponential utility functions.

The Brownian motion represents a bandit model with a continuum of correlated, deterministic arms. The correlation distinguishes us sharply from the classic bandit literature. It is well known that the analysis of correlated bandits is difficult. Yet it is correlation that captures the ability of agents to learn across alternatives, an ability that is important in practice. Correlation across actions is also necessary for a meaningful distinction between the knowledge frontier and the full body of accumulated knowledge. This distinction is essential for our results and is what separates our model from reduced-form models of R&D and growth in the IO and macro literatures, respectively. This richness in the decision problem requires a trade-off analytically, limiting our analysis

[3] As is well known, quadratic utility is not formally compatible with this environment. Additionally, quadratic utility exhibits increasing absolute risk aversion, a property at odds with empirical evidence.

[4] Another difference between the models is that Garfagnini and Strulovici (2015) apply the cost of experimentation *only* to actions on the flanks of knowledge and not those between previously chosen actions, whereas in our model the same utility function applies everywhere and the impact of risk aversion is felt in between known actions as well as on the flanks.

to a decision theoretic problem and limited planning horizons. Extending the model to allow for longer horizons and competitive markets for innovation offer the promise for further insight.

2 The Model

A countable set of agents take actions sequentially, one in each period $t = 1, 2, ...$ The action of the agent in period t is $a_t \in \mathbb{R}$ and generates outcome $m_t \in \mathbb{R}$. The agents are otherwise identical and their utility depends only on their own action. Our aim is to characterize the agents' optimal actions given the technology, preferences, and information structure we describe next.

Technology: The production function that maps action a_t into outcome, or performance, m_t is given by $m(a_t)$. We model the production function $m(a_t)$ as the realized path of a Brownian motion with drift $\mu > 0$ and variance $\sigma^2 > 0$. We refer to $a_0 = 0$ as the status quo action and denote its outcome by $m_0 = m(a_0)$. The realized path of the Brownian motion is determined by nature before the start of the game and does not change over time. Figure 1 depicts one possible realization of the Brownian motion.

Preferences: The utility function that maps outcome $m_t \in \mathbb{R}$ into utility $u \in \mathbb{R}$ is $u(m_t)$. The utility function is four times continuously differentiable and satisfies $u'(m_t) > 0$ and $u''(m_t) < 0$ for all $m_t \in \mathbb{R}$. The first condition implies non-satiation and the second risk aversion.

We further assume that the utility function satisfies *standard risk aversion* (Kimball 1993). This requires that a risk cannot be made more desirable by the imposition of an independent, loss-aggravating risk, which implies that independent risks are substitutes.[5] Standard risk aversion is satisfied by most commonly used utility functions, including exponential, logarithmic, and power functions.

Standard risk aversion is a stronger notion than decreasing absolute risk aversion as, in settings such as ours, the former implies the latter but not the reverse. It is necessary to assume the more restrictive notion as agents in our setting can control the degree of risk that they engage, and thus must compare risky alternatives to other risky alternatives. A lesson from the literature on decision making under uncertainty is that, for such comparisons, absolute risk aversion is insufficient to fully

[5] Formally, suppose there are two independent random variables x and y such that

$$\mathrm{E}\left[u(m_t + x)\right] - u(m_t) \leq 0 \quad \text{and} \quad \mathrm{E}\left[u'(m_t + y)\right] - u'(m_t) \geq 0.$$

The risk x is undesirable since the agent would turn it down if someone offered it to him. And the risk y is loss aggravating since it increases the agent's expected marginal utility; it therefore makes a sure loss more undesirable and is itself more undesirable if the agent is exposed to a sure loss. The utility function then satisfies standard risk aversion if and only if

$$\mathrm{E}\left[u(m_t + x + y) - u(m_t + y)\right] \leq \mathrm{E}\left[u(m_t + x) - u(m_t)\right],$$

that is, if and only if the loss aggravating risk y does not make the undesirable risk x more desirable.

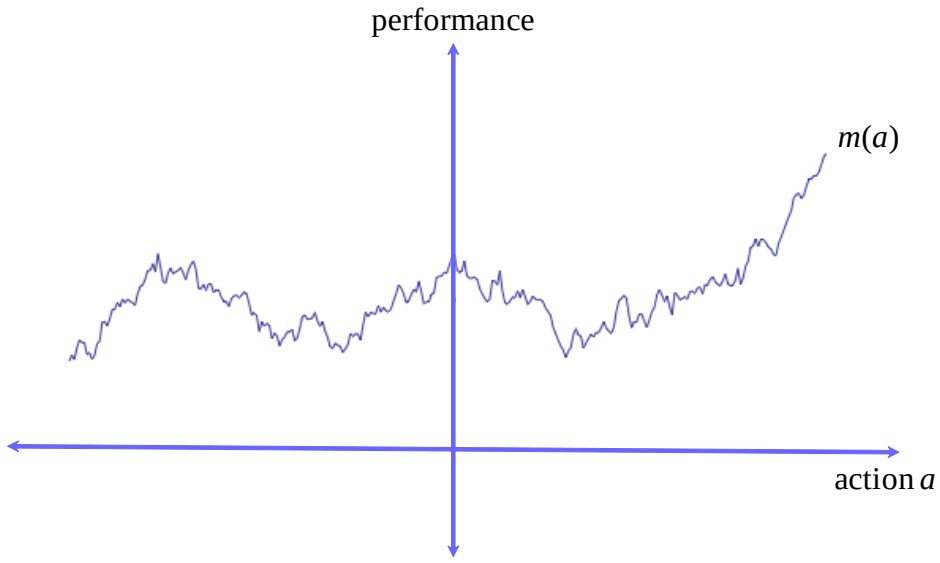


Figure 1: The production function m(a)

characterize behavior, being designed instead for the comparison of a risky action to a safe action (see, for instance, Chapter 9 in Gollier (2001)).

Nevertheless, standard risk aversion alone is insufficient to ensure well defined choice in our setting. We require an additional 'crossing condition' on absolute risk aversion. Denoting the coefficient of absolute risk aversion by $r(m_t) = -u''(m_t)/u'(m_t)$, such that decreasing absolute risk aversion implies $r'(m_t) \leq 0$ for all $m_t \in \mathbb{R}$, we assume that $r(m_t)$ crosses the value $2\mu/\sigma^2$, where μ and σ^2 are the drift and the variance of the Brownian motion. An example of a standard utility function that satisfies our conditions is given by $u(m_t) = \alpha m_t - \exp(-\beta m_t)$, where $\alpha > 0$ and $\beta > 2\mu/\sigma^2$.

Information: The agents do not know the realization of the Brownian motion. They do, however, know that the production function was generated by a Brownian motion with drift $\mu > 0$ and variance σ^2. And they know that the status quo action a_0 generates status quo action m_0. In addition, agents know all the actions that their predecessors took and the outcomes that these actions generated. The information set in period t is therefore given by $I_t = \{\mu, \sigma^2, (a_0, m_0), ... (a_{t-1}, m_{t-1})\}$.

Several points in the information set are of particular importance. We refer to the right-most known action as the "knowledge frontier" and the corresponding outcome as "frontier performance." As performance at the frontier may be a breakthrough or a setback, we reserve the distinct notion of "peak performance" for the maximum outcome attained by any action in I_t.

Optimal Learning Rule: In any period t, the agent chooses the action that maximizes his

expected utility. An optimal learning rule is therefore given by $(a_1^*, a_2^*, ...)$, where

$$a_t^* \in \arg\max_{a_t} \mathrm{E}\left[u\left(m_t\right)|I_t\right].$$

Our goal is to characterize the set of optimal learning rules.

3 Beliefs and Expected Utility

We start by examining agents' beliefs about outcomes over untried actions. Consider an agent in some period t and let l_t and h_t denote the left-most and right-most known actions. From the properties of the Brownian motion, beliefs for all untried actions are distributed normally. For an action a_t to the right of h_t, the mean is given by

$$\mathrm{E}\left[m\left(a_t\right)\right] = m\left(h_t\right) + \mu(a_t - h_t) \tag{1}$$

and variance

$$\mathrm{Var}\left(m\left(a_t\right)\right) = \left(a_t - h_t\right)\sigma^2. \tag{2}$$

The agent expects an action to generate a higher outcome the further it is to the right of h_t. At the same time, the further an action is to the right of h_t, the more uncertain the agent is about the outcome generated by that action. Thus, the more novel an experiment is, the more uncertain is the outcome. The value of σ^2 parameterizes how uncertain beliefs are and we interpret σ^2 as the complexity of the production process. Beliefs for actions to the left of the left-most action l_t are analogous.

That beliefs to the right of action h_t depend only on h_t is due to the Markov property of the Brownian motion. The Markov property also applies between known actions, although now beliefs depend on the closest known action in either direction. To avoid the need for more notation, suppose that there are no known actions between l_t and h_t. For any action $a_t \in [l_t, h_t]$, outcome $m(a_t)$ is then normally distributed with mean

$$\mathrm{E}\left[m\left(a_t\right)\right] = \frac{a_t - l_t}{h_t - l_t} m\left(h_t\right) + \frac{h_t - a_t}{h_t - l_t} m\left(l_t\right) \tag{3}$$

and variance

$$\mathrm{Var}\left(m\left(a_t\right)\right) = \frac{\left(a_t - l_t\right)\left(h_t - a_t\right)}{h_t - l_t}\sigma^2. \tag{4}$$

The agent's expected outcome is a convex combination of the outcome generated by l_t and h_t. Moreover, the further the action is from the closest known action, the more uncertain the agent is about the outcome generated by that action.

The Brownian motion formulation ensures that the agents' beliefs take a simple form that satisfies several intuitive and appealing properties. First, beliefs are normally distributed. This provides the advantage of tractability as it permits a simple mean-variance representation for expected utility, even within a large class of utility functions. Second, agents face a familiar risk-return trade-off in experimentation, one that is independent of the state of knowledge and, thus, one in which behavior is driven by risk aversion. Third, agents know more about an action the closer the action is to a known action and the less complex is the production process. Fourth, agents learn across alternatives as each new point that is observed reveals information about not only that action but of all neighboring points. Fifth, this leads learning to be directed: agents know where they can expect better actions and they focus their experimentation in that direction. Sixth, and finally, the model captures the reality that theoretical reasoning can only go so far as regardless of how many (finite) points are observed in the mapping, agents never can infer the full mapping and there remains an essential need to learn by trial and error.

For an agent with beliefs normally distributed of mean M and variance V, let z denote a random variable that is drawn from a standard normal distribution. The following lemma establishes the conditions necessary for expected utility to be representable in mean-variance space. The lemma is proven in Theorem 1 of Chipman (1973).

LEMMA 1 (Chipman 1973). *Suppose that* $|u(m)| \leq A\exp(Bm^2)$ *for some* $A > 0$ *and* $B > 0$. *Then the expected utility function*

$$W(M,V) = \mathrm{E}\left[u\left(M + \sqrt{V}z\right)\right]$$

exists for all $M \in (-\infty, \infty)$ *and* $V \in (0, 1/(2B))$.

Formally, the restriction in the lemma ensures that expected utility is integrable, and for the remainder of the paper we assume that it holds. Notice that since we are free to choose any positive parameters A and B, this restriction is mild.

4 Risk and Trial-and-Error Learning

We begin by characterizing the optimal action in the first period before turning to the second and subsequent periods. We then explore the implications of optimal learning for long-term performance.

4.1 The First Period

The first agent has the choice to play it safe or to experiment. The safe action is the status quo that realizes outcome m_0 with certainty. Any other action, $a_1 \neq a_0$, is an experiment as the outcome is unknown. Actions to the left of the status quo can be ruled out as their expected outcomes are below the status quo. The agent therefore experiments to the right or does not experiment at all.

Proposition 1 establishes when the agent experiments and how boldly he does so. It shows that the first agent's behavior satisfies a simple threshold rule, where the degree of experimentation, and whether to experiment at all, are determined by the performance level at the status quo. Let $\Delta_1 = a_1 - a_0 \geq 0$ denote the size of the step the agent takes in the first period, and let $\widehat{m}$ denote the largest outcome for which the coefficient of absolute risk aversion satisfies, $r(m_t) = 2\mu/\sigma^2$.

PROPOSITION 1. *The first agent's optimal action is unique and given by*

$$a_1^* = \begin{cases} a_0 + \Delta(m_0) & \text{if } m_0 \geq \widehat{m} \\ a_0 & \text{if } m_0 < \widehat{m}, \end{cases}$$

where $\Delta(m_0)$ *is implicitly defined by*

$$R(m_0, \Delta(m_0)) \equiv -\frac{\mathrm{E}\left[u''\left(m_0 + \mu\Delta_1 + \sqrt{\Delta_1}\sigma z\right)\right]}{\mathrm{E}\left[u'\left(m_0 + \mu\Delta_1 + \sqrt{\Delta_1}\sigma z\right)\right]} = 2\mu/\sigma^2$$

and satisfies $\Delta(m_0) = 0$ *if* $m_0 = \widehat{m}$ *and* $\Delta(m_0) > 0$ *if* $m_0 > \widehat{m}$. *The optimal step size* $\Delta(m_0)$ *is increasing without bound in status quo outcome* m_0 *and the drift* μ *and decreasing in the complexity of the production function* σ^2 *and the agent's risk aversion. The threshold* $\widehat{m}$ *is decreasing in* μ *and increasing in* σ^2 *and the agent's risk aversion.*

The threshold $\widehat{m}$ separates agents who experiment and those who play it safe. A status quo outcome below the threshold leads the agent to play it safe by settling for the status quo action. An agent experiments for a status quo outcome above $\widehat{m}$, where the boldness of his experiment increases in the performance level of the status quo outcome.

The non-experimentation of low performing agents represents a performance trap. Low performing agents become trapped not because they have fewer opportunities or lack access to a technology such as capital. Every agent faces exactly the same risk-return trade-off regardless of the status quo outcome and regardless of how much risk he has already engaged. The reason that a low performing agent forgoes the potential gains from experimentation is because of risk aversion. Low performing agents have lower income and face higher effective aversion to risk. For them, consequently, the risks of experimentation are more daunting and the down-side risks more costly.

The fact that the performance trap smothers experimentation completely contrasts with behavior in the related and well-known portfolio problem. The portfolio problem is similar to ours in that agents trade-off risk and return along a linear function. In that problem, however, the trade-off is linear in return and the standard deviation of risk. As a result, investors in that setting always find it optimal to put some income into the risky asset no matter the level of their risk aversion (see, for instance, Chapter 4 in Gollier (2001)). In our setting, in contrast, the linear trade-off is between return and variance, and this drives the lower performing agents away from experimentation altogether.

A status quo outcome above the threshold $\widehat{m}$ induces an agent to experiment. The size of the experiment—how much it departs from what is known—is increasing without bound in the performance level of the status quo outcome. A high performing status quo relaxes effective risk aversion and the downside risks are no longer so daunting. The knowledge of a good action, therefore, empowers the agent to experiment and the better his knowledge is, the more empowered he is and the larger is the size of the experimental step he takes.

The behavior of the first agent is intuitive and simple. Yet it is not immediate and it does not emerge from risk aversion alone. To see why, observe that each increment in the size of an experiment is equivalent to adding an additional independent risk. Moreover, each increment increases the expected outcome of the experiment, making the agent wealthier in expectation. As absolute risk aversion is decreasing, this raises the following question: If an agent finds it appealing to add one increment of risk, why does he not find it optimal to add a second identical increment when he has higher outcome in expectation, and a third increment, and so on ad infinitum? To put it more concisely, why is it that an agent is inevitably satiated by risk such that a finite optimal action exists?

The answer to this question is the combination of standard risk aversion and the crossing condition on absolute risk aversion. With these two conditions holding, an agent eventually is satiated by risk, regardless of the performance level at the status quo outcome, and a finite optimal action always exists. It turns out that the same two conditions are sufficient to guarantee the emergence of the performance trap. Thus, these two properties of behavior are inextricably intertwined. The nature of this interdependence emerges from the derivation of optimal behavior, which we now develop in some detail.

For an experiment of size Δ_1, we have from (1) and (2) that the expected outcome is $m_0 + \mu\Delta_1$ and the variance $\sigma^2\Delta_1$. Utilizing the mean-variance representation of Lemma 1, the first agent's

problem can be written:

$$\max_{\Delta_1 \geq 0} W\left(m_0 + \mu\Delta_1, \sigma^2\Delta_1\right).$$

The shape of the indirect utility function, $W(.)$, depends on how additional increments of risk affect utility. It is easy to show that risk aversion alone is insufficient to pin down the shape of $W(.)$. For $W(.)$ to be concave, it is necessary that each additional increment of risk is less attractive than the increment before. That is, it is necessary that the increments of risk are substitutes and not complements. Lemma 2 establishes that this is exactly the property guaranteed by standard risk aversion. Kimball (1993), in introducing standard risk aversion, showed that it ensures individuals save in the face of income uncertainty when they have standard risk aversion. We have shown that the same underlying force ensures that the marginal benefit of risk is decreasing in the distinct setting of experimentation and innovation.

LEMMA 2. *The expected utility function* $W\left(m_0 + \mu\Delta_1, \sigma^2\Delta_1\right)$ *is concave in* Δ_1.

Concavity of $W(.)$ implies that the first increment of risk is the most valuable, regardless of the performance level of the status quo outcome, and that the marginal benefit decreases as more and more risk is engaged. Concavity alone is insufficient to ensure existence of an optimal action as it does not guarantee that the marginal benefit ever reaches zero. Concavity does imply, however, that the question of existence reduces to whether the first derivative of expected utility ever crosses zero. To that end, differentiate $W(\cdot)$ to obtain

$$\frac{\mathrm{d}W\left(m_0 + \mu\Delta_1, \sigma^2\Delta_1\right)}{\mathrm{d}\Delta_1} = \mathrm{E}\left[u'\left(m_0 + \mu\Delta_1 + \sigma\sqrt{\Delta_1}z\right)\right]\frac{\sigma^2}{2}\left(\frac{2\mu}{\sigma^2} - R\left(m_0, \Delta_1\right)\right), \tag{5}$$

where

$$R\left(m_0, \Delta_1\right) \equiv -\frac{\mathrm{E}\left[u''\left(m_0 + \mu\Delta_1 + \sqrt{\Delta_1}\sigma z\right)\right]}{\mathrm{E}\left[u'\left(m_0 + \mu\Delta_1 + \sqrt{\Delta_1}\sigma z\right)\right]}. \tag{6}$$

and where we make use of the fact that $\mathrm{E}\left[u'(\cdot)\, z\right] = \sigma\sqrt{\Delta_1}\mathrm{E}\left[u''(\cdot)\right]$.

The sign of expected marginal utility is determined by the final term in (5). This compares the relative size of the ratio $2\mu/\sigma^2$and $R\left(m_0, \Delta_1\right)$. The ratio μ/σ^2 is a measure of the risk adjusted return from experimentation, summarizing the production side of the model, and $R\left(m_0, \Delta_1\right)$ is the coefficient of absolute risk aversion for the expected utility function $\mathrm{E}\left[u(\cdot)\right]$, summarizing the agent's preferences. Note that the right hand side expression is different from the expected coefficient of absolute risk aversion $-\mathrm{E}\left[u''(\cdot)/u'(\cdot)\right]$, except when $\Delta_1 = 0$. In that case both expressions are equivalent to the coefficient of absolute risk aversion; i.e., $R\left(m_0, 0\right) = r\left(m_0\right)$.

This last fact implies that the choice whether to experiment or not is determined by the coefficient of absolute risk aversion. This, combined with our crossing condition, implies:

$$\frac{\mathrm{d}W\left(m_0,0\right)}{\mathrm{d}\Delta_1}\begin{cases} >0 & \textit{if } m_0>\widehat{m} \\ \leq 0 & \textit{if } m_0\leq\widehat{m}. \end{cases} \tag{7}$$

The crossing condition requires that $r\left(m_0\right)$ crosses the value $2\mu/\sigma^2$ with $\widehat{m}$ denoting the largest outcome m for which $r(m)=2\mu/\sigma^2$. Thus, the crossing condition, combined with standard risk aversion, creates the performance trap. Agents with a status quo outcome below the threshold $\widehat{m}$ find the first increment of risk unpalatable, and concavity of expected utility ensures all subsequent increments are unpalatable as well.

The crossing condition also ensures that the performance trap applies only to lower performing agents. If $r(m)<2\mu/\sigma^2$ for all values of m then no agent would experiment. Instead, the crossing condition guarantees that the first increment of risk is attractive for agents who face high status quo outcomes. Less immediate, it is the same crossing condition that ensures that marginal utility eventually turns negative and that a finite optimum exists. If $r(m)<2\mu/\sigma^2$ for all values of m then not only does a performance trap not exist, but for no agent does a finite optimal action exist. The necessity of the crossing condition for the existence of an optimal action is stated in Proposition 2, which also captures the knife-edge case in which $r(m)=2\mu/\sigma^2$ for all values of m.

PROPOSITION 2. *If the coefficient of absolute risk aversion $r\left(m\right)$ satisfies $r\left(m\right)>2\mu/\sigma^2$ for all $m\in\mathbb{R}$, the first agent takes the status quo action, and so do all his successors. If, instead, $r\left(m\right)<2\mu/\sigma^2$ for all $m\in\mathbb{R}$, an optimal action does not exist. Finally, if $r\left(m\right)=2\mu/\sigma^2$ for all $m\in\mathbb{R}$, then in any period t the agent is indifferent between the right-most known action and any action to its right.*

Without the coefficient of absolute risk aversion ever dropping below the $2\mu/\sigma^2$ threshold, agents never tire of risk. Although each increment of risk is less and less satisfying—expected utility is still concave—they are never undesirable and an optimal action fails to exist.

4.2 The Second and Subsequent Periods

The second agent differs from the first only in that he observes the first agent's action and outcome before he must act. This advantage is immaterial if $m_0\leq\widehat{m}$ as then the first agent is caught in the performance trap and his action reveals no new information. In that case, the second agent faces the same decision problem as the first and he too chooses the safety of the status quo. This logic

recurs indefinitely and the hold of the performance trap is permanent. Once experimentation is abandoned it never restarts and low performing agents are forever condemned to low performance.

We presume hereafter that $m_0 > \widehat{m}$ and that the first agent experiments, revealing a second point in the production function. This additional information distinguishes the second agent's problem from that of the first agent. The new knowledge does not shed light on actions to the left of a_0 and these actions remain dominated by the status quo. For actions between a_0 and a_1^* beliefs are updated according to (3) and (4), and for actions to the right of a_1^* beliefs are given by (1) and (2), although now anchored at a_1^* rather than a_0.

For actions to the right of a_1^*, the second agent's problem is analogous to the first agent's problem. He can pursue additional increments of experimentation, and his expected utility will yield a finite optimum in that direction, just as it did for the first agent. It is tempting, therefore, to conjecture that the second agent's problem is a recurrence of the first, albeit with a different starting point. Yet this is not the case. The additional information available to the second agent provides him with an additional and potentially valuable option.

The additional option available to the second agent comes from experience. In addition to repeating the frontier action or experimenting to the right of there, the second agent has the option of exploiting actions away from the frontier and between known actions. He will not do this if peak performance is at the frontier as then the frontier dominates all other actions. If, however, knowledge at the frontier suffers a setback and peak performance resides with the original status quo, the agent may prefer to backtrack and choose an action away from the frontier.

This leads to two questions: When does the second agent backtrack away from the knowledge frontier? And which action does he choose when he does so? The second question is straightforward and we deal with it first: When the agent backtracks he chooses the action that delivers peak performance. All other actions behind the knowledge frontier are dominated as they yield lower expected outcomes and (weakly) greater risk.

The question of when the agent backtracks is more subtle. If the outcome of the first agent's action is below $\widehat{m}$ this leaves frontier knowledge within the region of the performance trap. The second agent, prefers to repeat the first agent's action rather than experiment further, and, thus, strictly prefers to backtrack to the higher status quo outcome. Thus, the performance trap is not exclusively a starting phenomenon. It is instead a frontier phenomenon. Experimentation is abandoned and learning stops whenever frontier knowledge falls within its grasp, regardless of what was previously possible.

Indeed, not only does the performance trap persists, it grows over time. For frontier knowledge

above the threshold $\widehat{m}$, the second agent prefer to experiment rather than repeat the first agent's action. Yet experimentation itself may be dominated by backtracking to the surety of peak performance. In this case the second agent would experiment if frontier knowledge were all that he knew, but, in possession of the full body of knowledge, he instead chooses to abandon experimentation. Lemma 3 characterizes the critical value of the first period outcome at which the second agent is indifferent between continuing to experiment and backtracking to the status quo.

LEMMA 3. *There exists a threshold outcome* $\widetilde{m}(m_0) \in (\widehat{m},\, m_0)$ *such that*

$$u\left(m_0\right) = W\left(\widetilde{m}(m_0) + \mu\Delta\left(\widetilde{m}(m_0)\right), \sigma^2\Delta\left(\widetilde{m}(m_0)\right)\right), \tag{8}$$

where $\Delta\left(\widetilde{m}(m_0)\right) > 0$. *The derivative of the threshold satisfies* $0 < \widetilde{m}'(m_0) \leq 1$. *Moreover, the threshold is increasing in* σ^2 *and the agents' risk aversion and decreasing in* μ.

Behavior in the second period follows, therefore, a threshold rule, albeit a different threshold to the first agent and with a twist to the logic. The second agent experiments when first period performance is above the threshold $\widetilde{m}(m_0)$, and when he does so, he follows exactly the prescriptions of Proposition 1. Below the threshold, however, the second agent abandons experimentation and reverts to a previously chosen action. As this reveals no new information, the third agent inherits exactly the same problem and he too seeks the safety of a known action. Thus, it takes only a single sufficiently poor outcome of the first experiment to bring experimentation to a permanent end.

Two properties of interest emerge from this threshold rule. First, the second agent never repeats the action of the first agent. He either continues to experiment or he reverts to the status quo action. Thus, experimentation cannot stop at the frontier itself. Second, the accumulation of knowledge over time places a lower bound on the incrementalism of experimentation. As the new threshold $\widetilde{m}(m_0)$ is strictly higher than $\widehat{m}$, and the size of an agent's optimal experiment is increasing in m, the second agent prefers to either experiment a non-trivial amount or to backtrack to the status quo. Therefore, the performance level of the status quo outcome, and peak performance more generally, bound the size of all future experimentation.

The logic for the second agent extends naturally to all subsequent agents. An agent either experiments to the right or back-tracks, and when he back-tracks he reverts to a known action. The only difference with the second agent is that a later agent reverts to the best previous action and not necessarily to the status quo action. Let $\overline{m}_t$ denote the highest known outcome in period t, that is, let

$$\overline{m}_t = \max\{m_0, m_1^*, m_2^*, ..., m_{t-1}^*\}.$$

Also, let $\overline{a}_t$ denote the action that generates $\overline{m}_t$, that is,

$$\overline{a}_t \in \{a_0, a_1^*, a_2^*, ..., a_{t-1}^*\} \text{ such that } m(\overline{a}_t) = \overline{m}_t.$$

And finally, recall that h_t denotes the right-most known action in period t. The optimal action for any agent in period $t \geq 2$ is then given as follows.

PROPOSITION 3. *The agents' optimal action in period $t \geq 2$ is unique and given by*

$$a_t^* = \begin{cases} h_t + \Delta(m(h_t)) & \text{if } m(h_t) > \widetilde{m}(\overline{m}_{t-1}) \\ a(\overline{m}_{t-1}) & \text{if } m(h_t) \leq \widetilde{m}(\overline{m}_{t-1}), \end{cases}$$

where $\Delta(m) > 0$ is the Δ that solves $R(m, \Delta) = 2\mu/\sigma^2$ and $\widetilde{m}(\cdot)$ is defined in (8).

The optimal action in any period depends on only two factors: frontier performance $m(h_t)$ and peak performance $m(\overline{a}_t)$. If frontier performance falls sufficiently short of peak performance, the agent reverts to the best known action. If, instead, frontier performance is sufficiently good relative to peak performance, the agent continues to experiment and he experiments more boldly the better frontier performance is. As the level of peak performance can only (weakly) increase, the performance threshold at which experimentation stops, in turn, can only increase.

4.3 Long-Run Performance

Propositions 1 and 3 characterize behavior *within* each period. In this section we turn to the effects of risk *across* periods. Our main result is to show that the effects of risk are cumulative and complementary. That the more early agents engage in experimentation, the more likely later agents do as well, and the better off they are. Formally, we establish that performance across agents diverge in expectation over time. This implies that initial differences in performance matter, that they persist and grow in expectation, and that they are determinative of long-run performance.

We state this result formally in Proposition 4. To more easily describe performance over time, we refer to a sequence of actions as representing a *field* of knowledge.

PROPOSITION 4. *Consider two fields of knowledge, H and L. The production function of field $k = H, L$, is characterized by status quo outcome m_0^k, drift μ, and variance σ^2, with $m_0^H > m_0^L > \widehat{m}$. Then*

$$\mathrm{E}_1\left[m_t^*(m_0^H) - m_t^*(m_0^L)\right] > m_0^H - m_0^L \text{ for all } t = 1, 2, ..., \tag{9}$$

where $\mathrm{E}_1[\cdot]$ are the expectations taken at the beginning of the first period.

Divergence in performance implies a strong path dependence to trial-and-error learning. It establishes not only that long-term performance depends on initial conditions, but that it depends

on initial conditions in a systematic way. The result holds only probabilistically, however, and it is nevertheless possible for an early leader to get a stroke of misfortune and fall behind. In this case the divergence result implies that this stroke of bad luck has long term ramifications as that field of knowledge is thereafter expected to remain behind. We explore the likelihood of this possibility numerically in the following section.

Performance divergence follows from behavior at the knowledge frontier and from the weight of history behind it. At the knowledge frontier the mechanism that drives divergence is clear. A complementarity develops between experimentation and performance. Better performance lowers risk aversion, which leads to bolder experimentation and better performance on average, which, in turn, further relaxes risk aversion and begins the cycle anew.

Away from the knowledge frontier the mechanism of divergence is more subtle. Experimentation is bolder in better performing fields and, thus, riskier, with greater chance of a significant setback. It may seem, therefore, that it is in these fields that agents are more likely to be tempted by past actions when peak performance no longer resides at the frontier. To understand why this isn't the case, observe that the trade-off in backtracking to a known action involves not only the quality of the safe action but also the potential benefit of experimenting further. Precisely because experimentation is timid in poorly performing fields, the gains from experimentation are small, and the amount agents are willing to give up to experiment (from a lower frontier point) is even smaller. For poorly performing fields, therefore, the undulations in performance may be smaller in an absolute sense, but in a relative sense they are ever more daunting. We establish this formally in the proof of Proposition 4, showing that experimentation will be abandoned with greater frequency the worse performance is in a field.

The decreasing rate at which learning is caught in the performance trap leads to the question of whether getting caught is, in fact, inevitable. Proposition 5 establishes that it is not.

PROPOSITION 5. *For $m_0 > \widehat{m}$, the probability that learning stops in finite time is strictly less than 1.*

Thus, with positive probability, innovation *escapes* the performance trap and learning continues indefinitely. This result requires that the probability of being caught in the performance trap converges on zero as performance increases and that it converges on zero sufficiently fast. This implies for low performing fields that learning is fragile. Innovation arrives at a slow rate and even small setbacks run the risk of putting a stop to growth altogether. If this can be navigated successfully, however, prospects improve on two fronts: experimentation is bolder and growth accelerates while at the same time the prospect of a setback derailing innovation fades away. In the following section

we explore this transition numerically and find that the transition is sharp.

4.4 Example and Numerical Simulations

To illustrate the optimal learning rule, and further explore the dynamics of knowledge discovery, let the utility function be given by

$$u\left(m_t\right) = \alpha m_t - \exp(-\beta m_t), \tag{10}$$

where $\alpha > 0$ and $\beta > 2\mu/\sigma^2$. It can be readily verified that this utility satisfies standard risk aversion and the crossing condition on absolute risk aversion. This class of utility functions is known as the linex class (as in, linear-exponential) and has been extensively analyzed as it possesses many attractive properties. Bell (1988) shows that it, along with the sum of two exponential utility functions (sumex), are the only utility classes that increase in wealth, exhibit decreasing absolute risk aversion, and satisfy an intuitive one-switch condition (such that preference over two gambles switches at most once as wealth increases). Bell (1988) provides two additional simple and intuitive conditions and shows that only linex utility satisfies either one. Bell and Fishburn (2001) provide a strengthening of the one-switch condition that also rules out sumex utility. Finally, and pertinently, the linex class of utility functions permit simple closed form expressions for the thresholds that characterize optimal behavior in our model.

The initial threshold for the performance trap is given by

$$\widehat{m} = -\frac{1}{\beta}\ln\left[\frac{2\alpha\mu}{\sigma^2\beta\left(\beta - 2\frac{\mu}{\sigma^2}\right)}\right] \tag{11}$$

and the ongoing threshold as experience grows is

$$\widetilde{m}\left(\overline{m}_{t-1}\right) = \widehat{m} + \frac{\beta - \frac{2\mu}{\sigma^2}}{\alpha\beta}\left(u\left(\overline{m}_{t-1}\right) - u(\widehat{m})\right).$$

When the agent does experiment, his optimal action is given by

$$h_t + 2\mu\frac{m\left(h_t\right) - \widehat{m}}{\sigma^2\left(\beta - \frac{2\mu}{\sigma^2}\right)}$$

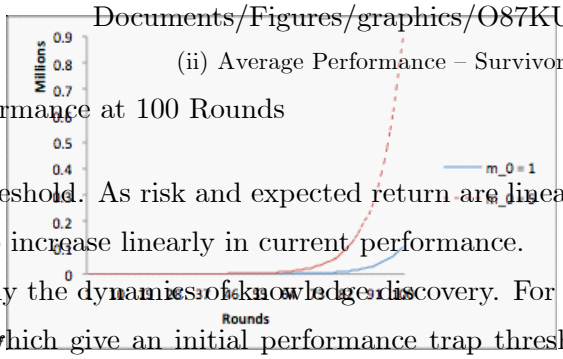
and his expected utility is

$$u(\widehat{m}) + \frac{\alpha\beta}{\beta - \frac{2\mu}{\sigma^2}}\left(m\left(h_t\right) - \widehat{m}\right).$$

As the parameters α, β, μ, and σ^2, are exogenous, an appealing property of this class of utility functions is that the optimal size of an experiment is linear in $\left(m\left(h_t\right) - \widehat{m}\right)$, the distance between

Documents/Figures/graphics/O87KU100

(i) Average Performance



Documents/Figures/graphics/O87KU101

(ii) Average Performance – Survivors

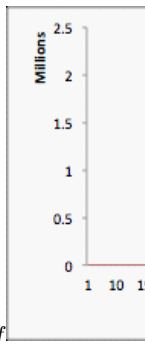


Figure 2: Average Performance at 100 Rounds

current performance and the initial knowledge threshold. As risk and expected return are linear in actions, this implies that both of these values also increase linearly in current performance.

We use these expressions to explore numerically the dynamics of knowledge discovery. For this exercise we set $\alpha = \beta = \sigma^2 = 1$ and $\mu = 1/5$, which give an initial performance trap threshold of $\widehat{m} = -\ln\left[\frac{2}{3}\right] = 0.41$. We calculate and contrast behavior for different values of the status quo outcome m_0. Specifically, we consider the two cases in which $m_0 = 1$ and $m_0 = 5$. For each case, we generated 1000 independent runs and recorded the agents' optimal actions and outcomes for the first 100 rounds.

The long run average performance is depicted in the left panel of Figure 2. The divergence of performance from the two starting outcomes is clearly evident. The initial advantage of a higher status quo outcome persists over time and grows in absolute terms.

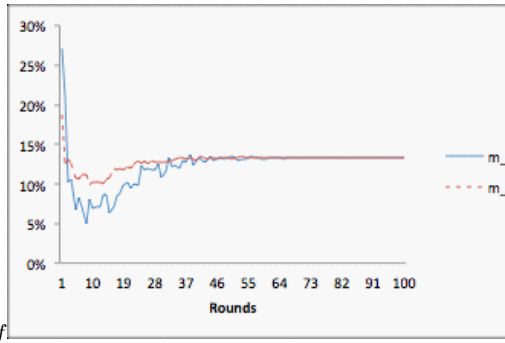
To delve deeper into divergence, Figure 3 indicates the growth rate of average performance over time. Growth rates are volatile early in the learning process before settling down and by round 50 they are remarkably stable. The early volatility is higher, and growth rates initially lower, for the lower status quo outcome. Ultimately, however, the growth rates converge to exactly the same level, driven by the linearity of equilibrium behavior in the linex utility class. With the same growth rate, the average performance of the high performance field stabilizes as a constant multiple of the low performance field. In our simulations this multiply is approximately nine times. Taking the performance trap threshold as the benchmark, an initial knowledge advantage of 7.7 times expands in the long-run to a 9 times advantage.[6]

The simulations also provide insight into the reasons for performance divergence and the chan-

[6] The initial advantage is calculated as: $\frac{5-(-\ln(2/3))}{1-(-\ln(2/3))} \approx 7.7$.

Documents/Figures/graphics/O87DDK4T

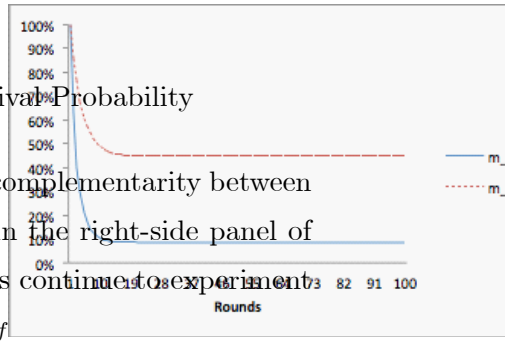
Figure 3: Growth Rate of Average Performance



3.*pdf*

Documents/Figures/graphics/O87DDK4X

Figure 4: Probability Learning Continues at 100 Rounds – Survival Probability

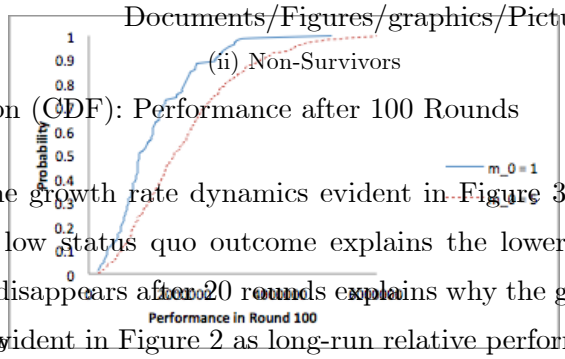


nels through which it operates. The first channel driving divergence—the complementarity between performance and experimentation at the knowledge frontier—is evident in the right-side panel of Figure 2. This figure restricts attention to simulation runs in which agents continue to experiment and shows that even within this set average performance is diverging. 4.*pdf*

The second channel driving divergence—the performance trap—can be seen in Figure 4. The figure shows the fraction of times that agents are still experimenting and learning after each round. This rate drops precipitously early on, yet it stabilizes by round 20. This stability represents the "escape probability" for innovation. As is evident, this probability can be high, hitting approximately 50% for the starting performance of $m_0 = 5$. The escape probability is considerably lower for $m_0 = 1$, reflecting the heightened risk of even a small setback derailing learning at low performance levels. This difference in survival complements the advantage at the knowledge frontier that a high status quo outcome carries and together they drive performance divergence.

Documents/Figures/graphics/Picture1

(i) Survivors



Documents/Figures/graphics/Picture2

(ii) Non-Survivors

Figure 5: Cumulative Density Function (CDF): Performance after 100 Rounds

The patterns in Figure 4 also rationalize the growth rate dynamics evident in Figure 3. The higher probability that learning stops for the low status quo outcome explains the lower early growth rates, and the fact that this probability disappears after 20 rounds explains why the growth rates stabilize and equalize. This effect is also evident in Figure 2 as long-run relative performance is much closer across the two cases when conditioning on survival.

An implication of the simulations is that knowledge discovery proceeds through two relatively distinct stages. During the early stage, a field is susceptible to getting caught in a performance trap. If this early stage can be survived, the risk of getting caught dissipates quickly. Knowledge then progresses indefinitely with each agent building on the knowledge discovered by his predecessors. It seems that learning either stops early or it doesn't stop at all.

The simulations also illuminate the substantial heterogeneity in performance over time. Although performance diverges on average across high and low status quo outcomes, there is substantial variation within each case. A field that starts with a low status quo can perform well, with innovation becoming self-sustaining, whereas a field with a high status quo outcome may very well stagnate. Even though average performance is significantly divergent, the low status quo outcome yields better performance than the high status quo outcome 8.3% of the time when the 1000 runs are pitted in head-to-head competition. Cutting the data more systematically, Figure 5 depicts the distribution of performance at 100 rounds for high and low status quo outcomes for when agents continue to experiment (survivors—left panel) and when learning has stopped (non-survivors—right panel). The possibility for overtaking in performance is reflected in the fact that the distributions overlap. Indeed, not only might a performance from the low status quo outcome surpass high status quo performance, but it can significantly outperform it. Combining this result with those

for long-run performance (specifically Figure 3), these results demonstrate how the randomness of outcomes has a lasting impact on performance. An initial setback does not change the long-term rate of growth but it does establish a lower starting point. Thus, a set-back is baked into long-run performance, so to speak, and this leads in expectation to significantly lower long-run performance.

4.5 Longer Planning Horizons

The analysis heretofore leads to the question of how experimentation is affected when agents possess longer planning horizons. In this section we allow for forward looking behavior and show that the features of optimal learning—including existence—are qualitatively similar to those in the main model, especially if agents are not too patient.

Suppose that each agent lives for two periods, after which he is replaced by another agent who lives for two periods, and so on. Each agent inherits the information discovered by his predecessors and he discounts any future payoff he expects by a discount factor $\delta \in [0,1]$ per period. Finally, let the agents' utility function be given by the exponential utility function (10) that we used in Section 4.4. All else is as in the main model, including the assumption that status quo outcome is given by $m_0 > \widehat{m}$.

In the second period of each agent's life the decision problem is identical to that in our main model. The question then is how the agents behave in the first period of their lives. To answer this question, we begin with the first agent. Actions to the left of the status quo now need not be dominated by the status quo but each is dominated by an action to the right of the status quo that produces the same variance but with higher expected outcome. Therefore, the agent will take an action only weakly to the right of the status quo and his first period problem is given by

$$\max_{\Delta_1 \geq 0} W\left(m_0 + \mu\Delta_1, \sigma^2\Delta_1\right) + \delta\left[\int_{-\infty}^{\widetilde{z}} u\left(m_0\right) \mathrm{d}F\left(z_1\right) + \int_{\widetilde{z}}^{\infty} W\left(m_1 + \mu\Delta_2\left(m_1\right), \sigma^2\Delta_2\left(m_1\right)\right) \mathrm{d}F\left(z_1\right)\right],$$

where $m_1 = m_0 + \mu\Delta_1 + \sigma\sqrt{\Delta_1}z_1$ is the outcome in the first period, $\Delta_2\left(m_1\right)$ is his optimal experiment in the second period if experimentation is optimal, $\widetilde{z}$ is the realization of z_1 such that $m_1 = \widetilde{m}\left(m_0\right)$, and $\widetilde{m}\left(m_0\right)$ is the threshold outcome above which it is optimal for the agent to experiment in the second period. The characterization of the threshold $\widetilde{m}\left(m_0\right)$ is in Lemma 3 and that of the optimal experiment $\Delta_2\left(m_1\right)$ is in Proposition 3. The first term in the objective function is the agent's expected utility in the first period, which is the only thing he cares about if he is myopic. Since the agent is now forward looking, however, he also cares about his expected utility in the second period of his life, which is given by the second and the third terms in the objective function. To understand these terms, note that if $z_1 \leq \widetilde{z}$, the agent's first period outcome is below the threshold

$\widetilde{m}(m_0)$ and in the second period it is optimal for him to take the status quo action, in which case he realizes $u(m_0)$. If, instead, $z_1 \geq \widetilde{z}$, the agent's first period outcome is sufficiently high for him to engage in experimentation again in the second period, in which case his expected utility is given by the integrand in the third term.

We characterize the solution to this problem formally in the appendix and describe it informally here. As in our main model, there is a threshold level of the status quo outcome below which the agent does not experiment and above which he does. If $\delta = 0$, the threshold and the size of the optimal experiment are the same as in the main model. As δ increases, the threshold falls and the size of the optimal experiment increases. These comparative statics reflect the option value of experimentation that make it more attractive. Importantly, however, both the threshold and the optimal experiment are finite, even if $\delta = 1$. Therefore, the option value of experimentation does not swamp the agent's risk aversion and an optimal action always exists.

Consider now the subsequent agents. In any period t in which an agent is in the first period of his life, his behavior is characterized by the following threshold rule. If the outcome $m(h_t)$ generated by the right-most action h_t is above an upper threshold, the agent experiments to the right of h_t. An optimal action exists and its characterization is analogous to that of the first agent's first period action.

If, instead, $m(h_t)$ falls below a lower threshold, the agent takes an untried action between a_0 and h_t. In contrast to a short-lived agent, a longer-lived agent sometimes finds it optimal to experiment between known actions. Even though doing so lowers the agent's expected utility when he is young, it offers the benefit that, should he find a better action, he can choose it again when he is old and benefit twice. In spite of this difference, it is still the case that once the agent has taken an action between a_0 and h_t, doing so will also be optimal in all future periods. Therefore, a single sufficiently low outcome realization is enough to put a permanent end to exploring to the right of h_t, even when agents are longer-lived.

The third possibility is that $m(h_t)$ falls between the two thresholds. In this case the agent may experiment to the left or to the right of h_t depending on the exact outcomes that have been realized by previous agents. Moreover, even if the agent experiments to the left of h_t, future agents may find it optimal to restart experimentation to the right of h_t. This contrasts to the main model as now experimentation may take place in waves. This intermediate region, however, exists only when the agents are sufficiently patient. In particular, there exists a $\overline{\delta} > 0$ such that the upper threshold is strictly above the lower threshold if and only if $\delta \geq \overline{\delta}$.

As one would expect, longer-lived agents engage in more experimentation than short-lived ones,

yet the qualitative features of optimal experimentation are very similar to those in the main model, especially when agents are not too patient. Even when agents are very patient, the option value of experimenting never swamps the agents' risk aversion and an optimal learning rule always exists.

4.6 Pilots

So far we have assumed that agents can only take a single action per period. If an agent wants to explore an untried action, he has to put his entire income at stake and cannot simply try out the action with a small pilot experiment. We make this assumption because we are motivated by situations in which experiments are by necessity at such a scale that the stakes are large. While experiments are often significant, they need not be quite as significant as we have assumed so far. Established firms, for instance, may be able to find out demand for a new product by introducing it on a small scale while continuing to produce existing ones. In this section we extend our model to explore how such pilot experiments affect optimal learning by trial and error.

To do so, suppose that each agent can take two actions. In particular, suppose that in any period t, the agent can put a fraction θ_t of his resources in any action a and the rest in any action a', as long as $\theta_t \in [\underline{\theta}, 1-\underline{\theta}]$, where $\underline{\theta} \in (0, 1/2]$. The assumption that $\underline{\theta} > 0$ captures the fact that the type of experiments we are concerned with cannot be infinitesimally small.

Consider now any period t. The agent's expected outcome is given by

$$\mathrm{E}\left[m_t\right] = \theta_t \mathrm{E}\left[m\left(a\right)\right] + \left(1-\theta_t\right)\mathrm{E}\left[m\left(a'\right)\right] \tag{12}$$

and the variance is given by

$$\mathrm{Var}\left(m_t\right) = \theta_t^2 \mathrm{Var}\left(m\left(a\right)\right) + \left(1-\theta_t\right)^2 \mathrm{Var}\left(m\left(a'\right)\right) + 2\left(1-\theta_t\right)\theta_t \mathrm{Cov}\left(m\left(a\right), m\left(a'\right)\right). \tag{13}$$

In the appendix, we show that in any period t, the agent will either put all his resources into the best known action $\overline{a}_t$ or he will put fraction $(1-\underline{\theta})$ into the best known action and the rest into action $h_t + \Delta_t(\cdot)$, where $\Delta_t(\cdot)$ solves

$$\max_{\Delta_t \in [0,\infty)} \mathrm{E}\left[u\left(\left(1-\underline{\theta}\right)m\left(\overline{a}_t\right) + \underline{\theta} m\left(h_t\right) + \underline{\theta}\mu\Delta_t + \underline{\theta}\sqrt{\Delta_t}\sigma z_1\right)\right]. \tag{14}$$

Notice that this problem is very similar to the agents' problem in the main model and, indeed, identical if we set $\underline{\theta} = 1$. The above expression shows that allowing for pilots increases the risk adjusted return from μ/σ^2 to $\left(\mu/\underline{\theta}\sigma^2\right)$. Adding pilots also increases the effective starting point for experimentation from $m\left(h_t\right)$ to $[(1-\underline{\theta})\, m\left(\overline{a}_t\right) + \underline{\theta} m\left(h_t\right)]$. This increase represents a wealth effect that lowers the agents' risk aversion, encouraging them to experiment more.

Documents/Figures/graphics/O87WJ505

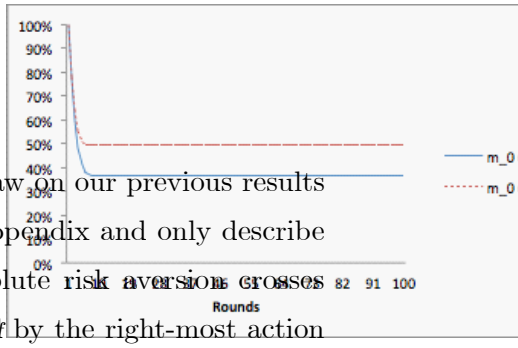


Figure 6: Survival Probability with Pilots

Given the similarity with the problem in the main model, we can draw on our previous results to characterize the agents' optimal actions. We do so formally in the appendix and only describe optimal behavior informally here: Suppose that the coefficient of absolute risk aversion crosses $2\mu/\left(\underline{\theta}\sigma^2\right)$ so that a non-trivial solution exists. If the outcome generated by the right-most action $m\left(h_t\right)$ falls below a threshold, the agent puts all his resources into the best known action. If, instead, $m\left(h_t\right)$ is above the threshold, the agent puts a fraction $\left(1-\underline{\theta}\right)$ of his resources into the best known action and a fraction $\underline{\theta}$ into a unique, finite action strictly to the right of the right-most known action.

The implications of this optimal learning rule differ from those without pilots in three ways. First, allowing for pilots favors experimentation. Other things equal, the threshold above which agents engage in experimentation is lower, and the optimal actions if they do engage in experimentation are larger, than in the model without pilots. The reason is the increase in the risk adjusted return and the wealth effect we noted above. To illustrate the positive effect of pilots on experimentation, we repeat the simulation we performed above for $m_0 = 1$ and set $\underline{\theta} = 0.5$.

Second, allowing for pilots changes how peak performance affects experimentation. In particular, and in contrast to the model without pilots, optimal step size $\Delta_t\left(\cdot\right)$ is increasing, and the threshold $\widetilde{m}\left(\overline{m}_t\right)$ can be decreasing, in peak performance, even when peak performance does not occur at the frontier. The reason is again the wealth effect: since agents are able to exploit the best known action at the same time as they explore untried actions, higher peak performance reduces their risk aversion and induces them to experiment more.

The third implication is that the ability to pilot experiments reduces long-run divergence in performance. Recall that in the simulations for the main model the long-run average performance

for the high status quo outcome was about nine times that for the low status quo outcome. Re-running these simulations for $\underline{\theta} = 0.5$, this gap is now only about three times. This may at first be surprising since high status quo runs tend to have higher peak performance and, as just noted, higher peak performance favors experimentation when pilots are feasible. The reason why pilots seem to help learning more when the status quo outcome is low is that it reduces the prospect that learning gets caught in the performance trap. Figure 6 plots the probability that learning continues after each round when pilots are possible. Comparing this to Figure 4 for the main model, it is clear that pilots increases survival rates much more significantly for the low status quo outcome, and it is this effect that limits the long-run divergence in performance.

5 Discussion

The model we develop provides a rich representation of the uncertainty agents face in their environment. Yet, in so doing, it points to other possible avenues of interest, beyond those we pursue here. Perhaps the most interesting next step is to enrich the underlying economic environment, both via the knowledge held by agents and by the generating process. In our model agents' theoretical knowledge is perfect and unchanging as they know precisely the value of the drift and variance terms. All agents lack is knowledge of the particular realized setting that they face. It is straightforward to suppose that the agents must learn also about how their world was created—about the drift and variance parameters—and use this knowledge to inform their actions. This deepening of the decision problem would add realism to the agents' problem and bring into play several possibilities of substantive interest, in particular whether momentum effects will emerge in experimentation and performance.

Another possibility is to model uncertainty via stochastic processes other than the Brownian motion. As we describe in the introduction, an appealing feature of the Brownian motion is that the risk-return trade-off is independent of performance and action. This neutrality is not present in other stochastic processes. In that case the effects of risk aversion become intertwined with the risk-return opportunities available given the search history. Behavior will then naturally vary as these forces rise and fall in relative importance and may give rise to interesting dynamics that fit particular applications more tightly than does the Brownian motion. The simplest way to see the possibilities is to suppose that the Brownian motion has concave rather than linear drift. Expected return then declines as agents experiment further to the right. For sufficient concavity, the expected return becomes sufficiently small that agents stop experimenting altogether once they have explored enough of the action space. In this formulation, early performance is not only predictive of long-

run performance but fully determinative. A field will only rise as far as it can in the early periods while drift is sufficiently attractive and inevitably all fields stagnate. This dynamic resonates with applications in which innovation has early lock-in.

A property of concave drift is that expected returns decline as a function of actions and not outcomes. The mechanism of action can be flipped by instead modeling the underlying generating process as a geometric Brownian motion. In this formulation, expected return and variance both vary in the performance level but are independent of the action. However, a geometric Brownian motion produces outcomes that are lognormally rather than normally distributed. This hinders tractability and a full analysis of behavior is not possible.[7] Nevertheless, some insight can be obtained by characterizing the marginal incentive for initial experimentation and examining how it varies in performance.

In the appendix we derive the ratio of expected return to risk for the geometric Brownian motion and show that it declines in the performance level. (For comparison, this ratio for the Brownian motion, $\frac{\mu}{\sigma^2}$, is constant in performance). Working against this decline is that better performing agents are less risk averse. Behavior depends, therefore, on the relative decline of innovation opportunities to risk aversion. If risk aversion declines sufficiently fast then the logic of our results will hold up. We provide in the appendix a class of sumex utility functions that satisfy standard risk aversion for which this is true. For this class, the decline in risk aversion dominates the simultaneous decline in growth opportunities such that better performers are more inclined to experiment, as is the case in our model.

On the other hand, if growth opportunities decline faster then does risk aversion, it may be that the high performing agents are amongst those who do not experiment. This possibility has obvious implications for long-run performance and it may be that performance instead converges over time. This would be the case, for example, if utility were exponential as then absolute risk aversion is constant and independent of performance. (Exponential utility does not, therefore, satisfy the requirements of our model).[8] This combination produces a reflection of our results. Rather than fixing opportunities for innovation and establishing that declining risk aversion drives performance divergence, this combination of assumptions fixes risk aversion and suggests that a decline in innovation opportunities as performance increases leads to performance convergence.

Perhaps the most interesting possibilities are where changing risk aversion and growth oppor-

[7] The exception is exponential utility. We discuss this special case below.

[8] In the Brownian motion setting, exponential utility implies either that no agent experiments (for risk aversion above $\frac{2\mu}{\sigma^2}$) or that an optimal action does not exist for any agent (for risk aversion below $\frac{2\mu}{\sigma^2}$) as all agents—regardless of performance level—desire an infinite experimental step.

tunities interact. In the appendix we combine geometric Brownian motion with the linex class of utility functions. For this combination, the relative decrease of risk aversion and innovation opportunities is not monotonic as performance increases. Intriguingly, this implies that it is the mid-level performers who are least likely to experiment. At the top end, decreased risk aversion dominates the diminished opportunities to innovate. At the bottom end, the higher risk aversion is itself dominated by the increasingly attractive opportunities to innovate. It is in the middle that risk aversion's relative decline is strongest. Putting the pieces together suggests that it is mid-level performers who will be first to be caught in the performance trap and to stop experimenting. This possibility resonates with the famous middle-income trap of economic growth where it is the mid-level countries that stagnate whereas the poor manage to rise and the rich continue to grow, at least until the poor reach the middle income trap or the rich are unlucky enough to fall back to mid-level performance.[9]

The connection to growth points towards the policy implications of our model. How can a government influence experimentation, learning, and performance, in a way to have a sustainable significant impact on innovation and growth? Returning to our main model, two immediate implications emerge, both deriving from the two-stage dynamic of innovation and growth. First, that a short-term boost to knowledge and performance can have a lasting impact on long-run performance. This is particularly relevant early in the growth process for fields that are barely above the performance trap. An early boost to knowledge substantially lowers the risk of being caught in the performance trap. Notably, the policy intervention need only be temporary to have a lasting impact, as once a field moves into the second stage of development, innovation will become effectively self-sustaining.

A second policy implication is to question the conventional wisdom that government policy should encourage experimentation. One might think that by encouraging experimentation society benefits from more innovations and, consequently, grows faster. The two stage dynamic to growth belies that intuition. Because learning can stop following a substantial set-back, slow and steady growth is superior to faster but more volatile growth.[10] A government policy that encourages bolder experimentation may very well increase the probability that innovation comes to an end. The policy advice that government should restrain experimentation is, however, context specific.

[9] The idea of economic growth as a learning problem has a long but underdeveloped history. See Matsuyama (1996) for an explicit but informal treatment. Exploring the connection between our model and economic growth is a particularly exciting avenue of research.

[10] This is not at odds with the finding that lower performers experiment less and are more likely to fall into the performance trap. The proximity of the performance trap is driven by the low performance and not the small experiment. The claim we are making here separates those effects.

Should learning already have stopped, or should growth have progressed to the second stage, then further experimentation is an unalloyed public good. The conditionality of this policy advice may go some way to explaining the mixed evidence on government innovation policies in practice.

6 Appendix

This appendix contains the proofs for all lemmas and for Propositions 1-3, which are the results that characterize the agents' optimal actions. The proofs for Propositions 4-8, which cover the divergence result and the results in the sections on pilots and longer planning horizons, are in the online appendix.

Recall that

$$R(m,\Delta) \equiv -\frac{\mathrm{E}\left[u''\left(m+\mu\Delta+\sqrt{\Delta}\sigma z\right)\right]}{\mathrm{E}\left[u'\left(m+\mu\Delta+\sqrt{\Delta}\sigma z\right)\right]}$$

and that $R(m,0)$ is equal to the coefficient of absolute risk aversion $r(m)$. For the proofs below it is convenient to also define

$$P(m,\Delta) \equiv -\frac{\mathrm{E}\left[u'''\left(m+\mu\Delta+\sqrt{\Delta}\sigma z\right)\right]}{\mathrm{E}\left[u''\left(m+\mu\Delta+\sqrt{\Delta}\sigma z\right)\right]} \tag{15}$$

and

$$T(m,\Delta) \equiv -\frac{\mathrm{E}\left[u''''\left(m+\mu\Delta+\sqrt{\Delta}\sigma z\right)\right]}{\mathrm{E}\left[u'''\left(m+\mu\Delta+\sqrt{\Delta}\sigma z\right)\right]}. \tag{16}$$

Notice that $P(m,0)$ is equal to the coefficient absolute prudence $-u'''(m)/u''(m)$ (Kimball 1990) and that $T(m,0)$ is equal to the coefficient of absolute temperance $-u''''(m)/u'''(m)$ (Gollier and Pratt 1996). We can now prove the following lemma.

LEMMA A1. *For any $m \in \mathbb{R}$ and $\Delta \geq 0$ we have*

$$R(m,\Delta) \leq P(m,\Delta) \leq T(m,\Delta).$$

Moreover, the first inequality is strict if either $\Delta > 0$ or $\Delta = 0$ and $r'(m) < 0$, where $r(m)$ is the coefficient of absolute risk aversion.

Proof of Lemma A1: (i.) Suppose first that $\Delta = 0$. Differentiating the coefficient of absolute risk aversion, we get

$$r'(m) = R(m,0)\left(R(m,0) - P(m,0)\right).$$

As we mentioned above, Kimball (1993) shows that standard risk aversion implies decreasing absolute risk aversion, that is, $r'(m) \leq 0$. It then follows from the above expression that $R(m,0) \leq P(m,0)$. Moreover, this inequality is strict if absolute risk aversion is strictly decreasing. Similarly, differentiating the coefficient of absolute prudence, we get

$$p'(m) = P(m,0)\left(P(m,0) - T(m,0)\right).$$

Kimball (1993) shows that in a setting such as ours, standard risk aversion is equivalent to decreasing absolute prudence, that is, $p'(m) \leq 0$ for all $m \in \mathbb{R}$. It then follows from the above expression that $P(m,0) \leq T(m,0)$.

(ii.) Suppose now that $\Delta > 0$. Property 5 in Meyer (1987) shows that decreasing absolute risk aversion implies $R(m,\Delta) \leq P(m,\Delta)$ and Result 1 in Eichner and Wagener (2003) shows that decreasing absolute prudence implies $P(m,\Delta) \leq T(m,\Delta)$ (see also page 116 in Gollier (2001)). All we need to do therefore is to show that our assumption that $r(m)$ crosses $2\mu/\sigma^2$, and thus $r'(m) < 0$ for some m, implies $R(m,\Delta) < P(m,\Delta)$. For this purpose, notice that $R(m,\Delta) < P(m,\Delta)$ is equivalent to

$$\mathrm{E}\left[u'\left(M+\sqrt{V}z\right)\right]\mathrm{E}\left[u'''\left(M+\sqrt{V}z\right)\right] - \mathrm{E}\left[u''\left(M+\sqrt{V}z\right)\right]^2 > 0, \tag{17}$$

where $M = m + \mu\Delta$ and $V = \Delta\sigma^2$. We can rewrite this inequality as

$$\mathrm{E}\left[u'\left(M+\sqrt{V}z\right)\right]\mathrm{E}\left[u''\left(M+\sqrt{V}z\right)z\right] - \mathrm{E}\left[u'\left(M+\sqrt{V}z\right)z\right]\mathrm{E}\left[u''\left(M+\sqrt{V}z\right)\right] > 0, \tag{18}$$

where we use the facts that $\mathrm{E}\left[u'\left(M+\sqrt{V}z\right)z\right] = \sqrt{V}\mathrm{E}\left[u''\left(M+\sqrt{V}z\right)\right]$ and $\mathrm{E}\left[u''\left(M+\sqrt{V}z\right)z\right] = \sqrt{V}\mathrm{E}\left[u'''\left(M+\sqrt{V}z\right)\right]$. We can then follow the same argument as in the proof of Property 5 in Meyer (1987) to show that this inequality holds. In particular, let z^* satisfy

$$z^*\int_{-\infty}^{\infty} u'\left(M+\sqrt{V}z\right)dF(z) = \int_{-\infty}^{\infty} u'\left(M+\sqrt{V}z\right)zdF(z),$$

where $F(z)$ is the cumulative density function of the standard normal distribution. We can then rewrite the left-hand side of (18) as

$$\int_{-\infty}^{\infty} u'\left(M+\sqrt{V}z\right)dF(z)\int_{-\infty}^{\infty} r\left(M+\sqrt{V}z\right)u'\left(M+\sqrt{V}z\right)(z^*-z)\,dF(z), \tag{19}$$

where $r\left(M+\sqrt{V}z\right)$ is the coefficient of absolute risk aversion. The first integral is strictly positive. To sign the second integral, notice that

$$\int_{-\infty}^{\infty} u'\left(M+\sqrt{V}z\right)(z^*-z)\,dF(z) = 0$$

and that the integrand changes sign from positive to negative once. Since $r\left(M+\sqrt{V}z\right)$ is everywhere decreasing and strictly decreasing for at least some outcome levels, the second integral in (19) is strictly positive. The overall expression in (19) is therefore strictly positive. ■

Proof of Lemma 1: This lemma is proven in Theorem 1 in Chipman (1973). ■

Proof of Proposition 1: In this proof we use the fact that expected utility is concave in Δ_1, which we prove in Lemma 2. The first-order condition for the agent's problem is given by

$$\frac{\mathrm{d}W\left(m_0+\mu\Delta_1,\Delta_1\sigma^2\right)}{\mathrm{d}\Delta_1}=\mathrm{E}\left[u'\left(\cdot\right)\right]\frac{\sigma^2}{2}\left(\frac{2\mu}{\sigma^2}-R\left(m_0,\Delta_1\right)\right)\begin{cases}=0 & \textit{if } \Delta_1>0\\ \leq 0 & \textit{if } \Delta_1=0.\end{cases} \tag{20}$$

With this condition in mind, we now prove the optimal actions for an m_0 such that (i.) $m_0<\widehat{m}_l$, where $\widehat{m}_l$ denotes the smallest m such that $r\left(m\right)=2\mu/\sigma^2$, (ii.) $m_0>\widehat{m}$, where $\widehat{m}$ denotes the largest m such that $r\left(m\right)=2\mu/\sigma^2$, and (iii.) $m_0\in[\widehat{m}_l,\widehat{m}]$. We then conclude the proof by performing the comparative statics that are summarized in the proposition.

(i.) Optimal action for $m_0<\widehat{m}_l$: In this case, $\mathrm{d}W\left(m_0,0\right)/\mathrm{d}\Delta_1<0$. Since expected utility $W\left(m_0+\mu\Delta_1,\Delta_1\sigma^2\right)$ is concave in $\Delta_1\geq 0$ it then follows that the status quo is the uniquely optimal action.

(ii.) Optimal action for $m_0>\widehat{m}$: In this case, $\mathrm{d}W\left(m_0,0\right)/\mathrm{d}\Delta_1>0$. If an optimal action exists, it is therefore strictly to the right of the status quo.

To prove that an optimal action does exist, it is sufficient to show that there is a $\Delta_1>0$ such that $\mathrm{E}\left[u\left(m_0+\Delta_1+\sigma\sqrt{\Delta_1}z\right)\right]<u\left(m_0\right)$. For this purpose, consider a status quo outcome $\underline{m}$ such that $r(\underline{m})>2\mu/\sigma^2$. For such an outcome the agent strictly prefers the status quo to any action that is strictly to the right of the status quo. Now let $k\left(\Delta_1\right)$ denote the "compensating premium" that would make the agent indifferent between, on the one hand, taking the status quo action and, on the other hand, receiving $k\left(\Delta_1\right)$ and taking an action that is a distance $\Delta_1\geq 0$ to the right of the status quo. Formally, $k\left(\Delta_1\right)$ is given by the k that solves

$$\mathrm{E}\left[u\left(\underline{m}+k+\Delta_1+\sigma\sqrt{\Delta_1}z\right)\right]=u\left(\underline{m}\right)\quad\text{for }\Delta_1\geq 0,$$

where the Implicit Function Theorem ensures that $k\left(\Delta_1\right)$ exists. Implicitly differentiating this expression, we get

$$\frac{\mathrm{d}k\left(\Delta_1\right)}{\mathrm{d}\Delta_1}=\mu\frac{\sigma^2}{2}\left(R\left(\underline{m}+k,\Delta_1\right)-\frac{2\mu}{\sigma^2}\right).$$

Notice that this derivative is strictly positive for $\Delta_1=0$. Differentiating again we get

$$\frac{\mathrm{d}^2 k(\Delta_1)}{\mathrm{d}\Delta_1^2} = R(\underline{m}+k,\Delta_1)\left(\frac{\sigma^2}{2}\right)^2 * \left[(R(\underline{m}+k,\Delta_1)-P(\underline{m}+k,\Delta_1))^2 + P(\underline{m}+k,\Delta_1)(T(\underline{m}+k,\Delta_1)-P(\underline{m}+k,\Delta_1))\right].$$

Lemma A1 implies that this expression is positive for all $\Delta_1 \geq 0$.

Consider now any $m_0 > \widehat{m}$. Since $k(\Delta_1)$ is strictly increasing and convex, there exists a $\Delta_1 > 0$ such that $m_0 = \underline{m} + k(\Delta_1)$. We then have

$$\mathrm{E}\left[u\left(m_0+\Delta_1+\sigma\sqrt{\Delta_1}z\right)\right] = u(\underline{m}) < u(m_0),$$

where the equality follows from the definition of $k(\Delta_1)$ and the inequality from the fact that $k(\Delta_1) > 0$. This implies that an optimal action exists.

Finally, we need to show that the optimal action is unique. It follows from Lemma A1 that the expression for $\mathrm{d}^2W(\cdot,\cdot)/\mathrm{d}\Delta_1^2$ in (23) is strictly negative for any $\Delta_1 > 0$ that satisfies the first-order condition (20). This, in turn, implies that the optimal action is unique.

(iii.) Optimal action for $m_0 \in [\widehat{m}_l, \widehat{m}]$: In this case, $\mathrm{d}W(m_0,0)/\mathrm{d}\Delta_1 = 0$. The status quo is therefore an optimal action. Moreover, it follows from Lemma A1 and (23) that the status quo is the unique optimum.

(iv.) Comparative statics for any $m_0 \geq \widehat{m}$: We first derive the comparative statics for the optimal step size $\Delta(m_0)$ and then turn to those for the threshold $\widehat{m}$. Above we showed that for any $m_0 \geq \widehat{m}$, the uniquely optimal action is given by $a_1^* = a_0 + \Delta(m_0)$, where $\Delta(m_0)$ is the $\Delta_1 \geq 0$ that solves

$$\frac{\mathrm{d}W\left(m_0+\mu\Delta_1,\Delta_1\sigma^2\right)}{\mathrm{d}\Delta_1} = \mathrm{E}\left[u'(\cdot)\right]\frac{\sigma^2}{2}\left(\frac{2\mu}{\sigma^2} - R(m_0,\Delta_1)\right) = 0. \tag{21}$$

Implicitly differentiating this expression we get

$$\frac{\mathrm{d}\Delta(m_0)}{\mathrm{d}m_0} = -\frac{\mathrm{d}^2W\left(m_0+\mu\Delta_1,\Delta_1\sigma^2\right)}{\mathrm{d}\Delta_1\mathrm{d}m_0}\left(\frac{\mathrm{d}^2W\left(m_0+\mu\Delta_1,\Delta_1\sigma^2\right)}{\mathrm{d}\Delta_1^2}\right)^{-1},$$

evaluated at $\Delta_1 = \Delta(m_0)$. We have already observed that the second term on the right hand side is strictly negative. The sign of $\mathrm{d}\Delta(m_0)/\mathrm{d}m_0$ is therefore the same as the sign of $\mathrm{d}^2W(\cdot,\cdot)/\mathrm{d}\Delta_1\mathrm{d}m_0$. Similarly, the signs of $\mathrm{d}\Delta(m_0)/\mathrm{d}\mu$ and $\mathrm{d}\Delta(m_0)/\mathrm{d}\sigma$ are the same as the signs of $\mathrm{d}^2W(\cdot,\cdot)/\mathrm{d}\Delta_1\mathrm{d}\mu$ and $\mathrm{d}^2W(\cdot,\cdot)/\mathrm{d}\Delta_1\mathrm{d}\sigma$. Differentiating (5) with respect to m_0, μ, and σ we get

$$\begin{aligned}
\frac{\mathrm{d}^2W\left(m_0+\mu\Delta_1,\Delta_1\sigma^2\right)}{\mathrm{d}\Delta_1\mathrm{d}m_0} &= \mathrm{E}\left[u''(\cdot)\right]\frac{\sigma^2}{2}\left(\frac{2\mu}{\sigma^2} - P(m_0,\Delta_1)\right) > 0,\\
\frac{\mathrm{d}^2W\left(m_0+\mu\Delta_1,\Delta_1\sigma^2\right)}{\mathrm{d}\Delta_1\mathrm{d}\mu} &= \mathrm{E}\left[u'(\cdot)\right] + \Delta_1\mathrm{E}\left[u''(\cdot)\right]\frac{\sigma^2}{2}\left(\frac{2\mu}{\sigma^2} - P(m_0,\Delta_1)\right) > 0, \text{ and}\\
\frac{\mathrm{d}^2W\left(m_0+\mu\Delta_1,\Delta_1\sigma^2\right)}{\mathrm{d}\Delta_1\mathrm{d}\sigma} &= \sigma\mathrm{E}\left[u''(\cdot)\right] + \mathrm{E}\left[u'''(\cdot)\right]\sigma\Delta_1\frac{\sigma^2}{2}\left(\frac{2\mu}{\sigma^2} - T(m_0,\Delta_1)\right) < 0,
\end{aligned}$$

where the signs follow from $T(m_0, \Delta_1) \geq P(m_0, \Delta_1) > R(m_0, \Delta_1) = 2\mu/\sigma^2$, which in turn follows from Lemma A1 and the first order condition for the optimal step size (21). For an unbounded step size, note that this not holding requires that a $\overline{\Delta}$ exists such that $R\left(m, \overline{\Delta}\right) > \frac{2\mu}{\sigma^2}$ for all m. As the distribution of outcomes for an experiment of size $\overline{\Delta}$ shifts only in mean and without bound as m increases, $r(m) < \frac{2\mu}{\sigma^2}$ for $m > \widehat{m}$ implies this cannot be true. This establishes the comparative statics of $\Delta(m_0)$ with respect to m_0, μ, and σ that are stated in the proposition.

For the comparative static with respect to the agent's risk aversion, consider an alternative utility function $\widehat{u}(m)$ which satisfies the same conditions as $u(m)$ but for which $-\widehat{u}''(m)/\widehat{u}'(m) \geq -u''(m)/u'(m)$ for all m. An agent with utility function $\widehat{u}(m)$ is then more risk averse than an agent with utility function $u(m)$. It follows from Proposition 20 in Gollier (2001) that

$$-\frac{\mathrm{E}\left[\widehat{u}''\left(m_0 + \mu\Delta_1 + \sigma\sqrt{\Delta_1}z\right)\right]}{\mathrm{E}\left[\widehat{u}'\left(m_0 + \mu\Delta_1 + \sigma\sqrt{\Delta_1}z\right)\right]} \geq -\frac{\mathrm{E}\left[u''\left(m_0 + \mu\Delta_1 + \sigma\sqrt{\Delta_1}z\right)\right]}{\mathrm{E}\left[u'\left(m_0 + \mu\Delta_1 + \sigma\sqrt{\Delta_1}z\right)\right]}.$$

It then follows from the first order condition for the optimal step size (21) and the concavity of the expected utility function, that a more risk averse agent takes a smaller action.

Finally, the comparative statics of $\widehat{m}$ with respect to μ, σ^2 and the agent's risk aversion follow immediately from the definition of $\widehat{m}$ and the fact that the coefficient of absolute risk aversion is everywhere decreasing. ■

Proof of Lemma 2: The first derivative of $W\left(m_0 + \mu\Delta_1, \Delta_1\sigma^2\right)$ with respect to Δ_1 is given by (5). Differentiating the expression again, we get

$$\frac{\mathrm{d}^2 W(\cdot,\cdot)}{\mathrm{d}\Delta_1^2} = \mu^2\left[\mathrm{E}\left[u''(\cdot)\right] + 2\left(\frac{\sigma^2}{2\mu}\right)\mathrm{E}\left[u'''(\cdot)\right] + \left(\frac{\sigma^2}{2\mu}\right)^2 \mathrm{E}\left[u''''(\cdot)\right]\right], \tag{22}$$

where we used the facts that $\mathrm{E}\left[u'(\cdot)\, z\right] = \sigma\sqrt{\Delta_1}\mathrm{E}\left[u''(\cdot)\right]$ and $\mathrm{E}\left[u''(\cdot)\, z\right] = \sigma\sqrt{\Delta_1}\mathrm{E}\left[u'''(\cdot)\right]$. We can then use the definitions of $P(m_0, \Delta_1)$ and $T(m_0, \Delta_1)$ in (15) and (16) to rewrite this expression as

$$\begin{aligned}\frac{\mathrm{d}^2 W(\cdot,\cdot)}{\mathrm{d}\Delta_1^2} &= -\mu^2\mathrm{E}\left[u'(\cdot)\right] R(m_0, \Delta_1)\left(\frac{\sigma^2}{2\mu}\right)^2 * \\ &\quad \left[\left(\frac{2\mu}{\sigma^2} - P(m_0, \Delta_1)\right)^2 + P(m_0, \Delta_1)\left(T(m_0, \Delta_1) - P(m_0, \Delta_1)\right)\right].\end{aligned} \tag{23}$$

Lemma A1 implies that this expression is negative for any $\Delta_1 \geq 0$, which proves that expected utility is concave. ■

Proof of Proposition 2: When $r(m) > 2\mu/\sigma^2$ for all $m \in \mathbb{R}$ the agent is too risk averse to engage any risk, regardless of m. The result follows from the definition of $\widehat{m}$ in Proposition 1. If, instead,

$r(m) < 2\mu/\sigma^2$ the agent is sufficiently risk tolerant to engage risk for any m. As this implies, however, that

$$-\mathrm{E}\left[u''\left(m_0+\mu\Delta_1+\sqrt{\Delta_1}\sigma z\right)\right] < 2\mu/\sigma^2\mathrm{E}\left[u'\left(m_0+\mu\Delta_1+\sqrt{\Delta_1}\sigma z\right)\right],$$

it follows from (5) that marginal expected utility is strictly positive for all $\Delta_1 \geq 0$. An optimum therefore doesn't exist. Finally, when $r(m) = 2\mu/\sigma^2$ for all $m \in \mathbb{R}$, the agent is indifferent whether to undertake risk at every wealth level, and an analogous argument to the previous case establishes the result. ■

Proof of Lemma 3: We first show that there exists a unique $\widetilde{m}(m_0) \in (\widehat{m},\, m_0)$ such that

$$u(m_0) = \mathrm{E}\left[u\left(\widetilde{m}(m_0)+\mu\Delta\left(\widetilde{m}(m_0)\right)+\sqrt{\Delta\left(\widetilde{m}(m_0)\right)}\sigma z\right)\right]. \tag{24}$$

For this purpose, notice that

$$\mathrm{E}\left[u\left(\widehat{m}+\mu\Delta(\widehat{m})+\sqrt{\Delta(\widehat{m})}\sigma z\right)\right] = u(\widehat{m}) < u(m_0)$$

and

$$\mathrm{E}\left[u\left(m_0+\mu\Delta(m_0)+\sqrt{\Delta(m_0)}\sigma z\right)\right] > u(m_0).$$

Expected utility $\mathrm{E}\left[u\left(m+\mu\Delta(m)+\sqrt{\Delta(m)}\sigma z\right)\right]$ is therefore strictly less than $u(m_0)$ for $m = \widehat{m}$ and strictly larger than $u(m_0)$ for $m = m_0$. To show the existence of a unique $\widetilde{m}(m_0)$ it is therefore sufficient to show that expected utility $\mathrm{E}\left[u\left(m+\mu\Delta(m)+\sqrt{\Delta(m)}\sigma z\right)\right]$ is strictly increasing in $m \in [\widehat{m}, m_0]$. Applying the Envelope Theorem we obtain

$$\frac{\mathrm{dE}\left[u\left(m+\mu\Delta(m)+\sqrt{\Delta(m)}\sigma z\right)\right]}{\mathrm{d}m} = \mathrm{E}\left[u'\left(m+\mu\Delta(m)+\sqrt{\Delta(m)}\sigma z\right)\right] > 0,$$

which completes the proof of the existence of a unique $\widetilde{m}(m_0) \in (\widehat{m},\, m_0)$.

To prove the comparative statics, we implicitly differentiate (24). Once again applying the Envelope Theorem, we have

$$\frac{\mathrm{d}\widetilde{m}(m_0)}{\mathrm{d}m_0} = \frac{u'(m_0)}{\mathrm{E}\left[u'\left(\widetilde{m}+\mu\Delta\left(\widetilde{m}\right)+\sqrt{\Delta\left(\widetilde{m}\right)}\sigma z\right)\right]} > 0,$$

where the inequality follows from non-satiation.

To show that $\mathrm{d}\widetilde{m}/\mathrm{d}m_0 \leq 1$ we need to establish that (24) implies

$$u'(m_0) \leq \mathrm{E}\left[u'\left(\widetilde{m}+\mu\Delta\left(\widetilde{m}\right)+\sqrt{\Delta\left(\widetilde{m}\right)}\sigma z\right)\right]$$

or, equivalently,

$$v\left(m_0\right) \geq \mathrm{E}\left[v\left(\widetilde{m}+\mu\Delta\left(\widetilde{m}\right)+\sqrt{\Delta\left(\widetilde{m}\right)}\sigma z\right)\right], \tag{25}$$

where we define $v\left(m_0\right)$ as the utility function $v\left(m_0\right)=-u'\left(m_0\right)$. Notice that (24) implies (25) if an agent with utility $v\left(\cdot\right)$ is more risk averse than an agent with utility function $u\left(\cdot\right)$. It is therefore sufficient to show that

$$-\frac{u'''\left(m\right)}{u''\left(m\right)} \geq -\frac{u''\left(m\right)}{u'\left(m\right)} \text{ for all } m \in \mathbb{R},$$

where the LHS is the coefficient of absolute risk aversion associated with $v\left(\cdot\right)$ and the RHS is the one associated with $u\left(\cdot\right)$. This inequality is satisfied since the utility function $u\left(\cdot\right)$ satisfies decreasing absolute risk aversion.

Finally, to obtain the comparative statics, we once again differentiate (24) to obtain

$$\frac{\mathrm{d}\widetilde{m}(m_0)}{\mathrm{d}\mu} = -\frac{1}{\Delta\left(\widetilde{m}\right)} < 0$$

and

$$\frac{\mathrm{d}\widetilde{m}(m_0)}{\mathrm{d}\sigma} = -\frac{\sigma\Delta\left(\widetilde{m}\right)\mathrm{E}\left[u''\left(\widetilde{m}+\mu\Delta\left(\widetilde{m}\right)+\sqrt{\Delta\left(\widetilde{m}\right)}\sigma z\right)\right]}{\mathrm{E}\left[u'\left(\widetilde{m}+\mu\Delta\left(\widetilde{m}\right)+\sqrt{\Delta\left(\widetilde{m}\right)}\sigma z\right)\right]} > 0.$$

To establish the comparative static with respect to the agent's risk aversion, note that we can interpret m_0 as the certainty equivalent of a normally distributed random variable with mean $\widetilde{m}+\mu\Delta\left(\widetilde{m}\right)$ and variance $\Delta\left(\widetilde{m}\right)\sigma^2$. The comparative static then follows from the fact that the certainty equivalent of a lottery is decreasing in the agent's risk aversion (see, for instance, Proposition 6.C.2 in Mas-Colell, Whinston, and Green (1995)). ■

Proof of Proposition 3: Follows immediately from the discussion in the text. ■

References

[1] Arrow, Kenneth. 1962. Economic Welfare and Allocation of Resources for Invention. In *The Rate and Direction of Inventive Activity: Economic and Social Factors. Princeton University.*

[2] Astebro, Thomas, Holger Herz, Ramana Nanda, and Roberto Weber. 2014. Seeking the Roots of Entrepreneurship: Insights from Behavioral Economics. *The Journal of Economic Perspectives*, 28(3): 49-69.

[3] Bell, David E. 1988. One-Switch Utility Functions and a Measure of Risk. *Management Science*, 34(12): 1416-1424.

[4] Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David Mckenzie, and John Roberts. 2013. Does Management Matter? Evidence from India. *Quarterly Journal of Economics*, 128(1): 1-51.

[5] Callander, Steven. 2011. Searching and Learning by Trial and Error. *American Economic Review*, 101(6): 2277-2308.

[6] Catmull, Ed. 2014. *Creativity, Inc.: Overcoming the Unseen Forces That Stand in the Way of True Inspiration.* Random House.

[7] Chang, Seok-Ho, Pamela C. Cosman, and Laurence B. Milstein. 2011. Chernoff-Type Bounds for the Gaussian Error Function. *IEEE Transactions on Communications*, 59(11): 2939-2944.

[8] Chipman, John. 1973. The Ordering of Portfolios in Terms of Mean and Variance. *The Review of Economic Studies*, 40(2): 167-190.

[9] Eichner, Thomas, and Andreas Wagener. 2003. More on Parametric Characterizations of Risk Aversion and Prudence. *Economic Theory*, 21(4): 895-900.

[10] Bell, David E., and Peter C. Fishburn. 2001. Strong One-Switch Utility. *Management Science*, 47(4): 601-604.

[11] Garfagnini, Umberto, and Bruno Strulovici. 2016. Social Experimentation with Interdependent and Expanding Technologies. *Review of Economic Studies 83: 1579-1613.*

[12] Gibbons, Robert and Rebecca Henderson. 2013. What Do Managers Do? In *Handbook of Organizational Economics*, eds. Robert Gibbons and John Roberts, Princeton University Press.

[13] Gollier, Christian. 2001. *The Economics of Risk and Time*. MIT Press.

[14] ——— and John Pratt. 1996. Risk Vulnerability and the Tempering Effect of Background Risk. *Econometrica*, 64(5): 1109-1123.

[15] Harris, Christopher, and John Vickers. 1987. Racing with Uncertainty. *Review of Economic Studies 54: 1-21.*

[16] Kimball, Miles. 1990. Precautionary Saving in the Small and in the Large. *Econometrica*, 58(1): 53-73.

[17] ———. 1993. Standard Risk Aversion. *Econometrica*, 61(3): 589-611.

[18] Klepper, Steven. 2015. *Experimental Capitalism: The Nanoeconomics of American High-Tech Industries.* Princeton University Press.

[19] Mas-Colell, Andreu, Michael Whinston, and Jerry Green. 1995. *Microeconomic Theory.* Oxford University Press.

[20] Matsuyama, Kiminori. 1996. Economic Development as Coordination Problems. In *The Role of Government in East Asian Development: Comparative Institutional Analysis*, eds. Masahiko Aoki, Hyung-Ki Kim, and Masahiro Okuno-Fujiwara, Oxford University Press.

[21] Merton, Robert K. 1936. The Unanticipated Consequences of Purposive Social Action. *American Sociological Review*, 1(Dec): 894-904.

[22] Meyer, Jack. 1987. Two-Moment Decision Models and Expected Utility Maximization. *American Economic Review*, 77(3): 421-430.

[23] Sheff, David. 1985. Interview with Steve Jobs. *Playboy.*

[24] Syverson, Chad. 2011. What Determines Productivity?, *Journal of Economic Literature*, 49(2): 326-65.

7 For Online Publication

This appendix proves the results on divergence, imitation, pilots, and longer planning horizons.

7.1 Long-Run Performance

Proof of Proposition 4: We can interpret our model as one in which, instead of a sequence of short-lived agents, there is a single, infinitely long-lived but myopic agent. For expositional convenience, we use this interpretation in this proof. We can then denote the agent in charge of Field $k = L, H$ and Agent k.

Suppose that Agent L engages in optimal experimentation. This is going to generate some outcome m_1^{L*} in the first period, m_2^{L*}in the second, and so on. Now take any period T and let τ denote the largest $t \in [1, T]$ in which Agent L engaged in experimentation. Note that since $m_0^L > \widehat{m}$, Agent L engages in experimentation in the first period and thus $\tau \in [1, T]$. We can now write the outcome Agent L realized each period as

$$m_t^{L*} = \begin{cases} m_{t-1}^{L*} + \mu\Delta\left(m_{t-1}^{L*}\right) + \sigma\sqrt{\Delta\left(m_{t-1}^{L*}\right)}Z_t \text{ for } t = 1, ..., \tau \\ \overline{m}_{\tau^L+1}^{L*} \text{ for } t = \tau + 1, ..., T, \end{cases}$$

where Z_t is the realization of a random variable z_t that is drawn from a standard normal distribution and where we streamline our notation by defining $m_0^{L*} \equiv m_0^L$.

Consider now Agent H. Since $m_0^H > \widehat{m}$ this agent also engages in experimentation in the first period. Suppose now that if Agent H engages in experimentation in a period $t = 1, ..., \tau$ he happens to realize the same Z_t that Agent L realized. We will show that it must then be the case that

$$\mathrm{E}_{T+1}\left[m_{T+1}^{H*} - m_{T+1}^{L*}\right] \geq m_0^H - m_0^L, \tag{26}$$

where the inequality is strict for some values of Z_1,...,Z_τ. Since this result holds for any T and any Z's, it implies (9).

To show (26), we first need to introduce two definitions. For the first definition, consider some period t in which both agents find it optimal to engage in experimentation. We know from above that each agent's outcome is given by

$$m_t^{k*} = m_{t-1}^{k*} + \mu\Delta\left(m_{t-1}^{k*}\right) + \sigma\sqrt{\Delta\left(m_{t-1}^{k*}\right)}Z_t \quad \text{for } k = H, L$$

We then define

$$\widehat{z}_t\left(m_{t-1}^{H*}, m_{t-1}^{L*}\right) \equiv -\frac{\mu\left(\Delta\left(m_{t-1}^{H*}\right) - \Delta\left(m_{t-1}^{L*}\right)\right)}{\sigma\left(\sqrt{\Delta\left(m_{t-1}^{H*}\right)} - \sqrt{\Delta\left(m_{t-1}^{L*}\right)}\right)} \tag{27}$$

as the value of Z_t such that $m_t^{H*} - m_t^{L*} = m_{t-1}^{H*} - m_{t-1}^{L*}$. For the second definition, recall that an agent engages in experimentation in period $t+1$ if and only if $m_t^* > \widetilde{m}\left(\overline{m}_t\right)$, where $\overline{m}_t = \max\left[m_0, m_1^*, ...m_{t-1}^*\right]$. We then define

$$\widetilde{z}_t\left(m_{t-1}^{L*}, \overline{m}_t^L\right) \equiv -\frac{m_{t-1}^{L*} + \mu\Delta\left(m_{t-1}^{L*}\right) - \widetilde{m}\left(\overline{m}_t^L\right)}{\sigma\sqrt{\Delta\left(m_{t-1}^{L*}\right)}} \tag{28}$$

as the value of Z_t such that $m_t^* = m_{t-1}^{L*} + \mu\Delta\left(m_{t-1}^{L*}\right) + \sigma\sqrt{\Delta\left(m_{t-1}^{L*}\right)}Z_t = \widetilde{m}\left(\overline{m}_t^L\right)$.

Since Agent L engages in experimentation in periods $2, ..., \tau$ it must be that

$$Z_t > \widetilde{z}_t\left(m_{t-1}^{L*}, \overline{m}_t^L\right) \text{ for all } t = 1, ..., \tau - 1. \tag{29}$$

In Lemma A2 below we show that if (29) holds then it must be that

$$\widetilde{z}_t\left(m_{t-1}^{L*}, \overline{m}_t^L\right) > \max[\widetilde{z}_t\left(m_{t-1}^{H*}, \overline{m}_t^H\right), \widehat{z}_t\left(m_{t-1}^{H*}, m_{t-1}^{L*}\right)] \text{ for all } t = 1, ..., \tau. \tag{30}$$

To see the implications of this result, suppose first that $\tau = T$, in which case Agent L is experimenting in all periods up to and including period T. Since $Z_t > \widetilde{z}_t\left(m_{t-1}^{L*}, \overline{m}_t^L\right)$ for all $t = 1, ..., T-1$, it follows from (30) that

$$Z_t > \widetilde{z}_t\left(m_{t-1}^{H*}, \overline{m}_t^H\right) \quad \text{and} \quad Z_t > \widehat{z}_t\left(m_{t-1}^{H*}, m_{t-1}^{L*}\right) \text{ for all } t = 1, ..., T-1.$$

Together with the fact that $m_0^H > \widehat{m}$, the first inequality implies that Agent H also engages in experimentation in all periods up to and including period T. And the second inequality implies that

$$m_{T-1}^{H*} - m_{T-1}^{L*} > m_{T-2}^{H*} - m_{T-2}^{L*} > ... > m_0^H - m_0^L. \tag{31}$$

Finally, since (30) holds for $t = T$ it must be that either (i.) $Z_T > \widetilde{z}_T\left(m_T^{L*}, \overline{m}_T^L\right)$, (ii.) $Z_T < \widetilde{z}_T\left(m_T^{H*}, \overline{m}_T^H\right)$, or (iii.) $Z_T \in \left(\widetilde{z}_T\left(m_T^{H*}, \overline{m}_T^H\right), \widetilde{z}_T\left(m_T^{L*}, \overline{m}_T^L\right)\right)$. We will show next that (26) holds in any one of those three cases.

Case (i.): If $Z_T > \widetilde{z}_T\left(m_T^{L*}, \overline{m}_T^L\right)$ both agents experiment in period $T+1$. We then have

$$\begin{aligned} \mathrm{E}_{T+1}\left[m_{T+1}^{H*} - m_{T+1}^{L*}\right] &= m_T^{H*} + \mu\Delta\left(m_T^{H*}\right) - m_T^{L*} - \mu\Delta\left(m_T^{L*}\right) \\ &> m_T^{H*} - m_T^{L*} \\ &> m_0^H - m_0^L, \end{aligned} \tag{32}$$

where the first inequality follows from the fact that the optimal experiment is strictly increasing in the outcome. To derive the second inequality, notice that since $Z_T > \widehat{z}_T\left(m_{T-1}^{H*}, m_{T-1}^{L*}\right)$ we have $m_T^{H*} - m_T^{L*} > m_{T-1}^{H*} - m_{T-1}^{L*}$. The second inequality then follows from (29).

Case (ii.): If $Z_T < \widetilde{z}_T\left(m_T^{H*}, \overline{m}_T^H\right)$ neither agent experiments in period T. We then have

$$\mathrm{E}_{T+1}\left[m_{T+1}^{H*} - m_{T+1}^{L*}\right] = \overline{m}_T^H - \overline{m}_T^L > m_0^H - m_0^L.$$

To see the inequality, let $\overline{\tau}^k \in \{0, 1, ..., T-1\}$ denote the period in which the outcome of Agent $k = H, L$ peaked, that is, in which $m_{\overline{\tau}^k}^{k*} = \overline{m}_T^{k*}$ (where $\overline{\tau}^k = 0$ is the case in which the outcome was below status quo outcome in $t = 1, 2, ..., T-1$). Suppose first that $\overline{\tau}^H = \overline{\tau}^L$. Then

$$\overline{m}_T^H - \overline{m}_T^L = m_{\overline{\tau}^H}^{H*} - \overline{m}_{\overline{\tau}^H}^{L*} = m_0^H - m_0^L \text{ if } \overline{\tau}^H = 0.$$

and

$$\overline{m}_T^H - \overline{m}_T^L = m_{\overline{\tau}^H}^{H*} - \overline{m}_{\overline{\tau}^H}^{L*} > m_0^H - m_0^L \text{ if } \overline{\tau}^H > 0,$$

where the inequality follows from (31). Suppose next that $\overline{\tau}^H \neq \overline{\tau}^L$. Then

$$\overline{m}_T^H - \overline{m}_T^L > m_{\overline{\tau}^L}^{H*} - \overline{m}_{\overline{\tau}^L}^{L*} > m_0^H - m_0^L,$$

where the first inequality follows from $\overline{m}_T^H > m_{\overline{\tau}^L}^{H*}$ and $\overline{m}_T^L = m_{\overline{\tau}^L}^{L*}$ and the second inequality follows from (31).

Case (iii.): If $Z_T \in \left(\widetilde{z}_T\left(m_T^{H*}, \overline{m}_T^H\right), \widetilde{z}_T\left(m_T^{L*}, \overline{m}_T^L\right)\right)$ Agent H engages in experimentation in period $T+1$ but Agent L does not. Since Agent H prefers engaging in experimentation to realizing his previous peak $\overline{m}_T^H$ it must be that $\mathrm{E}_{T+1}\left[m_{T+1}^{H*}\right] > \overline{m}_T^H$. We therefore have

$$\mathrm{E}_{T+1}\left[m_{T+1}^{H*} - m_{T+1}^{L*}\right] > \overline{m}_T^H - \overline{m}_T^L \geq m_0^H - m_0^L,$$

where the second inequality follows from our discussion in Case (ii.) above.

To complete the proof, suppose that $\tau < T$. Since τ is the last period in which Agent L engaged in experimentation, we have $m_t^{L*} = \overline{m}_\tau^L$ for all $t \geq \tau + 1$. Since $Z_t > \widetilde{z}_t\left(m_{t-1}^{L*}, \overline{m}_t^L\right)$ for all $t = 1, ..., \tau - 1$, it follows from (29) that

$$Z_t > \widetilde{z}_t\left(m_{t-1}^{H*}, \overline{m}_t^H\right) \quad \text{and} \quad Z_t > \widehat{z}_t\left(m_{t-1}^{H*}, m_{t-1}^{L*}\right) \text{ for all } t = 1, ..., \tau - 1.$$

Together with the fact that $m_0^H > \widehat{m}$, the first inequality implies that Agent H also engages in experimentation in all periods up to and including period τ. And the second inequality implies that

$$m_{\tau-1}^{H*} - m_{\tau-1}^{L*} > m_{\tau-2}^{H*} - m_{\tau-2}^{L*} > ... > m_0^H - m_0^L. \tag{33}$$

Finally, since (30) holds for $t = \tau$ it must be that either (a.) $Z_\tau < \widetilde{z}_\tau\left(m_\tau^{H*}, \overline{m}_\tau^H\right)$ or (b.) $Z_\tau \in \left(\widetilde{z}_\tau\left(m_\tau^{H*}, \overline{m}_\tau^H\right), \widetilde{z}_\tau\left(m_\tau^{L*}, \overline{m}_\tau^L\right)\right)$. We will show next that (26) holds in either of those cases:

Case (a.): If $Z_\tau < \widetilde{z}_\tau\left(m_\tau^{H*}, \overline{m}_\tau^H\right)$ then Agent H also does not experiment in periods $t = \tau+1, ..., T$. We then have

$$\mathrm{E}_{T+1}\left[m_{T+1}^{H*} - m_{T+1}^{L*}\right] = \overline{m}_\tau^H - \overline{m}_\tau^L \geq m_0^H - m_0^L,$$

where the inequality follows from our discussion in Case (ii.) above.

Case (b.): If $Z_\tau \in \left(\widetilde{z}_\tau\left(m_\tau^{H*}, \overline{m}_\tau^H\right), \widetilde{z}_\tau\left(m_\tau^{L*}, \overline{m}_\tau^L\right)\right)$ then Agent H does engage in experimentation in period $\tau+1$ and, possibly, in period $T+1$. Since, in period $T+1$, Agent H can guarantee himself $\overline{m}_\tau^H$ it must be that $\mathrm{E}_{T+1}\left[m_{T+1}^{H*}\right] \geq \overline{m}_\tau^H$. We therefore have

$$\mathrm{E}_{T+1}\left[m_{T+1}^{H*} - m_{T+1}^{L*}\right] \geq \overline{m}_\tau^H - \overline{m}_\tau^L \geq m_0^H - m_0^L,$$

where, once again, the second inequality follows from our discussion in Case (ii.) above. ■

LEMMA A2. *If*

$$Z_t > \widetilde{z}_t\left(m_{t-1}^{L*}, \overline{m}_t^L\right) \text{ for all } t = 1, ..., \tau-1. \tag{34}$$

then

$$\widetilde{z}_t\left(m_{t-1}^{L*}, \overline{m}_t^L\right) > \max[\widetilde{z}_t\left(m_{t-1}^{H*}, \overline{m}_t^H\right), \widehat{z}_t\left(m_{t-1}^{H*}, m_{t-1}^{L*}\right)] \text{ for all } t = 1, ..., \tau. \tag{35}$$

Proof of Lemma A2: To prove this lemma, we first show that (35) holds for $t = 1$. We then show that if (35) holds for $1, ..., x$, where $1 < x \leq \tau - 1$, then it also holds for $t = x+1$. Together these facts imply that (35) holds for $t = 1, ..., \tau$ as claimed in the lemma.

Suppose first then that $t = 1$. From the definitions of $\widehat{z}_t(\cdot)$ and $\widetilde{z}_t(\cdot)$ in (27) and (28) we have

$$\widetilde{z}_1\left(m_0^L, m_0^L\right) - \widehat{z}_1\left(m_0^H, m_0^L\right) = \frac{\widetilde{m}\left(m_0^L\right) + \mu\sqrt{\Delta\left(m_0^L\right)}\sqrt{\Delta\left(m_0^H\right)} - m_0^L}{\sigma\sqrt{\Delta\left(m_0^L\right)}}. \tag{36}$$

and

$$\widetilde{z}_1\left(m_0^L, m_0^L\right) - \widetilde{z}_1\left(m_0^H, m_0^H\right) = \frac{m_0^H - m_0^L + \widetilde{m}\left(m_0^L\right) - \widetilde{m}\left(m_0^H\right)}{\sigma\sqrt{\Delta\left(m_0^H\right)}} \tag{37}$$

$$+\frac{\sqrt{\Delta\left(m_0^H\right)} - \sqrt{\Delta\left(m_0^L\right)}}{\sigma\sqrt{\Delta\left(m_0^L\right)}\sqrt{\Delta\left(m_0^H\right)}}\left(\widetilde{m}\left(m_0^L\right) + \mu\sqrt{\Delta\left(m_0^L\right)}\sqrt{\Delta\left(m_0^H\right)} - m_0^L\right).$$

To see that (36) is strictly positive notice that

$$\begin{aligned} & \widetilde{m}\left(m_0^L\right)+\mu\sqrt{\Delta\left(m_0^L\right)}\sqrt{\Delta\left(m_0^H\right)}-m_0^L \\ > & \widetilde{m}\left(m_0^L\right)+\mu\Delta\left(m_0^L\right)-m_0^L \\ > & \widetilde{m}\left(m_0^L\right)+\mu\Delta\left(\widetilde{m}\left(m_0^L\right)\right)-m_0^L \\ > & 0, \end{aligned} \tag{38}$$

where the first inequality follows from $\Delta\left(m_0^H\right) > \Delta\left(m_0^L\right)$ and the second follows from $m_0^L > \widetilde{m}\left(m_0^L\right)$. To see the third inequality, recall that $\widetilde{m}\left(m_0^L\right)$ is defined as the outcome at which the agent is indifferent between receiving m_0^L and a normally distributed gamble that pays $\widetilde{m}\left(m_0^L\right) + \mu\Delta\left(\widetilde{m}\left(m_0^L\right)\right)$ on average and has a strictly positive variance $\Delta\left(\widetilde{m}\left(m_0^L\right)\right)\sigma^2$. Since the agent is risk averse, it must then be that $\widetilde{m}\left(m_0^L\right)+\mu\Delta\left(\widetilde{m}\left(m_0^L\right)\right) > m_0^L$.

To show that (37) is also strictly positive, consider first the second term on the RHS of (37). Since $\Delta\left(m_0^H\right) > \Delta\left(m_0^L\right)$ it follows from (38) that this term is strictly positive. Consider next the first term on the RHS of (37). This term has to be weakly positive since $m_0^H > m_0^L$ and $\mathrm{d}\widetilde{m}\left(m\right)/\mathrm{d}m \in (0,1]$. We therefore have $\widetilde{z}_1\left(m_0^L, m_0^L\right) > \max[\widetilde{z}_1\left(m_0^H, m_0^H\right) > \widehat{z}_t\left(m_0^H, m_0^L\right)]$.

Suppose now that (35) holds for $t = 1, ..., x$, where $1 < x \leq \tau - 1$. We will show that (35) then also holds for $t = x + 1$. For this purpose, notice first that if (35) holds for $t = 1, ..., x$, then it follows from (34) that (i.) both agents are engaging in experimentation in periods $t = 1, ..., x+1$ and (ii.) it must be that

$$m_{x+1}^{H*} - m_{x+1}^{L*} > m_x^{H*} - m_x^{L*} > ... > m_0^H - m_0^L. \tag{39}$$

Furthermore, we know from the definitions of $\widehat{z}_t\left(\cdot\right)$ and $\widetilde{z}_t\left(\cdot\right)$ in (27) and (28) that

$$\widetilde{z}_{x+1}\left(m_x^{L*}, \overline{m}_{x+1}^L\right) - \widehat{z}_{x+1}\left(m_x^{H*}, m_x^{L*}\right) = \frac{\widetilde{m}\left(\overline{m}_{x+1}^L\right) + \mu\sqrt{\Delta\left(m_x^{L*}\right)}\sqrt{\Delta\left(m_x^{H*}\right)} - m_x^{L*}}{\sigma\sqrt{\Delta\left(m_x^{L*}\right)}}.$$

To see that this expression is strictly positive, notice that

$$\begin{aligned} & \widetilde{m}\left(\overline{m}_{x+1}^L\right)+\mu\sqrt{\Delta\left(m_x^{L*}\right)}\sqrt{\Delta\left(m_x^{H*}\right)}-m_x^{L*} \\ \geq & \widetilde{m}\left(m_x^L\right)+\mu\sqrt{\Delta\left(m_x^{L*}\right)}\sqrt{\Delta\left(m_x^{H*}\right)}-m_x^{L*} \\ > & \widetilde{m}\left(m_x^L\right)+\mu\Delta\left(m_x^{L*}\right)-m_x^{L*} \\ > & \widetilde{m}\left(m_x^L\right)+\mu\Delta\left(\widetilde{m}\left(m_x^L\right)\right)-m_x^{L*} \\ > & 0, \end{aligned} \tag{40}$$

where the first inequality follows from $\overline{m}_{x+1}^L \geq m_x^L$, the second from $\Delta\left(m_x^{H*}\right) > \Delta\left(m_x^{L*}\right)$, and the third from $m_x^{L*} > \widetilde{m}\left(m_x^L\right)$. To derive the last inequality, recall that $\widetilde{m}\left(m_x^L\right)$ is defined as the

outcome at which the agent is indifferent between receiving m_x^L and a normally distributed gamble that pays $\widetilde{m}\left(m_x^L\right)+\mu\Delta\left(\widetilde{m}\left(m_x^L\right)\right)$ on average and has a strictly positive variance $\Delta\left(\widetilde{m}\left(m_x^L\right)\right)\sigma^2$. Since the agent is risk averse, it must then be that $\widetilde{m}\left(m_x^L\right)+\mu\Delta\left(\widetilde{m}\left(m_x^L\right)\right)-m_x^{L*}$. We therefore have that if (35) holds for $t=1,...,x$, then $\widetilde{z}_{x+1}\left(m_x^{L*},\overline{m}_{x+1}^L\right)>\widehat{z}_{x+1}\left(m_x^{H*},m_x^{L*}\right)$.

Next, we know from the definition of $\widetilde{z}_t(\cdot)$ in (28) that

$$\widetilde{z}_{x+1}\left(m_x^{L*},\overline{m}_x^{L*}\right)-\widetilde{z}_{x+1}\left(m_x^{H*},\overline{m}_x^{H*}\right)=\frac{m_x^{H*}-m_x^{L*}+\widetilde{m}\left(\overline{m}_{x+1}^{L*}\right)-\widetilde{m}\left(\overline{m}_{x+1}^{H*}\right)}{\sigma\sqrt{\Delta\left(m_x^{H*}\right)}} \tag{41}$$
$$+\frac{\sqrt{\Delta\left(m_x^{H*}\right)}-\sqrt{\Delta\left(m_x^{L*}\right)}}{\sigma\sqrt{\Delta\left(m_x^{L*}\right)}\sqrt{\Delta\left(m_x^{H*}\right)}}\left(\widetilde{m}\left(\overline{m}_{x+1}^{L*}\right)+\mu\sqrt{\Delta\left(m_x^{L*}\right)}\sqrt{\Delta\left(m_x^{H*}\right)}-m_x^{L*}\right).$$

We know from (40) that the first term on the RHS is strictly positive. To show that the second term is weakly positive, we first show that

$$m_x^{H*}-m_x^{L*}\geq\overline{m}_{x+1}^{H*}-\overline{m}_{x+1}^{L*}.$$

For this purpose, let $\overline{\tau}^k\in\{0,1,...,x\}$denote the period in which the outcome of Agent $k=H,L$ peaked, that is, in which $m_{\overline{\tau}^k}^{k*}=\overline{m}_{x+1}^{k*}$ (where $\overline{\tau}^k=0$ is the case in which the outcome was below the status quo outcome in $t=1,2,...,x$). Suppose now that $\overline{\tau}^H=\overline{\tau}^L$. Then

$$m_x^{H*}-m_x^{L*}\geq m_{\overline{\tau}^H}^{H*}-m_{\overline{\tau}^H}^{L*}=\overline{m}_{x+1}^{H*}-\overline{m}_{x+1}^{L*},$$

where the inequality follows from (39) and the equality follows from the definition of $\overline{\tau}^H$ and $\overline{\tau}^L$.

Suppose next that $\overline{\tau}^H\neq\overline{\tau}^L$. Then

$$m_x^{H*}-m_x^{L*}\geq m_{\overline{\tau}^H}^{H*}-m_{\overline{\tau}^H}^{L*}>m_{\overline{\tau}^H}^{H*}-m_{\overline{\tau}^L}^{L*}=\overline{m}_{x+1}^{H*}-\overline{m}_{x+1}^{L*},$$

where the first inequality follows from (39), the second follows from $m_{\overline{\tau}^L}^{L*}>m_{\overline{\tau}^H}^{L*}$, and the equality follows from the definition of $\overline{\tau}^H$ and $\overline{\tau}^L$. We therefore have $m_x^{H*}-m_x^{L*}\geq\overline{m}_{x+1}^{H*}-\overline{m}_{x+1}^{L*}$ as claimed above.

Finally, notice that

$$\overline{m}_{x+1}^{H*}-\overline{m}_{x+1}^{L*}\geq\widetilde{m}\left(\overline{m}_{x+1}^{H*}\right)-\widetilde{m}\left(\overline{m}_{x+1}^{L*}\right),$$

where the inequality follows from $\overline{m}_x^{H*}>\overline{m}_x^{L*}$ and $\mathrm{d}\widetilde{m}(m)/\mathrm{d}m\in(0,1]$. We therefore have

$$m_x^{H*}-m_x^{L*}\geq\overline{m}_{x+1}^{H*}-\overline{m}_{x+1}^{L*}\geq\widetilde{m}\left(\overline{m}_{x+1}^{H*}\right)-\widetilde{m}\left(\overline{m}_{x+1}^{L*}\right)$$

which implies that the second term on the RHS of (41) is weakly positive. ■

Proof of Proposition 5: The logic of escape is easiest to see for the example utility function of Section 4.4. We prove the result in this special case. The key property in the argument is that

the experiment step size increases without bound in m. As this property holds generally, it is straightforward to extend the argument to the full model.

We establish the result by proving a stronger result: That the probability agents achieve at least half of the expected gain in performance in every period is strictly bounded away from zero. Beginning at $m_0 > \hat{m}$, this critical threshold we denote $m_1^{\#}$, is $m_0 + \frac{1}{2}(m_0 - \hat{m})\, c\mu$, where $c > 0$ is a constant. The probability of success is $1 - CDF\left(m_1^{\#}\right) = \frac{1}{2}\left[1 - \operatorname{erf}\left(m_1^{\#}\right)\right] = 1 - \frac{1}{2} erfc\left(-m_1^{\#}\right)$, where erf and $erfc$ are the error function and complementary error function, respectively, and using the identities $\operatorname{erf}(x) = -\operatorname{erf}(-x)$ and $erfc(x) = 1 - \operatorname{erf}(x)$. From Chang, Cosman, and Milstein (2011) we have $erfc(x) \leq e^{-x^2}$, and thus the success probability satisfies:

$$\Pr_e(1) \geq 1 - \frac{1}{2} e^{-\left[\frac{\frac{1}{2}(m_0 - \hat{m})c\mu}{\sigma\sqrt{2(m_0 - \hat{m})c}}\right]^2}$$

Take $m_1^{\#}$ as the realized outcome and iterate. The same set of calculations give:

$$\Pr_e(2) \geq 1 - \frac{1}{2} e^{-\left[\frac{\frac{1}{2}(m_0 - \hat{m})\left(\frac{1}{2}c\mu + 1\right)c\mu}{\sigma\sqrt{2(m_0 - \hat{m})\left(\frac{1}{2}c\mu + 1\right)c}}\right]^2} = 1 - \frac{1}{2} e^{-\left[\frac{\mu^2}{8\sigma^2}\right](m_0 - \hat{m})\left(\frac{1}{2}c\mu + 1\right)c}$$

And this generalizes to:

$$\Pr_e(t) \geq 1 - \frac{1}{2} e^{-\left[\frac{\mu^2}{8\sigma^2}\right](m_0 - \hat{m})\left(\frac{1}{2}c\mu + 1\right)^{t-1}c}$$

The escape probability is:

$$\Pr_e \geq \prod_{t=1}^{\infty} \Pr_e(t) = \prod_{t=1}^{\infty} \left[1 - \frac{1}{2} e^{-wz^{t-1}}\right]$$

where $w > 0$ and $z > 1$ are constants.

As $0 < 1 - e^{-wz^{t-1}} < 1$,

$$\prod_{t=1}^{T} \left(1 - \frac{1}{2} e^{-wz^{t-1}}\right)$$

converges as $T \to \infty$. Taking the logarithm of this product, we get that the product converges to a positive value if

$$\sum_{t=1}^{T} \ln\left(1 - \frac{1}{2} e^{-wz^{t-1}}\right)$$

converges as $T \to \infty$. To prove that, we first note the classical inequality[11]

$$\ln(1 - x) \geq -\frac{3}{2}x$$

[11] For example, see `http://functions.wolfram.com/ElementaryFunctions/Log/29/`

for $0 \leq x \leq 1/2$. We also note that since $z > 1$, $z^t/t \to \infty$ and so there exists T_0 such that if $t \geq T_0$, $z^{t-1} > t$. Hence, for any $T > T_0$, we have the inequality

$$\sum_{t=T_0}^{T} \ln\left(1 - \frac{1}{2}e^{-wz^{t-1}}\right) \geq \sum_{t=T_0}^{T} \ln\left(1 - \frac{1}{2}e^{-wt}\right) \geq -\frac{3}{2}\sum_{t=T_0}^{T}\frac{1}{2}e^{-wt} \geq -\frac{3}{4}\sum_{t=0}^{\infty}e^{-wt} = -\frac{3}{4}.\frac{1}{1-e^{-w}}$$

as each term $\ln\left(1 - e^{-wz^{t-1}}\right)$ is negative. This establishes the necessary convergence.

7.2 Longer Planning Horizons

We assume that $\beta > 2\frac{\mu}{\sigma^2}$, which ensures that the crossing condition is satisfied, and that $m_0 > \widehat{m}$, which ensures that the first agent engages in experimentation.

PROPOSITION 6. *In the second period of their lives, agents behave as in the main model. An optimal action therefore exists and is given by*

$$a_t^* = \begin{cases} \overline{a}_t & \text{if } m(h_t) \leq \widetilde{m}(m(h_t)) \\ h_t + \Delta(m(h_t)) & \text{if } m(h_t) \geq \widetilde{m}(m(h_t)), \end{cases}$$

where

$$\Delta(m(h_t)) = \max\left\{0, 2\frac{m(h_t) - \widehat{m}}{\sigma^2\left(\beta - \frac{2\mu}{\sigma^2}\right)}\right\}, \tag{42}$$

$$\widehat{m} = -\frac{1}{\beta}\ln\left[\frac{2\alpha\mu}{\sigma^2\beta\left(\beta - 2\frac{\mu}{\sigma^2}\right)}\right], \tag{43}$$

and

$$\widetilde{m}(m(h_t)) = \widehat{m} + \frac{\beta - \frac{2\mu}{\sigma^2}}{\alpha\beta}\left(u(m(\overline{a}_t)) - u(\widehat{m})\right).$$

The agent's expected utility from taking the best action is given by

$$W\left(m(h_t) + \mu(a_t^* - h_t), (a_t^* - h_t)\sigma^2\right) = \begin{cases} u(m(\overline{a}_t)) & \text{if } m(h_t) \leq \widetilde{m}(m(h_t)) \\ u(\widehat{m}) + (m(h_t) - \widehat{m})\frac{\alpha\beta}{\left(\beta - \frac{2\mu}{\sigma^2}\right)} & \text{if } m(h_t) \geq \widetilde{m}(m(h_t)). \end{cases} \tag{44}$$

Proof: In the second period of their lives, the agents only care about expected utility from that period. They therefore behave just like agents in the main model. Given the exponential utility function, expected utility from taking action $h_t + \Delta_t$ is given by

$$W\left(m(h_t) + \mu\Delta_t, \Delta_t\sigma^2\right) = \alpha(m(h_t) + \mu\Delta_t) - \exp\left(-\beta(m(h_t) + \mu\Delta_t) + \frac{1}{2}\beta^2\Delta_t\sigma^2\right). \tag{45}$$

The expressions in the proposition then follow from Propositions 1-3. ■

PROPOSITION 8. *In period $t = 1$ an optimal action exists. If, as we assume, $m_0 > \widehat{m}$, any optimal action is strictly to the right of a_0 and increasing in the discount factor δ, where $\widehat{m}$ is defined in* (43)*. If, instead, m_0 were weakly smaller than $\widehat{m}$, the agent would take the status quo action a_0.*

Proof: In the first period it can never be optimal for the agent to take an action strictly to the left of a_0. If it exists, the optimal first period action is therefore weakly to the right of a_0. The problem of characterizing the optimal first period actions that are weakly to the right of a_0 is a special case of the problem of characterizing the optimal actions in any period t in which a agent is in the first period of his life and is constrained to taking an action weakly to the right of the right-most action h_t. Since this more general problem is relevant for the proof of the next proposition, we examine it here.

Consider then any period t in which a agent is in the first period of his life and suppose that he has to take an action $a_t \geq h_t$. We know from the previous proposition that in $t+1$ the agent will then experiment to the right of a_t if and only if

$$m(a_t) \geq \widetilde{m}(\overline{m}_{t+1}). \tag{46}$$

This optimal learning rule is equivalent to the agent experimenting to the right if and only if

$$m(a_t) \geq \widetilde{m}(\overline{m}_t). \tag{47}$$

To see this, notice that the two inequalities are only different if

$$m(a_t) > \max\{m_0, ..., m_{t-1}\}. \tag{48}$$

Since

$$m(a_t) \geq \widetilde{m}(m(a_t))$$

and

$$\max\{m_0, ..., m_{t-1}\} > \widetilde{m}(\max\{m_0, ..., m_{t-1}\}).$$

inequality (48) implies (46) and (47) which, in turn, implies that the two learning rules are equivalent.

Next, we can write $m(a_t)$ as

$$m(a_t) = m(h_t) + \mu\Delta_t + \sigma\sqrt{\Delta_t}z_t.$$

Substituting this expression into (47) and rearranging, we have that in period $t+1$ the agent experiments to the right if and only if

$$z_t \geq \widetilde{z}_t,$$

where

$$m\left(h_t\right) + \mu\Delta_t + \sigma\sqrt{\Delta_t}\widetilde{z} = \widetilde{m}\left(\overline{m}_t\right)$$

or equivalently

$$\widetilde{z}_t = \left(-\frac{m\left(h_t\right) - \widetilde{m}\left(\overline{m}_t\right) + \mu\Delta_t}{\sigma\sqrt{\Delta_t}}\right). \tag{49}$$

We can therefore write the agent's problem as

$$\max_{\Delta_t \geq 0} V_r,$$

where

$$\begin{aligned} V_r &= \alpha\left(m\left(h_t\right) + \mu\Delta_t\right) - \exp\left(-\beta\left(m\left(h_t\right) + \mu\Delta_t\right) + \frac{1}{2}\beta^2\Delta_t\sigma^2\right) \\ &\quad +\delta\left[\int_{-\infty}^{\widetilde{z}} u\left(\overline{m}_t\right)\mathrm{d}F\left(z_t\right) + \int_{\widetilde{z}}^{\infty} u\left(\widehat{m}\right) + \left(m\left(h_t\right) + \mu\Delta_t + \sigma\sqrt{\Delta_t}z_t - \widehat{m}\right)\frac{\alpha\beta}{\left(\beta - \frac{2\mu}{\sigma^2}\right)}\mathrm{d}F\left(z_t\right)\right] \end{aligned} \tag{50}$$

and the subscript 'r' stands for 'to the right of h_t.' Differentiating V_r we get

$$\begin{aligned} \frac{\mathrm{d}V_r}{\mathrm{d}\Delta_t} &= \alpha\mu - \frac{1}{2}\beta\sigma^2\left(\beta - \frac{2\mu}{\sigma^2}\right)\exp\left(-\beta\left(m\left(h_t\right) + \mu\Delta_t\right) + \frac{1}{2}\beta^2\Delta_t\sigma^2\right) \\ &\quad +\delta\frac{\alpha\beta}{\left(\beta - \frac{2\mu}{\sigma^2}\right)}\left(\mu\left(1 - F\left(\widetilde{z}_t\right)\right) + \sigma\frac{1}{2\sqrt{\Delta_t}}f\left(\widetilde{z}_t\right)\right). \end{aligned}$$

Taking limits we further have that

$$\lim_{\Delta_t \to 0}\frac{\mathrm{d}V_r}{\mathrm{d}\Delta_t} = \begin{cases} > 0 & \text{if } m\left(h_t\right) > \widehat{m} \\ = 0 & \text{if } m\left(h_t\right) = \widehat{m} \\ < 0 & \text{if } m\left(h_t\right) < \widehat{m} \end{cases} \tag{51}$$

and

$$\lim_{\Delta_t \to \infty}\frac{\mathrm{d}V_r}{\mathrm{d}\Delta_t} = -\infty.$$

This implies that if $m\left(h_t\right) > \widehat{m}$ there exists an optimal $\Delta > 0$ that maximizes the agent's expected utility. Moreover, since

$$\frac{\mathrm{d}^2V_r}{\mathrm{d}\Delta_t\mathrm{d}\delta} = \frac{\alpha\beta}{\left(\beta - \frac{2\mu}{\sigma^2}\right)}\left(\mu\left(1 - F\left(\widetilde{z}_t\right)\right) + \sigma\frac{1}{2\sqrt{\Delta_t}}f\left(\widetilde{z}_t\right)\right) \geq 0$$

the optimal action is increasing in δ.

To establish the agent's optimal action if $m(h_t) \leq \widehat{m}$ notice that

$$\begin{aligned}\frac{\mathrm{d}^2 V_r}{\mathrm{d}\Delta_t \mathrm{d}m(h_t)} &= \frac{1}{2}\beta^2\sigma^2\left(\beta - \frac{2\mu}{\sigma^2}\right)\exp\left(-\beta\left(m(h_t)+\mu\Delta_t\right)+\frac{1}{2}\beta^2\Delta_t\sigma^2\right) \\ &\quad +\delta\frac{\alpha\beta}{\left(\beta-\frac{2\mu}{\sigma^2}\right)}\left(\mu\Delta_t+\widetilde{m}\left(\overline{m}_t\right)-m(h_t)\right)\frac{1}{\sigma 2\sqrt{\Delta_t}\Delta_t}f\left(\widetilde{z}_t\right)\end{aligned}$$

which is strictly positive for all $\Delta_t \geq 0$ if $m(h_t) \leq \widetilde{m}(\overline{m}_t)$. Since $\widehat{m} < \widetilde{m}(\overline{m}_t)$ it then follows from (51) that if $m(h_t) \leq \widehat{m}$ the agent's optimal action to the right of h_t is given by h_t.

PROPOSITION 9. *Consider any period $t = 3, 5, ...$ in which one of the agents that follows the first agent is in the first period of his life. In any such period, an optimal action exists. Any optimal action is strictly to the right of h_t if $m(h_t) \geq m_{h,t}$ and it is an unknown action to the left of h_t if $m(h_t) \leq m_{l,t}$, where $m_{l,t}$ and $m_{h,t} \geq m_l$ are defined in the proof. Moreover, there exists a $\overline{\delta} > 0$ such that $m_{l,t} = m_{h,t}$ for all $\delta \leq \overline{\delta}$.*

Proof: In the proof of the previous proposition we characterized the solution to the problem that an agent in the first period of his life faces if he has to take an action to the right of h_t. Consider now an alternative constrained problem in which the agent has to take an action between two neighboring actions a_l and a_h with $a_l < a_r \leq h_t$ and $m(a_l) \geq m(a_h)$. The agent's expected utility is then given by

$$\begin{aligned}V_l &= \mathrm{E}\left[u\left(\mathrm{E}\left[m\left(a_t\right)\right]+\sqrt{\mathrm{Var}\left(m\left(a_t\right)\right)}z_t\right)\right]+\delta\int_{\widetilde{z}_t}^{\infty}u\left(\mathrm{E}\left[m\left(a_t\right)\right]+\sqrt{\mathrm{Var}\left(m\left(a_t\right)\right)}z_t\right)\mathrm{d}F\left(z_t\right) \quad (52)\\ &\quad +\delta\int_{-\infty}^{\widetilde{z}_t}\max\left\{u\left(m\left(\overline{a}_t\right)\right),\mathrm{E}\left[u\left(m\left(h_t\right)+\mu\Delta\left(h_t\right)+\sigma\sqrt{\Delta\left(h_t\right)}z_{t+1}\right)\right]\right\}\mathrm{d}F\left(z_t\right),\end{aligned}$$

where

$$\mathrm{E}\left[m\left(a_t\right)\right] = \frac{a_t - a_l}{a_r - a_l}m\left(a_r\right)+\frac{a_r - a_t}{a_r - a_l}m\left(a_l\right),$$

$$\mathrm{Var}\left(m\left(a_t\right)\right) = \frac{\left(a_t - a_l\right)\left(a_r - a_t\right)}{a_r - a_l}\sigma^2,$$

$$\mathrm{E}\left[m\left(a_t\right)\right]+\sqrt{\mathrm{Var}\left(m\left(a_t\right)\right)}\widetilde{z}_t = \overline{\overline{m}},$$

$$u\left(\overline{\overline{m}}\right) = \max\left\{u\left(m\left(\overline{a}_t\right)\right),\mathrm{E}\left[u\left(m\left(h_t\right)+\mu\Delta\left(h_t\right)+\sigma\sqrt{\Delta\left(h_t\right)}z_{t+1}\right)\right]\right\},$$

and where the subscript 'l' stands for 'to the left of h_t.' Differentiating V_l and taking limits we get

$$\lim_{a_t\to a_l}\frac{\mathrm{d}V_l}{\mathrm{d}a_t} = \begin{cases}\infty & \text{if } m(a_l) = \overline{\overline{m}} \\ \frac{1}{2}\sigma^2 u'\left(a_l\right)\left(-\frac{2}{\sigma^2}\frac{m(a_l)-m(a_r)}{(a_r-a_l)} - r\left(m\left(a_l\right)\right)\right) < 0 & \text{if } m(a_l) < \overline{\overline{m}},\end{cases} \quad (53)$$

where $r(\cdot)$ is the coefficient of absolute risk aversion. If $m(a_l) = \overline{\overline{m}}$, therefore, there exists an action strictly between a_l and a_r that the agent strictly prefers to both a_l and a_r.

Suppose now that $m(h_t) \leq \widehat{m}$. We have established in the proof of the previous proposition that a young agent then prefers h_t to any action to the right of h_t. Moreover, if $m(h_t) \leq \widehat{m}$, it must be that $m(\overline{a}_t) = \overline{\overline{m}}$, in which case there exists an unknown action between $\overline{a}_t$ and its neighboring actions that the agent prefers to $\overline{a}_t$ and thus all all other known actions. For any $m(h_t) \leq m_{l,t} \equiv \widehat{m}$, it is therefore optimal for the agent to take an unknown action to the left of h_t.

Next, it is immediate that if h_t is the best known action, there always exists an action to the right of h_t that the agent prefers to any action to the left of h_t. For any $m(h_t) \geq m_{h,t} \equiv \overline{m}_{t-1}$, it is therefore optimal for the agent to take an unknown action to the right of h_t.

Suppose next that $m(h_t) \in (\widehat{m}, \widetilde{m}(\overline{m}_{t-1}))$. The agent's expected utility from taking the best action to the left of h_t is bounded from below by $u(m(\overline{a}_t))$. Suppose now that $\delta = 0$. The agent's optimal action to the right of h_t is then given by $\Delta(m(h_t))$ and his expected utility from taking this action is given by $W\left(m(h_t) + \mu\Delta(m(h_t)), \Delta(m(h_t))\sigma^2\right) < u(m(\overline{a}_t))$, where $\Delta(m(h_t))$ and $W(\cdot,\cdot)$ are given by (42) and (45). Moreover, it follows from (50) that at $\delta = 0$, the derivative of the agent's expected utility from experimenting to the right of h_t is given by

$$\frac{\mathrm{d}V_r}{\mathrm{d}\delta} = \int_{-\infty}^{\widetilde{z}} u(\overline{m}_t)\,\mathrm{d}F(z_t) + \int_{\widetilde{z}}^{\infty} u(\widehat{m}) + \left(m(h_t) + \mu\Delta(m(h_t)) + \sigma\sqrt{\Delta(m(h_t))}z_t - \widehat{m}\right)\frac{\alpha\beta}{\left(\beta - \frac{2\mu}{\sigma^2}\right)}\mathrm{d}F(z_t),$$

where $\widetilde{z}$ is defined in (49). Since this derivative is finite, it follows that there exists a $\overline{\delta} > 0$ such that for all $\delta \leq \overline{\delta}$ the expected utility from taking the best action to the left of h_t is strictly larger than the expected utility from taking the best action to the right of h_t.

Finally, suppose that $m(h_t) \in (\widetilde{m}(\overline{m}_{t-1}), \overline{m}_{t-1})$, in which case $m(\overline{a}_t) < \overline{\overline{m}}$. As we observed above, there then exists a $\overline{\delta} > 0$ such that for all $\delta \leq \overline{\delta}$ the agent never finds it optimal to take an action strictly in between two known actions. For any such δ, the best action to the left of h_t is therefore given by $\overline{a}_t$. The expected utility from taking the best action to the right of h_t is bounded from below by $W\left(m(h_t) + \mu\Delta(m(h_t)), \Delta(m(h_t))\sigma^2\right)$, where $\Delta(m(h_t))$ is the best myopic action to the right of h_t. Since $m(h_t) \in (\widetilde{m}(\overline{m}_{t-1}), \overline{m}_{t-1})$, we know that $W\left(m(h_t) + \mu\Delta(m(h_t)), \Delta(m(h_t))\sigma^2\right) > u(m(\overline{a}_t))$, which implies that for all $\delta \leq \overline{\delta}$ the expected utility from experimenting to the right of h_t is strictly larger than that from experimenting to the left of h_t.

There therefore exists a $\overline{\delta} > 0$ such that for all $\delta \leq \overline{\delta}$, $m_{l,t} = m_{h,t} = \widetilde{m}(\overline{m}_{t-1})$. ■

7.3 Pilots

PROPOSITION 7. *In any period $t \geq 1$ there exists a unique, optimal action for the agent. If $m(h_t) \leq \widetilde{m}(\overline{m}_t)$, the agent puts all his income into the best known action $\overline{a}_t$, where $\widetilde{m}(\overline{m}_t)$ is defined in the proof. If $m(h_t) > \widetilde{m}(\overline{m}_t)$, the agent puts a fraction $(1-\underline{\theta})$ of his income into the best known action $\overline{a}_t$ and the rest into action $h_t + \Delta(m(h_t))$, where $\Delta(m(h_t)) > 0$ is defined in the proof. The optimal step size $\Delta(m(h_t))$ is increasing in μ and $\overline{m}_t$ and decreasing in σ^2 and the agent's risk aversion. The threshold $\widetilde{m}(\overline{m}_t)$ is decreasing in μ, increasing σ^2 and the agent's risk aversion, and can be increasing or decreasing in $\overline{m}_t$. Moreover, an increase in the minimum feasible scale of a pilot $\underline{\theta}$ leads to an increase in the $\widetilde{m}(\overline{m}_t)$ and a reduction in $\Delta(m(h_t))$.*

Proof: Notice first that the agent will never take a known action other than the best known action, that he will never take an unknown action to the left of the right-most action h_t, and that the agent will also never take two unknown action to the left of h_t. The first two claims are immediate. To prove the last claim, suppose that in some period t the agent puts $(1-\theta_t)$ of his income into an action $a' > h_t$ and the rest into another action $a > a'$, where $\theta_t \in (0,1)$. Let $\Delta' = a' - h_t$ and $\Delta = a - a'$. The agent's expected outcome and its variance are then given by

$$\mathrm{E}[m_t] = m(h_t) + \mu\left(\Delta' + \theta_t \Delta\right)$$

and

$$\mathrm{Var}(m_t) = \sigma^2\left(\Delta' + \theta_t^2 \Delta\right),$$

where we are using (13) and the fact that $\mathrm{Cov}(m(a'), m(a)) = \mathrm{Var}(m(a'))$. The agent can then ensure himself the same expected outcome at a strictly lower variance by reducing Δ' and increasing Δ appropriately. This proves that the agent will never take two unknown actions to the left of h_t.

We already observed in the text that if the agent does put some income into an unknown action, he will never put in more than $\underline{\theta}$. In any period t, the agent will therefore put all his income into the best known action or he will put a fraction $(1-\underline{\theta})$ into the best known action and the rest into an unknown action to the right of h_t. Consider first the constrained problem in which the agent has to experiment to the right of h_t, which is given by

$$\max_{\Delta_t \in [0,\infty)} \mathrm{E}\left[u\left((1-\underline{\theta})\, m(\overline{a}_t) + \underline{\theta} m(h_t) + \underline{\theta}\mu\Delta_t + \underline{\theta}\sqrt{\Delta_t}\sigma z_t\right)\right].$$

This is exactly the same problem as the constrained problem in the main model if the drift and the variance of the Brownian motion were given by $\underline{\theta}\mu$ and $\underline{\theta}^2\sigma^2$ and the outcome generated by the right-most action were $(1-\underline{\theta})\, m(\overline{a}_t) + \underline{\theta} m(h_t)$. Moreover, the agent's unconstrained problem– which involves comparing the agent's expected utility from taking the constrained optimal action

with his utility from taking the best action–is also the same as appropriately specified version of the main model. Except for the comparative statics with respect to $\overline{m}_t$ and $\underline{\theta}$, all the claims in the proposition therefore follow from Propositions 1-3.

For the comparative static with respect to $\underline{\theta}$, suppose that the Δ_t that maximizes the agent's constrained problem is strictly positive and denote it by Δ^*. If the agent does experiment, his expected outcome is then given by $\mathrm{E}[m_t] = m(h_t) + \mu\underline{\theta}\Delta^*$ and the variance is given by $\mathrm{Var}(m_t) = \sigma^2\underline{\theta}^2\Delta^*$. Suppose now that the minimum feasible scale $\underline{\theta}$ is reduced to $\underline{\theta} < \underline{\theta}$. The agent can then achieve the same expected outcome with a lower variance by reducing the fraction of income he invests in the unknown action to $\underline{\theta}$ and increasing the step size. This has two implications. First, since expected utility is concave in the step size, the new optimal step size is strictly larger than Δ^*. This proves that reduction in $\underline{\theta}$ increases the optimal step size. Second, the above observation implies that a reduction in $\underline{\theta}$ increases the agent's expected utility from taking the optimal action. It then follows from the proof of Proposition 3 that a reduction in $\underline{\theta}$ reduces $\widetilde{m}(\overline{m}_t)$.

Finally, consider the comparative statics with respect to $\overline{m}_t$. The result that the optimal step size $\Delta(m(h_t))$ is increasing in $\overline{m}_t$ follows immediately from the fact that in the main model the optimal step size is increasing in $m(h_t)$. To see that the threshold $\widetilde{m}(\overline{m}_t)$ can now be decreasing in $\overline{m}_t$, suppose that the utility function is given by (10). It then follows from Section 4.4 that

$$\widetilde{m}(\overline{m}_{t-1}) = \frac{1}{\underline{\theta}}\widehat{m} - \frac{(1-\underline{\theta})}{\underline{\theta}}\overline{m}_{t-1} + \frac{\left(\beta - \frac{2\mu}{\underline{\theta}\sigma^2}\right)}{\underline{\theta}\alpha\beta}\left(u(\overline{m}_{t-1}) - u(\widehat{m})\right).$$

Differentiating this expression we get

$$\frac{\mathrm{d}\widetilde{m}(\overline{m}_{t-1})}{\mathrm{d}\overline{m}_{t-1}} = -\frac{(1-\underline{\theta})}{\underline{\theta}} + \frac{\left(\beta - \frac{2\mu}{\underline{\theta}\sigma^2}\right)}{\underline{\theta}\alpha\beta}\left(\alpha + \beta\exp(-\beta\overline{m})\right).$$

This expression will be negative if

$$\left(\beta - \frac{2\mu}{\underline{\theta}\sigma^2}\right)\beta\exp(-\beta\overline{m}) < \alpha\left(\frac{2\mu}{\underline{\theta}\sigma^2} - \beta\underline{\theta}\right)$$

which, in turn, will be the case if

$$\frac{2\mu}{\underline{\theta}^2\sigma^2} > \beta > \frac{2\mu}{\underline{\theta}\sigma^2}$$

and $\overline{m}$ is sufficiently large. ■

7.4 Stochastic Processes

In this section we analyze the case in which the underlying environment is the realized path of a geometric Brownian motion. Recall that the agent's problem is

$$\max_{\Delta} E\left[u(m(\Delta))\right]$$

where $m(\Delta)$ is now given by

$$m(\Delta) = m_0 \exp(\mu\Delta + \sigma W(\Delta))$$

and $W(\Delta)$ is a standard Brownian motion.

Now let z denote a random variable with a standard lognormal distribution. We can then write

$$m(\Delta) = M(\Delta) + \sqrt{\frac{V(\Delta)}{Var(z)}}(z - E[z]),$$

which allows us to rewrite the problem as

$$\max_{\Delta} E\left[u\left(M(\Delta) + \sqrt{\frac{V(\Delta)}{Var(z)}}(z - E[z])\right)\right]$$

The first derivative is

$$\frac{dE[u(...)]}{d\Delta} = M'(\Delta)E\left[u'(..)\right] + \frac{1}{2}\sqrt{\frac{Var(z)}{V(\Delta)}}V'(\Delta)E\left[u'(...)(z - E[z])\right].$$

We wish to examine the derivative at $\Delta = 0$. For this purpose, rewrite the derivative as

$$\frac{dE[u(...)]}{d\Delta} = M'(\Delta)E\left[u'(..)\right] + \frac{1}{2}\sqrt{Var(z)}V'(\Delta)\left[\frac{E[u'(...)(z - E[z])]}{\sqrt{V(\Delta)}}\right].$$

At $\Delta = 0$, both the numerator and the denominator of the term in squared brackets are zero. Applying l'Hopital to the second term on the RHS we get

$$\begin{aligned}&\sqrt{V(\Delta)}\sqrt{Var(z)}\frac{V''(\Delta)}{V'(\Delta)}E\left[u'(...)(z - E[z])\right]\\&+\sqrt{Var(z)}\left(\sqrt{V(\Delta)}M'(\Delta)E\left[u''(...)(z - E[z])\right] + \frac{1}{2}\sqrt{Var(z)}V'(\Delta)E\left[u''(...)(z - E[z])^2\right]\right)\end{aligned}$$

At $\Delta = 0$ this becomes

$$Var(z)^2\frac{1}{2}V'(0)u''(m_0)$$

So at $\Delta = 0$ we have

$$\begin{aligned}\frac{dE[u(...)]}{d\Delta} &= M'(0)u'(m_0) + Var(z)^2\frac{1}{2}V'(0)u''(m_0)\\&= \frac{1}{2}V'(0)u'(m_0)\left[\frac{2M'(0)}{V'(0)} - \left(-\frac{u''(m_0)}{u'(m_0)}\right)Var(z)^2\right]\end{aligned}$$

Note that this expression is exactly the same as for the Brownian motion in which case z is distributed normally and $Var(z) = 1$.

From the properties of the geometric Brownian motion we know that

$$M'(0) = m_0\left(\mu + \frac{1}{2}\sigma^2\right)$$

and

$$V'(0) = m_0^2\sigma^2$$

And so we have

$$\begin{aligned}\frac{dE\left[u\left(...\right)\right]}{d\Delta} &= M'(0)u'(m_0) + Var(z)^2\frac{1}{2}V'(0)u''(m_0) \\ &= \frac{1}{2}V'(0)u'(m_0)\left[\frac{2\left(\mu+\frac{1}{2}\sigma^2\right)}{m_0\sigma^2} - \left(-\frac{u''(m_0)}{u'(m_0)}\right)Var(z)^2\right]\end{aligned}$$

The key difference to the Brownian motion case is that the risk adjusted return is now decreasing in m_0 (the first expression in the brackets). The richer you are, the worse therefore the opportunity for innovation. The sign of the derivative, however, depends on how this compares to declining absolute risk aversion, given by the second term in the brackets. The technological opportunities for innovation would dominate if the coefficient of absolute risk aversion were constant, that is, if we had a regular exponential utility function $u(m) = -\exp(-\beta m)$. In this case, the above becomes

$$\frac{dE\left[u\left(...\right)\right]}{d\Delta} = \frac{1}{2}V'(0)u'(m_0)\left[\frac{2\left(\mu+\frac{1}{2}\sigma^2\right)}{m_0\sigma^2} - \beta Var(z)^2\right].$$

and a threshold value of m_0 exists such that the first derivative is negative at $\Delta = 0$ for starting performance above this threshold.

To capture both declining innovation opportunities *and* declining risk aversion, apply instead the linex utility function used in Section 4.4 Setting $u(m) = am - \exp\left(-\beta m\right)$, the above becomes

$$\frac{dE\left[u\left(...\right)\right]}{d\Delta} = \frac{1}{2}V'(0)u'(m_0)\left[\frac{2\left(\mu+\frac{1}{2}\sigma^2\right)}{m_0\sigma^2} - \left(\frac{\beta^2\exp\left(-\beta m_0\right)}{a+\beta\exp\left(-\beta m_0\right)}\right)Var(z)^2\right]$$

or

$$\frac{dE\left[u\left(...\right)\right]}{d\Delta} = \frac{1}{2}V'(0)u'(m_0)m_0Var(z)^2\left[\frac{2\left(\mu+\frac{1}{2}\sigma^2\right)}{\sigma^2Var(z)^2} - \left(m_0\frac{\beta^2\exp\left(-\beta m_0\right)}{a+\beta\exp\left(-\beta m_0\right)}\right)\right]$$

Numerical calculations show that the second expression in the brackets is single peaked, reaching a maximum for moderate levels of performance. Thus, if the first expression is large—the innovation opportunities are good—then all agents have a marginal incentive to experiment. For less appealing technological opportunities (lower drift or higher variance of the stochastic process), it will be the moderate performers who first experience a disincentive for initial experimentation.

How agents trade-off the attractiveness of innovation opportunities and risk varies, therefore, in the agent's tolerance for risk. Take the following utility function that is the sum of two standard utility functions that are frequently used when outcomes are lognormally distributed.

$$\begin{aligned} u(m) &= a\log m - \frac{1}{m^b} \\ &= a\log m - \exp\left(-b\log m\right) \end{aligned}$$

Straightforward calculations show that the coefficient of risk aversion and prudence are given by

$$\begin{aligned} r &= \frac{b + am^b + b^2}{m\left(b + am^b\right)} \\ p &= \frac{2b + 2am^b + 3b^2 + b^3}{m\left(b + am^b + b^2\right)} \end{aligned}$$

As both are decreasing, and DARA and DAP imply standardness, this utility function satisfies standard risk aversion and the requirements of our model.

Looking again at the first derivative at $\Delta = 0$,

$$\begin{aligned} \frac{dE\left[u\left(...\right)\right]}{d\Delta} &= \frac{1}{2}V'(0)u'(m_0)\left[\frac{2\left(\mu + \frac{1}{2}\sigma^2\right)}{m_0\sigma^2} - \left(\frac{b + am_0^b + b^2}{m_0\left(b + am_0^b\right)}\right)Var(z)^2\right] \\ &= \frac{1}{2m_0}V'(0)u'(m_0)\left[\frac{2\left(\mu + \frac{1}{2}\sigma^2\right)}{\sigma^2} - \left(\frac{b + am_0^b + b^2}{\left(b + am_0^b\right)}\right)Var(z)^2\right]. \end{aligned}$$

which reveals that, for this utility function, the declining innovation opportunities of the geometric Brownian motion are dominated by declining risk aversion. Thus, within our model, there exists preferences such that the comparative static remains, even with a geometric Brownian motion, that if any agents do not have the incentive to experiment, it is the lowest performing agents only. This, unfortunately, is not a complete analysis of behavior, and we must leave that task for future work.