Management Science

Chasing a Moving Target: Exploitation and Exploration in Dynamic Environments

Hart E. Posen, Daniel A. Levinthal,

Please scroll down for article—it is on subsequent pages

Chasing a Moving Target: Exploitation and Exploration in Dynamic Environments

Hart E. Posen
Ross School of Business, University of Michigan, Ann Arbor, Michigan 48109, hposen@umich.edu

Daniel A. Levinthal
The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104, levinthal@wharton.upenn.edu

A common justification for organizational change is that the circumstances in which the organization finds itself have changed, thereby eroding the value of utilizing existing knowledge. On the surface, the claim that organizations should adapt by generating new knowledge seems obvious and compelling. However, this standard wisdom overlooks the possibility that the reward to generating new knowledge may itself be eroded if change is an ongoing property of the environment. This observation in turn suggests that environmental change is not a self-evident call for strategies of greater exploration. Indeed, under some conditions the appropriate response to environmental change is a renewed focus on exploiting existing knowledge and opportunities. We develop a computational model based on the canonical multiarmed bandit formulation of exploration and exploitation. We endeavor to understand the mechanisms by which environmental change acts to make purposeful efforts at organizational adaptation less (or more) valuable.

1. Introduction

A common justification for organizational change is that the circumstances in which the organization finds itself have changed. This idea is broadly reflected in the contemporary literature's focus on, for example, how the pace of environmental change accentuates the need for exploratory search (see, for instance, March 1991) and the importance and desirability of dynamic capabilities (see, for instance, Teece et al. 1997). Claims about the value of organizational action aimed at adaptation in the face of environmental change have a taken-for-granted quality and, on the surface, seem obvious and compelling. Yet as we will argue, change in the external environment does not necessarily imply the need for, or benefit of, expanded efforts at organizational adaptation. Indeed, under some conditions the appropriate response to environmental change is a renewed focus on exploiting existing knowledge and opportunities.

In examining the merits of organizational adaptation, our work is closely related to March (1991), who identifies the need to allocate limited resources across both the exploitation of the known and exploration of the novel as a central strategic trade-off. If balancing exploration and exploitation is central to firm performance, as the consensus now suggests (Gupta et al. 2006), then what factors affect the appropriate balance? To answer this question, March (1991) focuses on the role of environmental change, reflecting a long tradition of interest in the external environment and the ability and willingness of organizations to adapt (see, for instance, Benner 2009, Hannan and Freeman 1984, Henderson 1993, Lant and Mezias 1992, Sorenson and Stuart 2000, Tushman and Romanelli 1985). March (1991, p. 80) concludes that "exogenous environmental change makes adaptation essential," but he then tempers this conclusion by noting that it also "makes learning from experience difficult."

The broader strategy literature is generally less circumspect. In recognizing the importance of Schumpeterian dynamics, the literature increasingly focuses attention on dynamic capabilities (Teece et al. 1997), high-velocity markets (Brown and Eisenhardt 1997), and hypercompetition (D'Aveni and Gunther 1994). Although in substance this work is nuanced in its treatment of strategic responses to environmental change, in practice it is often interpreted as suggesting a simple positive relationship between the extent of environmental change and the merits of organizational adaptation. This idea is diffused in the conventional wisdom of the practitioner literature, and managerial sentiment seems to reflect Peters and Waterman's (1982) call for a "bias for action." Although we agree with the Schumpeterian focus of much recent research, our analysis suggests that an exhortation to action in dynamic environments

is *not always* appropriate. As such, we endeavor to understand the mechanisms by which environmental change makes purposeful efforts at organizational adaptation less (or more) valuable.

We anchor our formal development on the multiarmed bandit model. This model is the canonical representation of the exploration–exploitation problem across literatures as diverse as economics and statistics (Berry and Fristedt 1985, Gittins 1979, Robbins 1952) and computer science (Holland 1975, Sutton and Barto 1998). In the management literature, March (1991) originally formulated his analysis in terms of a specific form of a genetic algorithm;[1] subsequently, he formulated his discussion of learning in terms of the bandit model (Denrell and March 2001; March 1996, 2003, 2010).

Consider the bandit model in the context of a simple organization, where management consists of making choices under uncertainty. In each period, the organization must choose from a set of policy alternatives; it may choose to continue with the policy choice made in the prior period, or it may select a different policy choice. The payoff for choosing a particular alternative reflects a draw from a probability distribution with an unknown, alternative-specific, mean. In pursuing its activity, the organization seeks to maximize the flow of returns over time. The organization makes its choice in each period based on its beliefs about the expected returns to each of the alternatives. If these subjective beliefs were known to be objectively correct, then the organization would choose to exploit the alternative known to have the highest expected return. In the absence of such (perfect) knowledge, the organization employs a strategy that allocates effort between (a) exploiting the alternative currently believed to be superior and (b) exploring alternatives that currently seem less promising in hope of identifying a superior alternative to the current preferred alternative.

We examine this bandit model of organizational choice and focus on the implications of environmental turbulence that alters the returns to particular policy alternatives. Through our model, we highlight two main observations. First, the appropriate response to environmental change is not necessarily a strategic shift of effort toward exploration—indeed, a shift toward exploitation is sometimes superior.[2] Environmental change is a two-sided "blade" with respect

to the benefits of organizational adaptation. Scholars have tended to focus on the diminution of the firm's current knowledge stock and the value of existing practices, when change renders "a prior strategic orientation...no longer effective" (Tushman and Romanelli 1985, p. 180). Yet there is another, less thoroughly considered, side of the blade. It reflects the possibility that recurring, though not necessarily predictable, change erodes the future benefits of additional knowledge gained via exploration, suggesting a more circumspect conclusion as to the merits of organizational adaptation in dynamic environments.

Second, organizations are biased toward action in the face of environmental change, even in the absence of a strategic shift toward exploration. This action bias arises endogenously in the process of experiential learning where choices are based on beliefs about the merits of the alternatives. An organization exploits a particular policy alternative when it is believed to be markedly superior to others. Environmental change naturally undermines existing beliefs. Alternatives that were believed to be rather good are likely to seem worse. Moreover, there is the possibility that less-examined alternatives may have improved. This compression of beliefs leads to an action bias in dynamic environments.

This paper proceeds as follows. In §2, we begin by discussing the theoretical background of our model. Then we describe the multiarmed bandit computational model. In §3, we present the results of three simulation experiments. The first experiment examines the baseline properties of the model by focusing on a simple static environment. The second experiment adds turbulence to the baseline experiment. In doing so, we examine the mechanisms that underlie the action bias and the appropriateness of purposeful efforts at adaptation in the face of change. The third experiment examines the implications of nonneutral turbulence that not only changes the relative merits of alternatives but also makes the average better or worse. In §4, we discuss the implications of these results for theory and practice.

2. Model
The multiarmed bandit model takes its name from a slot machine analogy. In the analogy, a gambler sits in front of a slot machine with N arms, each having different underlying payoff probabilities that are unknown to the player. The gambler's objective is to maximize the returns to playing the machine. This model has been the subject of significant study because its underlying structure closely resembles a variety of realistic economic situations ranging from research settings, such as research and development (Hardwick and Stout 1992), to strategic issues, such

---

[1] The model in March (1991) is a specific form of a genetic algorithm in the sense that individuals do not recombine attributes of other individuals directly, but rather they do so indirectly, mediated by the "organizational code."

[2] Keller and Rady (1999) developed a related result of a monopolist responding to a demand curve that shifts between two states, showing that the monopolist may engage in less experimentation when the frequency of the shift between the two states increases.

as product pricing (Bergemann and Välimäki 1996), to consumer choice (Gans et al. 2007). Common to all of these settings is that information about the returns to an alternative can only be gathered by selecting it. As such, the primary feature of the bandit model is the process of search and learning (Denrell and March 2001, March 2010). As a consequence, as Holland (1975, p. 77) notes, in the bandit model "the trade-off between gathering more information and exploiting it appears in the simplest terms."

Formalizing the discussion above, the bandit model reflects a sequential choice problem where, at each point in time, t, an organization must choose between N alternatives. The realized outcome of a particular choice is a draw from an unknown (alternative specific) probability distribution. If the process is Bernoulli, then the choice of alternative i results in an outcome of $\sigma = (1, 0)$—with $\sigma = 1$ reflecting a positive outcome with probability p_i and $\sigma = 0$ reflecting a negative outcome with probability $1 - p_i$. As such, the state of the environment can be described by the vector of payoffs to the alternatives, $P_t = [p_{1,t}, \ldots, p_{N,t}]$. In the typical bandit model formulation, the environment is stable across time such that for all t, $P_t = P$. We relax this assumption to allow for environmental turbulence.[3] Consistent with the definition of turbulence as reflecting both the frequency and unpredictability of environmental change (Baum and Wally 2003, Dess and Beard 1984, Miller 1987, Wholey and Brittain 1989), we model the level of turbulence as the probability of a shock to the payoff values of alternatives.

We assume that an organization is a boundedly rational actor engaged in a process of search and learning (Denrell and March 2001, March 2010). Choice behavior in the bandit model is driven by the relationship between two key organizational constructs: beliefs and strategies. We discuss each in turn below.

In the bandit model, the first key feature of organizational behavior is that an organization is assumed to make its choice (select an alternative) based on its subjective assessment of the merits of the alternatives. In employing beliefs to facilitate choice, the model portrays a boundedly rational organization as possessing a mental model or cognitive map that encapsulates its understanding of the merits of the available set of choices (Barr et al. 1992, Gavetti and Levinthal 2000, Nelson 2008, Weick 1969). In particular, an organization's belief about the merits of a given alternative is $q_{i,t}$, indexed by alternative i at time t, such that

its set of beliefs across alternatives is characterized by $Q_t = [q_{1,t}, \ldots, q_{N,t}]$ with $0 \le q_{i,t} \le 1$.

Two dimensions of beliefs are particularly relevant. First, subjective beliefs may (or may not) have objective validity. To the extent that beliefs accurately characterize the state of the world and reflect the true underlying value of the alternatives, then they embody knowledge of the choice opportunities that the organization faces. Thus, we characterize knowledge based on the sum of squared errors in beliefs relative to the true alternative payoffs as $1 - \sum_1^N (p_i - q_i)^2$. Second, beliefs may embody strong or weak opinions. The strength of opinions reflects the extent to which an organization believes that some alternatives are strongly superior to others. We characterize this as the variance across beliefs, $\sum_1^N (\bar{q} - q_i)^2$.

To refine its beliefs and maximize the value of its stream of rewards, an organization engages in learning. In the initial period $t = 0$, the organization has prior beliefs, $Q_0 = [q_{1,0}, \ldots, q_{N,0}]$, about the value of the N reward outcome probabilities $P = [p_1, \ldots, p_N]$. In each subsequent period, the organization makes a choice from the set of alternatives. By acting—making a choice of alternative i—a firm receives feedback from the environment in the form of an outcome $\sigma_{i,t} = (1, 0)$ associated with a reward of $r_{i,t} = (r^+, r^-)$. This reward may, for example, take the form of sales and profits (or lack thereof), which augments the firm's stock of assets such that $S_t = \sum_1^t r_t$.

Equally important, from the learning perspective, performance feedback provides new information about action-outcome linkages (the value of alternatives) that serve to update beliefs. In particular, beliefs are updated as the organization chooses one option i and observes the outcome, $\sigma_t = (1, 0)$. We assume that beliefs at any given point in time reflect the average reward over an organization's entire history of trials of a given alternative (March 1996, Sutton and Barto 1998). This simple average updating is a special case of the more general Bush and Mosteller (1955) updating methodology. In a given period, the firm updates its beliefs, $q_{i,t}$, of the selected alternative following a Bush–Mosteller fractional adjustment methodology where $q_{i,t+1} = q_{i,t} + \delta(\sigma_t - q_{i,t})$ and δ is the rate of updating. Average updating is the special case of $\delta = 1/(k_i + 1)$, where k_i is the number of times alternative i has been tried at time t. For alternatives that are not selected in a given period, $q_{i,t+1} = q_{i,t}$.

The second key feature of organizational behavior in the bandit model is that an organization is assumed to possess a search strategy. The organization would like to select, in each period, the alternative with the highest expected value, $\max(p_1, \ldots, p_N)$. However, the expected payoffs are unknown to the organization. Thus, conditional on a given set of beliefs

---

[3] A static environment is typically assumed in the economics literature because of the challenges associated with obtaining closed form solutions. The rare exception here is Whittle's (1988) "restless bandit" model.

about the relative merits of the alternatives, an organization requires a mechanism for choosing among the alternatives. An organization's search strategy can be conceived of as a rule or routine (Cyert and March 1963, Nelson and Winter 1982) that maps the organization's beliefs about the merits of alternatives to specific choices from the set of alternatives.

Under an appropriate set of assumptions, Gittins and Jones (1974) demonstrate that an optimal strategy exists. The basic property of this rational choice strategy in the bandit model is as follows. Consider a slot machine with two arms; arm 1 has an unknown payoff probability, and arm 2 has a known payoff probability (or a fixed payoff). The rational choice solution involves identifying the optimal stopping time—the number of trials to sample arm 1 before deciding to select one or the other arm in perpetuity. In the context of a multiarm bandit (with N unknown arms), the solution is not a stopping rule; rather, it is a dynamic allocation index (Gittins index) that defines the optimal arm choice in any given period (Gittins 1979).

Although the optimal strategy described by the Gittins index is sufficiently computationally intensive such that it is intractable except under a set of restrictive assumptions (Gans et al. 2007), it is worth pointing out its central characteristic. The "index for an arm is equal to the expected immediate reward from the arm, with an upward adjustment reflecting any uncertainty about the prospects of obtaining rewards from the arm, and the possibilities of resolving those uncertainties by selecting that arm" (Gittins and Wang 1992, p. 1625). The index suggests a pattern of choice over time that is sensitive to both the mean of beliefs and the uncertainty surrounding that mean.

Behavioral approaches that seek to identify an organizationally plausible, although not necessarily optimal, strategy date back to Schmalensee (1975) and Radner and Rothschild (1975). Although there is a wide variety of plausible strategies, perhaps the simplest and best-known strategy is that of selecting, at each point in time, the alternative with the highest belief, $\max(q_1, \ldots, q_N)$, reflecting the highest expected reward (Auer et al. 2002, Sutton and Barto 1998). This rule is referred to as greedy because at each point in time it myopically maximizes the expected reward in the next period. The greedy rule always exploits current knowledge; it never explores alternatives that are apparently inferior but may in fact be superior.

As Brezzi and Lai (2002, p. 105) note in their discussion of behavioral approximations to the Gittins index, a good rule "introduces adjustments into the myopic (greedy) rule so that some active experimentation is used to generate information about the unknown parameters." Thus, a simple variation on the greedy rule is to behave greedily a majority of

the time but with some small probability ε select an alternative at random, independent of beliefs (Sutton and Barto 1998). This rule is typically referred to as ε-greedy. However, the ε-greedy rule has an important undesirable feature. The rule is both overly sensitive to beliefs and completely insensitive to beliefs. When the organization exploits ($1 - \varepsilon$ percent of the time), it chooses the alternative with the highest belief, even if that differential to the second highest belief is trivially small. In contrast, when it explores (ε percent of the time), it is equally likely to select among all alternatives, independent of beliefs about the merits of a particular alternative. As such, in exploring, the ε-greedy rule is equally likely to pick the worst alternative as it is to pick the next-best alternative.

A well-defined and examined strategy that addresses the above deficiency is found in the softmax choice rule attributable to Luce (1959) and employed widely (Camerer and Ho 1999, Gans et al. 2007, Sutton and Barto 1998, Vermorel and Mohri 2005, Weber et al. 2004). This strategy formulation takes the form of random choice based on a Gibbs (Boltzmann) distribution. Here, the probability of selecting alternative i, m_i, is defined as

$$m_i = \frac{e^{(q_i/(\tau/10))}}{\sum_{i=1}^{N} e^{(q_i/(\tau/10))}}, \tag{1}$$

where q_i is the belief regarding alternative i, and τ is a parameter reflecting the organization's strategy.[4] A strategy, τ, maps the organization's beliefs $Q = [q_1, \ldots, q_N]$ to choice probabilities $M = [m_1, \ldots, m_N]$ where $\sum_{i=1}^{N} m_i = 1$. When $\tau \to 0$ even small differences in beliefs are weighted heavily. Thus, if one alternative is viewed even marginally more favorably than are the others, it will be selected with near certainty, and thus we observe a very greedy strategy (highly exploitative). For example, consider a three-armed bandit with beliefs across arms of $Q = [0.7, 0.8, 0.7]$. A purely exploitive strategy would map this to a set of choice probabilities of $M = [0.0, 1.0, 0.0]$ so that the firm chooses with probability one the alternative with the highest beliefs. In contrast, when $\tau \to \infty$, differences in beliefs are shrunk to the point that the organization is indifferent between alternatives (even if it believes that one alternative is far superior). This leads to exploration as a random choice among all alternatives, where each alternative is chosen with probability $1/N$. Continuing the example, at very high τ, the beliefs $Q = [0.7, 0.8, 0.7]$ would map to the choice probabilities $M = [0.33, 0.33, 0.33]$. In such a setting, the organization would be insensitive to the fact that its current beliefs suggest that

---

[4] Note that the "/10" in Equation (1) acts simply as a scaling factor.

alternative two is superior and with probability 0.66 would choose an alternative other than the one currently believed best. At moderate levels of τ, alternatives that appear to be good (high beliefs) are more likely to be sampled. At intermediate levels of τ, alternatives that appear to be good (high beliefs) are more likely to be selected.[5]

Before proceeding with the analysis, we wish to highlight the relationship between the bandit model and that of the genetic algorithm model in March (1991). We focus on two key features of strategy in these models: (a) the mechanism governing organizational choice and (b) the meaning of exploration. First, March's (1991) genetic algorithm and the bandit model are fundamentally concerned with the trade-off between exploitation of existing knowledge and exploring for new knowledge in the process of organizational learning. Nevertheless, the level of analysis is different. March (1991) generates differences in the balance between exploration and exploitation by altering internal characteristics of the organization. The strategy features of his model are tunable parameters reflecting turnover and learning rate at the level of the individual. In our bandit model, a strategy τ encapsulates these internal organizational dynamics in a single parameter reflecting the sensitivity of organizations to their beliefs. Yet regardless of the level at which the tunable strategy parameter is implemented, the implications for tuning the explore-exploit trade-off remain consistent.

Second, March (1991) argues that exploration not only suggests a willingness to forgo the opportunity to engage in an action that one believes superior, but also to privilege alternatives with less certain beliefs. A search strategy in the bandit model implies a similar understanding of the notion of exploration because choice must focus on the mean value of an alternative with "an upward adjustment reflecting any uncertainty" (Gittins and Wang 1992, p. 1625). It is less obvious that the softmax formulation of strategy in our bandit model, although explicitly only incorporating the first moment of beliefs (q_i), endogenously weights the mean belief by uncertainty. We discuss the characteristic of our bandit model in the analysis of the baseline characteristics of the model in the subsequent section.

3. Analysis

We use the multiarmed bandit model described above as an analytical tool to study the efficacy of alternative adaptation strategies under turbulence. The structure of the simulation requires initializing two sets of variables associated with (a) the environment and turbulence rate and (b) organizational strategies and beliefs.

The opportunity structure of the environment is defined by initializing the payoff probabilities for each alternative. We formulate a 10-arm bandit model such that the vector of alternatives' payoff probabilities is $P_t = [p_{1,t}, \ldots, p_{10,t}]$. The choice of a particular alternative i leads to an outcome, positive/negative such that $\sigma_{i,t} = (1, 0)$, with probabilities $(p_{i,t}, 1 - p_{i,t})$ and an associated reward of $r_{i,t} = (1, -1)$. Each alternative is allocated an initial payoff probability, $p_{i,0}$, that is taken to be a draw from a beta distribution ($\alpha = 2$ and $\beta = 2$). This produces a distribution of payoff probabilities (across alternatives) that is symmetrical with mean of 0.50 and standard deviation of 0.22. The distribution resembles in shape a normal distribution but has a distribution over probability values that it is bounded between zero and one.[6]

Models incorporating turbulence employ a probabilistic shock in a given period where the probability of a shock, η, ranges from 0.0 to 0.32. When $\eta = 0$, the environment is stable, and $P_t = P$. For $\eta > 0$, conditional on the occurrence of a shock, the payoff to each of the alternatives is reset with probability 0.5. To reset an alternative, we redraw the payoff, $p_{i,t}$, from the same initial beta distribution ($\alpha = 2$ and $\beta = 2$). As a consequence, such turbulence is neutral in the sense that the expected value of the set of alternatives remains unchanged after a shock, whereas the payoff to individual alternatives change.

All simulations are seeded with 25,000 organizations, each facing a unique environment with alternatives drawn from a beta distribution ($\alpha = 2$ and $\beta = 2$). We set the initial asset stock to zero and run the simulation for 500 periods. We specify initial beliefs across these alternatives, $Q_0 = [q_{1,0}, \ldots, q_{10,0}]$, to be homogeneous and set the value to 0.5, which is equal to the mean value of the actual distribution of payoffs. The organizations are divided equally among five different strategies. The strategies, τ, reflect different balances between exploration and exploitation. The strategies implemented, which range from pure exploitation to high exploration, are $\tau = (0.0, 0.25, 0.5, 0.75, 1.0)$.[7]

---

[5] Organizations are assumed to be risk neutral. They select an alternative solely based on the expected value of an alternative. Although per March (1996) and Le Mens and Denrell (2011), risk aversion emerges endogenously in a model of experiential learning.

[6] We examined alternative initial arm distributions. In particular, we reran the models in this paper with arms drawn from a uniform distribution. The uniform is a special case of the beta distribution with ($\alpha = 1$ and $\beta = 1$). The results are qualitatively the same.

[7] Equation (1) does not permit the use of a value of zero for the strategy. As τ approaches zero, the value of Equation (1) approaches infinity. Therefore, we set as a pure exploitation strategy the computationally feasible value of $\tau = 0.02$. This value is sufficiently close to zero that the strategy reflects pure exploitation in the sense that the alternative with the highest expected payoff is chosen in all runs. As such, the figures and discussion in the remainder of the paper simply refer to this as $\tau = 0$.

The analysis proceeds over three experiments in which we examine the performance implications of alternative strategies of balancing exploration and exploitation across different environmental contexts: (1) stable world, (2) turbulent world (mean neutral), and (3) turbulent world (nonneutral: mean devaluing or mean enhancing).

3.1. Experiment 1: Adaptation Strategies in a Stable World

We begin our analysis by examining the bandit model in a stable environment. Our objective is to examine the basic properties of the model and to demonstrate that it generates results that are consistent with the prior literature, in terms of both the March (1991) model and also the broader bandit literature.
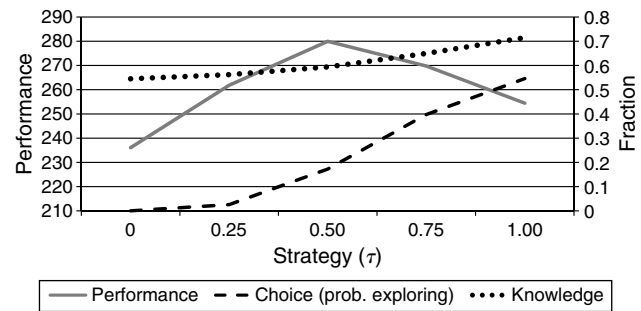
In Figure 1, we examine the implications of altering adaptation strategy, τ, across three measures: (a) exploratory choice, (b) performance, and (c) knowledge generation. Each is measured at the end of the simulation. First, the dashed line (right axis) represents the extent to which the choices taken by organizations are exploratory, measured as the probability of choosing a different alternative in period $t+1$ than was chosen in period t. As τ increases, the level of exploration increases. At the most exploratory level of search strategy in our model, $\tau = 1$, organizations select a new alternative roughly every second period.[8]

Second, the solid line (left axis) represents performance measured as the accumulated asset stock. The graph highlights that the best strategy, which displays a moderate level of exploration ($\tau = 0.5$), accumulates on average 280 units, which reflects an average frequency of positive payoffs across periods of 78%. Third, the dotted line (right axis) reflects the accuracy of knowledge embodied in firms' subjective beliefs about the value of alternatives at the end of the final period. We measure knowledge as one minus the sum of squared error of beliefs relative to the true arm payoffs.[9] The graph highlights that the level of knowledge obtained is increasing with the extent of exploration, from 0.55 at $\tau = 0$ to 0.72 at $\tau = 1$.

---

[8] We also examined alternative metrics of exploration. For example, the probability of not choosing the arm that is believed to offer the highest payoff. The qualitative results on exploration behavior do not change with this, or other, alternative measures considered.

[9] This measure of knowledge can be interpreted as reflecting an explicit comparison to the level of knowledge that would be obtained by a firm with beliefs chosen randomly. Knowledge is zero if beliefs are set randomly from a beta $(2, 2)$ distribution matching reality, and knowledge is one if beliefs perfectly reflect reality such that SSE $= 0$. Thus, this measure of knowledge has the attractive property that knowledge is greater than zero when the organization has knowledge above that of an organization with random beliefs. Organizations pursuing a search strategy that embodies a moderate degree of adaptiveness, $\tau = 0.5$, accumulate, on average, knowledge over the 500 periods of the simulation of 0.59.

Figure 1 Choice, Performance, and Knowledge Across Strategies in Stable Environment
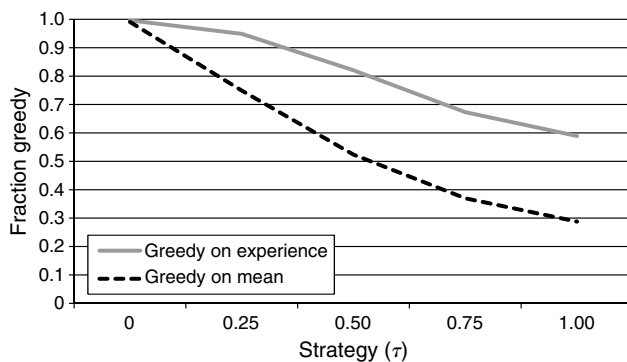

Taken together, the inverse U-shaped performance curve across strategies and the near linear increase in knowledge demonstrates the trade-off between generating new knowledge and leveraging existing knowledge. At low levels of exploration, little knowledge is attained and there is little knowledge to exploit. In contrast, at high levels of exploration, significant knowledge is attained, but it is never exploited to enhance performance. Thus, in this static environment, the results reflect March's (1991, p. 71) conclusion that organizations that "engage in exploration to the exclusion of exploitation are likely to find that they suffer the costs of experimentation without gaining many of its benefits," whereas organizations that "engage in exploitation to the exclusion of exploration are likely to find themselves trapped in suboptimal stable equilibrium."

Although our model replicates the main results of the March (1991) model, we demonstrate that that it also replicates key results in the bandit literature. We focus on two features of the model: (a) identifying the extent to which exploration in our model embodies choice based on uncertainty in beliefs and (b) docking our results to a bandit model with an ε-greedy strategy.

Gittins and Wang (1992, p. 1625) note that in selecting an alternative, it is appropriate to consider the "expected immediate reward from the arm, with an upward adjustment reflecting any uncertainty about the prospects of obtaining rewards from the arm, and the possibilities of resolving those uncertainties by selecting that arm." This intuition, which arises from the Gittins index, is consistent with the notion of exploration in March (1991) and the broader management literature. Nevertheless, although the softmax choice strategy we employ considers only the mean of beliefs, neglecting direct consideration of uncertainty, we demonstrate that it endogenously weights the level of uncertainty in its choice behavior.

In the bandit literature, strategy is defined as greedy if, in each period, it chooses the alternative

with the highest mean belief. Analogously, we construct a measure of greediness based on experience with an alternative (as a proxy for uncertainty) rather than on the first moment of beliefs. Greedy on experience measures the probability that an organization will choose an alternative with which it has the most experience. The results of these measures, at period 500, are presented in Figure 2. As exploration strategy, τ, increases, organizations become less greedy, on the basis both of mean and of uncertainty in beliefs. As such, organizations are more likely to explore alternatives that are presently believed to be inferior and for which the beliefs are relatively imprecise.

The intuition behind this result rests on the observation that alternatives that appear to be good are more likely to be sampled in the process of experiential learning (March 1996, Denrell and March 2001). Moreover, the more an alternative is sampled, the more certain the estimate of its mean value. As a result, there is a positive correlation between the mean belief regarding the payoff to a given alternative and the level of experience with that alternative. When an organization chooses a low mean belief alternative, it is also likely to select an alternative that is known with less certainty. Exploration strategies (high τ) are not only more likely to forgo alternatives with high mean belief expectations, they are also likely to forgo more certain alternatives in favor of less certain alternatives.

For the purposes of model docking, we rerun our simulation employing a simple ε-greedy strategy, rather than our softmax strategy. This enables comparisons across the two choice mechanisms. We expect, and observe, that as $\tau \to 0$, the performance results converge to the greedy strategy with $\varepsilon = 0$. Moreover, as expected from the above discussion of the characteristics of the softmax choice rule, it performs somewhat better than the ε-greedy choice rule. Additionally, Sutton and Barto (1998, p. 41) examine an $\varepsilon = 0.1$ greedy strategy on a 10-arm Gaussian (rather than Bernoulli) bandit model and measure the percentage of optimal action. This measure reflects

the probability that the alternative chosen in a given period is the alternative with the highest payoff probability. We examine this metric at period 500. In our implementation, both the $\tau = 0$ and $\varepsilon = 0$ strategies exhibit the same optimal action result of 39%. With $\varepsilon = 0.1$, our model exhibits 63% optimal action, consistent with the Sutton and Barto (1998) result. Thus, the softmax strategy of our bandit model docks to the ε-greedy variant of the bandit, and our ε-greedy implementation docks to that of Sutton and Barto (1998).

In sum, in a static environment, the properties of the bandit model are consistent with March's (1991) claim that superior performance arises from a strategy that balances exploration and exploitation.
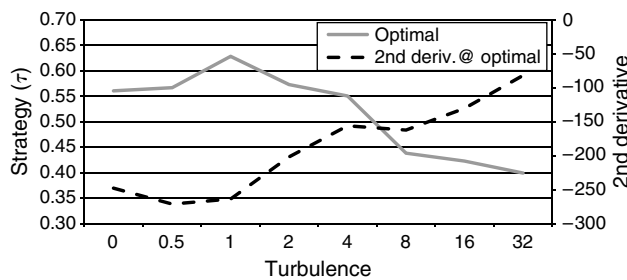
3.2. Experiment 2: Adaptation in a Changing World

In the second experiment, we add turbulence to the model, at levels ranging from very infrequent to near continuous (0–32%).[10] We examine how and why turbulence alters the optimal level of adaptiveness embodied in a search strategy.

The main results of the turbulence model are depicted in Figure 3. This figure plots the optimal search strategy calculated by fitting a third order polynomial to the results (that is, Figure 1 replicated across levels of turbulence) and extracting the strategy, τ, associated with that peak performance (maximum stock of accumulated assets). The zero turbulence result reflects the conclusions from Experiment 1: Peak performance occurs at a search strategy of $\tau = 0.56$, reflecting a moderate degree of adaptiveness. As the probabilistic frequency of turbulence increases, the optimal level of exploration takes an inverse U-shaped form. The curve is highly skewed in the sense that over most of the range of turbulence examined, the optimal degree of adaptiveness embodied in a search strategy is decreasing with increasing frequency of turbulence.

In addition to altering the optimal level of exploration strategy, turbulence may also alter the extent of the cost associated with failing to match the optimal exploration strategy. To examine this issue, we calculate the flatness of the performance versus τ curves by plotting the second derivative of performance at the optimal τ value (see Figure 3). If the curve is sufficiently flat, because little or no knowledge can be accumulated via experiential learning, then the distinction between exploration and exploitation becomes less relevant. Over the interesting range of turbulence (0–8%), the curve is significantly concave. At very low levels of turbulence (below 0.5%),

---

[10] At higher levels of turbulence, performance becomes invariant to exploration strategy.

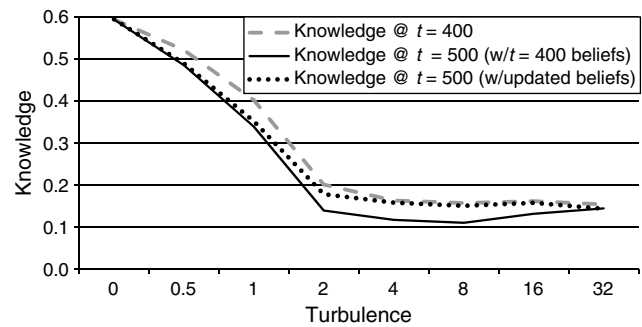**Figure 3   Optimal Exploration Strategy (τ) Across Turbulence Levels**

Figure 4   Knowledge Devaluation Across Turbulence Levels

Note. Holding strategy fixed ($\tau = 0.5$).

the second derivative of performance at the optimal τ value is increasingly negative, implying the performance versus τ curve is increasingly concave at the optimal τ. This suggests that increasing turbulence enhances the rewards to choosing the optimal τ. As turbulence increases further (at levels above 0.5%), the second derivative starts to increase (becomes less negative), implying that the performance versus τ curve at the optimal τ is becoming flatter. This implies that identifying the optimal strategy becomes less important for performance at higher levels of turbulence.[11]

Because the discussion of turbulence and exploration in March (1991) is so prominent in the literature, it is worthwhile to directly compare these results to his. In contrast to March (1991), the results in Figure 3 suggest a contingency in the relationship between turbulence and the level of exploration. However, March (1991) examines only a single, relatively modest level of turbulence (2%). His results, like our own, show that organizational performance is improved by increasing the level of exploration as turbulence increases from 0 to 2%. Like March (1991), we find that under conditions of very infrequent environmental change, a strategy stressing adaptation via exploration and new knowledge accretion is optimal; but as change becomes more frequent, the optimal response is, in fact, one of greater inertia.

We engage in additional analysis to understand the mechanisms driving this response to turbulence. First, we focus on the tension between the obsolescence of existing knowledge and the accretion of new knowledge. Second, we focus on the impact of turbulence on the organizations' strength of opinions surrounding its beliefs about the relative merits of alternatives. We examine each mechanism in turn below.

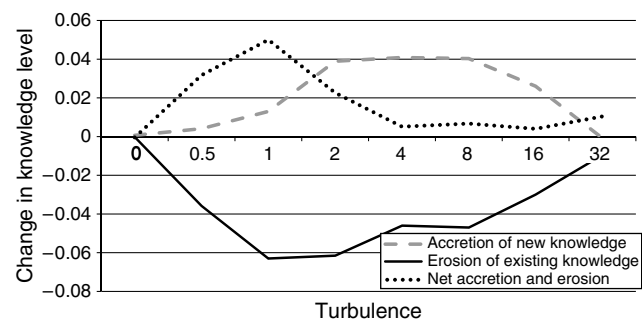Our working hypothesis, addressing the tension between knowledge obsolescence and accretion, is

that the characterization of preferred search strategies in Figure 3 arises because turbulence erodes knowledge asymmetrically. That is, turbulence acts to erode performance via two knowledge mechanisms. First, it reduces the future value of existing knowledge because it changes the underlying state of the world. Second, turbulence reduces the value of efforts to generate new knowledge because the lifespan of returns to new knowledge is reduced in a world in which change is more frequent. We examine the possibility of an asymmetry in the effect of turbulence on these two mechanisms in Figures 4 and 5.

We engage in a differences-in-differences analysis of knowledge levels. Holding beliefs constant at period 400 levels, we examine the resulting level of knowledge erosion over the subsequent 100 periods. We then compare the period 400 knowledge results to the period 500 knowledge results (assuming constant beliefs), and the period 500 results assuming learning at a strategy of moderate adaptation, $\tau = 0.5$. The dashed line in Figure 4 reflects the level of knowledge at period 400, across levels of turbulence. In the absence of turbulence, the knowledge level is 0.59 (as per the analysis in Experiment 1). By the time turbulence reaches 8%, knowledge is reduced to 0.44. The solid line presents the results of using period 400 beliefs through period 500 (that is, period 400 beliefs are used but not updated). As such, the difference

Figure 5   Knowledge Accretion and Erosion Across Turbulence Levels

Note. Holding strategy fixed ($\tau = 0.5$).

[11] Even a flat curve performance versus τ curve does not imply that an organization should be indifferent between the levels of exploration strategy. In our model, exploration and exploitation have no explicit cost (only opportunity costs). If, in contrast to our model, exploration has a direct cost that is higher than exploitation, then erring on the side of too little exploration is superior to erring on the side of too much exploration under high levels of turbulence.

between the solid and dashed lines reflects the erosion of existing knowledge over the 100-period time span. Finally, the dotted line reflects knowledge levels at period 500 if learning (updating of beliefs) continued through to period 500. The difference between the solid and dotted lines reflects the improved knowledge associated with efforts at new knowledge accretion over a 100-period time span.

In Figure 5, the solid and dashed lines reflect, respectively, (a) existing knowledge that is eroded by turbulence (the difference between the dashed and solid lines in Figure 4) and (b) new knowledge accretion (the difference between the solid and dotted lines in Figure 4). The dotted line in Figure 5 depicts the net knowledge change.

Over the range from no turbulence through a frequency of 1%, existing knowledge is eroded rapidly. The more rapidly existing knowledge is eroded, other things being equal, the greater the impetus to explore. The alternative to employing additional efforts at exploration is using existing knowledge. As such, the erosion of existing knowledge defines the opportunity cost of exploring—which is decreasing rapidly over this range. At the same time, additional exploration is generating new knowledge at an increasing rate. With decreasing opportunity cost and increasing generation of new knowledge, there is a substantial net knowledge benefit to exploring (positive slope of the dotted line), and as such the appropriate response is one of increased exploration at low levels of turbulence.

Over the range of turbulence from 1% to 2%, the existing knowledge erosion level is flattening and, indeed, slightly decreasing. Relative to this low turbulence range, this reflects a substantial decrease in the *rate* at which opportunity cost is decreasing. At the same time, the rate of new knowledge accretion is increasing. However, the increase in the rate of new knowledge accretion is substantially less than the decrease in the rate of existing knowledge erosion (and increase in opportunity cost). Thus, the appropriate response, as reflected by the negative slope of the net knowledge change curve, is one of decreased exploration over the 1%–2% turbulence levels.

Over the range of turbulence in excess of 2%, existing knowledge continues to erode less strongly (opportunity cost is increasing), and new knowledge accretion levels flatten, then decline. As a result, the appropriate response at higher levels of turbulence is one of further decreases in exploration. At the very highest levels of turbulence, there is little knowledge available to erode, and very limited potential for new knowledge accretion.

In sum, the appropriate adaptation strategy depends on the effect of turbulence on both the payoff
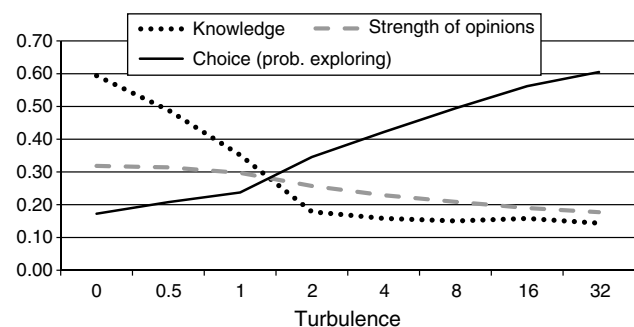
to utilizing current knowledge and the payoff to generating new knowledge. The value of current knowledge reflects the possibilities of continuing to use existing knowledge into the future and thereby drives the opportunity costs associated with engaging in additional exploration. However, turbulence does not simply alter the value of existing knowledge. It also alters the payoff to efforts at gathering new knowledge. The effect of increasing turbulence depends crucially on the relative consequences of turbulence along these two dimensions.

The second mechanism underlying the results in this experiment is that turbulence endogenously gives rise to an action bias, even when the exploration strategy itself is fixed. In our discussion of organizations' beliefs, we have focused on the extent these subjective beliefs embody knowledge. In addition, there is another important dimension of beliefs—the strength of opinions. The strength of opinions reflects how strongly a firm believes one alternative is superior to other alternatives. We measure this as the variance across an organization's beliefs. At the start of the simulation, firms have homogeneous beliefs about the alternatives; they simply believe all alternatives have a 50% payoff probability and, thus, the variance in beliefs across alternatives is zero. However, organizations quickly start to form divergent beliefs about these alternatives.

In Figure 6, we plot, at period 500, the accuracy of knowledge, strength of opinions, and extent of exploration across the range of turbulence for the $\tau = 0.5$ strategy. The dotted line highlights the devaluation of knowledge with increasing turbulence. The dashed line plots the strength of opinions (variance across beliefs). The variance in beliefs across alternatives is reduced by nearly one-third as we move from a stable world to one in which turbulence is set at 8%. It appears turbulence acts not only to devalue knowledge but also to homogenize beliefs.

The reduction in the strength of opinions in the face of turbulence has important consequences for

Figure 6 Knowledge, Opinions, and Choices Across Turbulence Levels

Note. Holding strategy fixed ($\tau = 0.5$).

the nature of organizational choice. As the strength of opinions is reduced, organizations endogenously increase the extent to which their choices are exploratory—even holding search strategy constant. The mechanism driving this action bias exists because in the experiential learning process, in making choices, organizations are sensitive to their beliefs about the merits of the alternatives available to them. Organizations exploit a given policy choice when they believe it is markedly superior to others. Environmental change naturally undermines existing beliefs; alternatives that were believed to be rather good are likely to seem worse. Moreover, there is the possibility that less-examined alternatives may have gotten better. As the variance across alternatives in beliefs is reduced by turbulence, organizations have fewer bases on which to differentiate among alternatives. As a consequence of this compression in beliefs, organizations exhibit an action bias in dynamic environments.

In sum, turbulence acts to alter the optimal level of adaptation by altering the returns to utilizing existing knowledge and generating new knowledge. Yet as we demonstrate, turbulence does not necessarily give rise to a requisite shift toward a more exploratory strategy. Indeed, as the level of turbulence increases, the appropriate balance between exploitation and exploration shifts in the direction of exploitation. In addition, we note a second mechanism at work—turbulence tends to endogenously generate exploratory behavior even in the absence of an overt change in adaptation strategy. In the next experiment, we examine further implications of this action bias.

3.3. Experiment 3: Adaptation Strategies Under Nonneutral Turbulence

Turbulence alters the appropriate adaptation strategy because it alters organizational knowledge. In the implementation of environmental change in Experiment 2, turbulence reflects a shock to the payoffs to the alternatives, but the shock is neutral. That is, the preshock and postshock distribution of payoffs across alternatives is drawn from the same beta distribution, and as such the expected value of the set of alternatives is held constant at 0.5. In this experiment, we examine turbulence that is nonneutral in that it makes the world more (or less) munificent (increases or decreases the expected value of the set of alternatives). We hold turbulence at the level of a 5% probability of a shock and examine turbulence that changes the mean payoff values (cumulatively) over the range from $+/-20\%$ of the neutral turbulence level.[12]
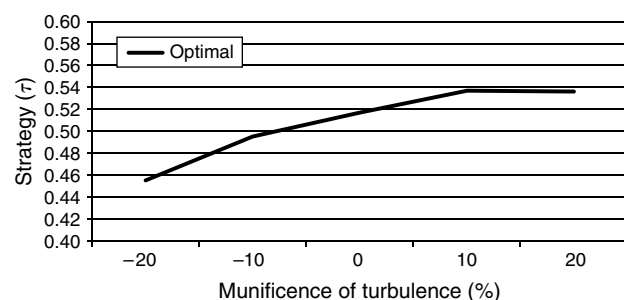
---

[12] We do so by shifting the parameters of the beta distribution. These shocks are cumulative in that a subsequent shock resets the payoff probability of an alternative from a progressively worse (better) distribution.

To illustrate munificence-altering turbulence, consider the emergence of (sufficiently fast) flash-based computer memory and the consequent changes in the digital data storage industry. This change reduced the payoff to producing magnetic spinning disks. It also enhanced the returns to a wide variety of applications that were, in the past, deemed to be inferior. These newly enhanced choices include not only solid-state disk drives as a direct replacement for spinning hard drives but also alternative applications of long-term storage outside the realm of the traditional PC (e.g., music player, smart phone, auto). If this turbulence had been neutral, then to the extent that the returns to selling spinning hard drives had been reduced, the expected returns to the alternatives would have been similarly enhanced. In this sense, neutral turbulence shuffles the deck but does not change the expected reward. Yet the turbulence in the long-term digital storage industry was nonneutral. The decrease in the expected returns to engaging in designing and selling spinning hard drives was more than offset by the enhanced returns to the many new uses for solid-state drives. As such, this turbulence was munificence enhancing because it not only reshuffled the deck, it also increased the expected return to the set of alternatives available to organizations in the industry.

Figure 7 displays the optimal search strategy across the range of the munificence of turbulence, holding the level of turbulence constant. Note that the zero percentage point reflects the neutral turbulence result from Experiment 2. The main result is that if turbulence increases the level of munificence, the optimal level of adaptiveness embodied in a search strategy increases. Because we hold the level of turbulence constant, the rates of erosion of existing knowledge and accretion of new knowledge is held fixed. As such, a different mechanism must be driving this result.
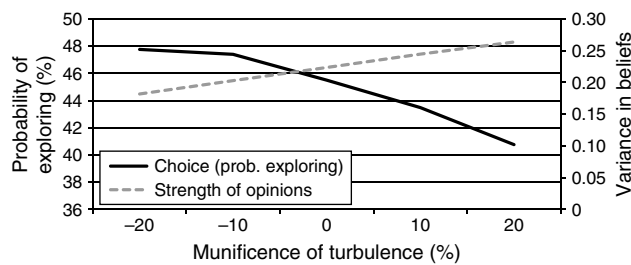
To identify the mechanism, we plot in Figure 8 the strength of organizations' opinions (dashed line) and the level of exploration (solid line) across the levels of the munificence of turbulence. With the search

Figure 7 Optimal Exploration Strategy (τ) Across Munificence of Turbulence

Note. Turbulence fixed at 5%.

Figure 8   Opinions and Choices: Across Munificence of Turbulence

Note. Turbulence fixed at 5% and strategy fixed at $\tau = 0.5$.

strategy held constant at $\tau = 0.5$, increasing munificence of turbulence increases the strength of opinions. This occurs because the payoff to all alternatives is, on average, increasing with munificence-enhancing turbulence. However, in the process of experiential learning, organizations are more likely to select an alternative that is known to be good (and as such the set of alternatives considered tends to be decreasing over time). This raises the possibility that they may fail to realize, on average, all alternatives have improved. Thus, in the experiential learning process, organizations may interpret the enhanced performance results as information about the quality of the alternative that they currently favor.

This increase in the strength of opinions under munificence-enhancing turbulence leads to a decrease in the extent of exploration (per the discussion in Experiment 2). But as turbulence, neutral or munificence enhancing, continues to devalue the returns to existing knowledge and to new knowledge accretion in the same manner, the appropriate level of τ has not changed. Thus, this endogenous decrease in exploration requires a countervailing change in adaptation strategy.

In the face of munificence-enhancing turbulence, organizations are subject to an "inaction bias." This bias is not driven by any inherent behavioral bias but rather by the nature of belief formation in the experiential learning process.[13] An increased effort at exploration (a more exploratory strategy) is needed to overcome the endogenous reduction in exploration that is associated with the mechanism described above. Indeed, the opposite is true for munificence-reducing turbulence, which tends to exaggerate the bias toward action.

In sum, this experiment serves to highlight the observation that turbulence devalues knowledge embodied in beliefs. This affects not only the return to

---

[13] This outcome is akin to a competence trap. However, instead of inertia resulting from accumulated competence with a particular alternative, it results from enhanced positive beliefs about the attractiveness of a given alternative in the absence of information about other alternatives.

an explicit strategic shift toward exploration, as discussed in Experiment 2, but also the nature of organizational action itself. When organizations are engaged in learning, and their choices are sensitive to their beliefs regarding the relative merits of alternatives, turbulence can engender organizational action (that is increasingly exploratory or exploitative) even in the absence of a shift in exploration strategy. As a result, the appropriate organizational response to turbulence is sensitive to the nature of turbulence—not only its severity but also the extent to which it enhances or degrades the value of possible organizational actions.

3.4. Sensitivity Analysis
We examine the sensitivity of our results to (a) alternative specifications of the key parameters of the bandit model, (b) episodic turbulence, and (c) selection. In each case, the results are robust. First, two key elements of the bandit model have been altered. We examine a bandit model in which belief updating is implemented following a Bayesian methodology rather than the simple average updating methodology employed in the discussion above. We also examine a model where the alternatives are drawn from a uniform distribution rather than the near-normal beta ($\alpha = 2$, $\beta = 2$) distribution.[14] Second, we examine a model in which turbulence is episodic. In particular, we examine turbulence that occurs in the middle third of the 500-period simulation across a wide range of severity. We also examine turbulence that occurs in a 50-period window, where the window occurs earlier or later in the simulation. Third, we examine a model in which there is selection. We examine noncompetitive selection in which an organization exits when its asset stock falls below zero. We also examine competitive selection, where an organization's probability of failing is a function of its performance relative to other organizations. In each of these additional experiments, the results are robust.

4.   Discussion
We formally consider the effect of changing environmental conditions on firm search strategy and, in particular, the extent to which an organization should pursue a strategy of adaptation under conditions of environmental change. On the surface, the claim that organizations should respond to changing circumstances by generating new knowledge seems obvious and compelling. This logic is reflected in a broader exhortation to action in both the academic and popular literatures—and its validity deserves examination. Is the exhortation to action in a dynamic environment

---

[14] Note that the uniform distribution is a special case of the beta distribution parameterized as $\alpha = 1$ and $\beta = 1$.

always appropriate? Our modeling efforts focus attention on the implications of turbulence for organizational beliefs about the relative merits of alternatives and suggest a more circumspect conclusion about the value of organizational adaptation.

Our model points to an inverse U-shaped relationship between the optimal degree of adaptiveness embodied in a search strategy and the frequency of turbulence. This result stems from two opposing mechanisms. On one hand, intuition suggests environmental change devalues existing knowledge. The failure to respond to change in the environment implies decreased performance. This mechanism is well known in the existing theoretical literature and underlies much of the wisdom regarding the appropriate response to change. On the other hand, less well developed and understood in the literature is a mechanism that acts in the opposite direction. Environmental change erodes the rewards to exploratory efforts at accumulating new knowledge. Our analysis suggests that the delicate balance between these two mechanisms, and the possibility of their asymmetrical response to turbulence, determines the appropriate organizational response to environmental change.

Environmental change has a second, arguably more subtle, implication. We conceive of strategies as reflecting managerial and organizational attempts to understand the world and act appropriately. This view is reflected in the bandit model where a strategy is a function that maps knowledge about the world onto particular choices. Thus, a search strategy does not determine a specific action but rather action conditional on beliefs about the state of the world. For a given search strategy, belief structures that are more highly skewed, that place greater weight on some alternatives and lesser weight on others, will generate choices that are more exploitive. By implication, factors that alter organizational beliefs may lead to changes in choice behavior even if the search strategy is fixed and unchanging.

When an organization is involved in a process of experiential learning, environmental change will tend not only to erode the extent of knowledge embodied in beliefs but also lead to more diffuse, less distinct beliefs about the relative merits of alternatives. Once-attractive alternatives are likely to become less appealing; as a result, once-positive beliefs about these initiatives will tend to be driven lower with repeated experience. By the same token, should the organization happen to explore alternatives it previously viewed as less attractive, it may find they have improved in the intervening periods. This homogenization of beliefs that arises in the experiential learning process under conditions of turbulence leads to an action bias.

Consider the case of adaptation to extreme turbulence in the auto industry. In the post-2007 period, oil prices rose markedly and the U.S. housing market collapsed. This led to rapid changes in consumer preferences; they shifted from large, heavy, all-wheel drive vehicles to smaller, lighter vehicles with better gas mileage.

The taken-for-granted strategy during this period of turbulence was that of significant action aimed at adaptation. Existing knowledge about designing and building large, heavy, and powerful vehicles was devalued. The response entailed altering the balance between exploitation and exploration by shifting resources toward the development of less well understood technologies (e.g., electric vehicles or light weight carbon fiber bodies), which were expected to be a better fit with the new environment.

This strategic response is reflected in the General Motors (GM) effort to bring the all-electric Volt to market in the post-2007 period. To allocate resources to the Volt, GM reduced efforts to exploit existing knowledge (cutting, at the margin, investment in mid-size fuel efficient vehicles; Welch 2008). The process of developing this highly exploratory technology was very slow not because of organizational impediments to change but rather because of the limited rate of new knowledge accretion that has been an ongoing feature of this technology domain (batteries in particular). By the time the Volt reached the market in 2010, oil prices had dropped to the point where the vehicle was no longer economically viable, and an ongoing recession made consumers highly price sensitive (Blanco 2011). Thus, ongoing turbulence had devalued not only GM's return to deploying existing knowledge and technologies (embodied in large sport utility vehicles) but also the near-term returns to its exploratory investment in an electric vehicle.[15]

Although the organizational impetus to act was strong, because this turbulence was seen as munificence reducing, the returns to exploring were devalued by ongoing turbulence. Yet there was an alternative course of action in the face of this environmental change. Ford, for example, chose to reduce its efforts at exploration, allocating a larger share of resources to exploitation. It implemented a strategy that focused more effort on refining existing technologies and designs, for example, deploying gas-electric hybrid technology in key models and modifying the highly successful European Fiesta for the U.S. market. As CNN reported in January 2009, "In the near term, the carmakers' [Ford's] so-called EcoBoost engines—turbocharged engines with highly sophisticated fuel

---

[15] Although subsequent increases in fuel prices may render GM's investments worthwhile, the key point is that the possibility of future turbulence reduces the expected returns from exploratory investment.

injection systems—will provide the greater fuel economy Americans want as gas prices rise gradually" (Valdes-Dapena 2009). Ford did not forgo exploration in its entirety, continuing an investment program in electric vehicle technologies; but whereas GM massively shifted resources toward exploration in the face of extreme turbulence, Ford shifted resources, at the margin, toward exploitation.

Our theory suggests that the appropriateness of such a strategy is a function of the extent to which existing knowledge is devalued by turbulence as well as the extent to which ongoing turbulence may devalue investments in exploration aimed at the generation of new knowledge. The challenge for organizations is the ex ante assessment of the magnitudes of these two competing forces. Organizations should be aware of the bias to act and view the decision to shift the exploit or explore balance toward exploration as a strategic choice rather than a strategic imperative.

Our theory, that there are conditions under which the appropriate response to environmental change is to heighten the degree of inertia, is broadly consistent with arguments within the population ecology literature that question the plasticity of organizations. But the mechanism in our theory is quite different. Population ecology holds that organizations are subject to inertial pressures that may arise from embedded organizational structures, routines, and processes that inhibit adaptation (Hannan and Freeman 1984). As a result, inertia may not only be a salient property of organizations but also may be performance enhancing in the face of environmental turbulence.

Whereas population ecology questions the feasibility of organizational adaptation, our theory questions the desirability of adaptation and points to the constraints on beliefs and their revision given some fixed set of experiences. This distinction between structural inertia, on one hand, and cognitive limitations on the rate of learning, on the other, is important because the two problems necessitate different solutions. If the limiting factor is structure, then the managerial implications are related to designing a more flexible organization, one amenable to modification commensurate with the new state of the world (Davis et al. 2009). For example, Brown and Eisenhardt (1997, p. 1) argue for a response that "blends limited structure around responsibilities and priorities with extensive communication and design freedom to create improvisation within current projects." If, in contrast, the limiting factor is a bound on the rate of learning, then the managerial implications are related to actions that enable organizations to more rapidly acquire and effectively process new information.[16] For

example, this may take the form of building enhanced knowledge management capabilities through additional efforts at environmental scanning. In technology industries, this may include corporate venture capital investments that provide enhanced access to information on emerging technologies (Benson and Ziedonis 2009). Alternatively, this may take the form of purposeful organizational forgetting, selectively discarding old knowledge in order to foster more rapid accretion of new knowledge (Hatch and Dyer 2004, Hedberg et al. 1976). McNamara and Baden-Fuller (1999), in a case study of publicly traded biotechnology firms, find that newer firms may be more adaptable not only because they have more flexible organizational structures but also because they learn more rapidly because prior (inappropriate) knowledge does not hinder the process of new knowledge accretion.

Our multiarmed bandit model opens up the possibility of a number of potentially interesting extensions. Two extensions are of particular interest. First, one might model competitive interactions across organizations. In the model presented here, organizations are effectively competing independently against a common environment and subsequently comparing their performance. More realistically, organizations compete directly with one another. Two types of competitive interactions are promising additions to our model. Competition may take the form of competitive selection, in which an organization's probability of failing is a function of its performance relative to other organizations. Alternatively, competition may take the form of competitive performance feedback such that instead of receiving direct feedback on the success or failure of a trial, the organization receives feedback that is a function of its own trial success or failure and also the extent to which others in the population were successful in a given period. This would make feedback about the true value of an alternative noisy and in turn alter both the nature of beliefs and subsequent choices.

Second, one might model organizational awareness of environmental change. Organizations in our model are not directly aware of the existence of environmental change and form no inferences about it. Yet the usefulness of our discussion is predicated on a limited notion of awareness, in the sense that an organization must be sufficiently aware so that it can appropriately tune its adaptation strategy. One might consider a more complex model in which there is second-order learning about the existence, magnitude, and munificence of turbulence. Clearly organizations have beliefs regarding such factors, although there does not seem to be strong evidence that organizations reliably forecast such matters.

---

[16] Of course, there may be some relationship between interventions that address the forces of structural inertia and those that facilitate learning.

Finally, additional extensions include (a) heterogeneity across organizations in the capability to learn; (b) nonfixed search strategies that respond to either the stock of accumulated resources, the nature of existing beliefs, or the level of performance aspirations; (c) interdependence across choices; and (d) endogenous environmental change that derives from interdependence across organizations.

These possible extensions provide important opportunities to build on the insights provided here, but would not negate the basic tensions characterized in the current work. Our objective is to first and foremost understand how knowledge is eroded by turbulence and the consequences of this erosion for the appropriate level of adaptiveness. Adaptive search is multifaceted. It is not only a strategy but also a set of enacted choices. Strategies map beliefs about the relative merits of alternatives to enacted choices. Examining the interplay between these factors greatly enhances our understanding of the dynamics of exploration, and of exploitation, and the rewards to organizational adaptation. Further, the manner in which these factors play out, not surprisingly, depends on the context in which they operate, in particular the degree to which environments are changing. However, the nature of these dependencies is surprising. Environmental change is not a self-evident call for strategies of greater exploration. The question of the appropriate organizational response to environmental change is a central challenge for organizations, and unpacking and understanding this tension remains a fertile and important line of inquiry for organizational theorists and strategy scholars.

Acknowledgments

References

Auer, P., N. Cesa-Bianchi, P. Fischer. 2002. Finite-time analysis of the multiarmed bandit problem. *Machine Learn.* **47**(2–3) 235–256.

Barr, P., J. Stimpert, A. Huff. 1992. Cognitive change, strategic action, and organizational renewal. *Strategic Management J.* **13**(S1) 15–36.

Baum, J., S. Wally. 2003. Strategic decision speed and firm performance. *Strategic Management J.* **24**(11) 1107–1129.

Benner, M. J. 2009. Dynamic or static capabilities? Process management practices and response to technological change. *J. Product Innovation Management* **26**(5) 473–486.

Benson, D., R. Ziedonis. 2009. Corporate venture capital as a window on new technologies: Implications for the performance of corporate investors when acquiring startups. *Organ. Sci.* **20**(2) 329–351.

Bergemann, D., J. Välimäki. 1996. Learning and strategic pricing. *Econometrica* **64**(5) 1125–1149.

Berry, D. A., B. Fristedt. 1985. *Bandit Problems: Sequential Allocation of Experiments*. Chapman and Hall, London.

Blanco, S. 2011. *Consumer Reports*: Chevy Volt "is going to be a tough sell to the average consumer." *AutoblogGreen* (March 1), http://green.autoblog.com/2011/03/01/consumer-reports-chevy-volt-is-going-to-be-a-tough-sell-to-the/.

Brezzi, M., T. L. Lai. 2002. Optimal learning and experimentation in bandit problems. *J. Econom. Dynam. Control* **27**(1) 87–108.

Brown, S., K. Eisenhardt. 1997. The art of continuous change: Linking complexity theory and time-paced evolution in relentlessly shifting organizations. *Admin. Sci. Quart.* **42**(1) 1–34.

Bush, R., F. Mosteller. 1955. *Stochastic Models for Learning*. John Wiley & Sons, New York.

Camerer, C., T. H. Ho. 1999. Experience-weighted attraction learning in normal form games. *Econometrica* **67**(4) 827–874.

Cyert, R., J. March. 1963. *A Behavioral Theory of the Firm*. Prentice-Hall, Englewood Cliffs, NJ.

D'Aveni, R. A., R. E. Gunther. 1994. *Hypercompetition: Managing the Dynamics of Strategic Maneuvering*. Simon and Schuster, New York.

Davis, J., K. Eisenhardt, C. Bingham. 2009. Optimal structure, market dynamism, and the strategy of simple rules. *Admin. Sci. Quart.* **54**(3) 413–452.

Denrell, J., J. March. 2001. Adaptation as information restriction: The hot stove effect. *Organ. Sci.* **12**(5) 523–538.

Dess, G., D. Beard. 1984. Dimensions of organizational task environments. *Admin. Sci. Quart.* **29**(1) 52–73.

Gans, N., G. Knox, R. Croson. 2007. Simple models of discrete choice and their performance in bandit experiments. *Manufacturing Service Oper. Management* **9**(4) 383–408.

Gavetti, G., D. A. Levinthal. 2000. Looking forward and looking backward: Cognitive and experiential search. *Admin. Sci. Quart.* **45**(1) 113–137.

Gittins, J. C. 1979. Bandit processes and dynamic allocation indices. *J. Roy. Statist. Soc.. Ser. B (Methodological)* **41**(2) 148–177.

Gittins, J. C., D. M. Jones. 1974. A dynamic allocation index for the sequential design of experiments. J. Gani, ed. *Progress in Statistics*. North-Holland, Amsterdam, 241–266.

Gittins, J. C., Y. Wang. 1992. The annals of statistics. *Ann. Statist.* **20**(3) 1625–1636.

Gupta, A., K. Smith, C. Shalley. 2006. The interplay between exploration and exploitation. *Acad. Management J.* **49**(4) 693–706.

Hannan, M., J. Freeman. 1984. Structural inertia and organizational change. *Amer. Sociol. Rev.* **49**(2) 149–164.

Hardwick, J. P., Q. F. Stout. 1992. Bandit strategies for ethical sequential allocation. *Proc. 23rd Sympos. Interface*, American Statistical Association, Alexandria, VA, 421–424.

Hatch, N. W., J. H. Dyer. 2004. Human capital and learning as a source of sustainable competitive advantage. *Strategic Management J.* **25**(12) 1155–1178.

Hedberg, B., C. Nystrom, W. Starbuck. 1976. Camping on seesaws—Prescriptions for a self-designing organization. *Admin. Sci. Quart.* **21**(1) 41–65.

Henderson, R. 1993. Underinvestment and incompetence as responses to radical innovation: Evidence from the photolithographic alignment equipment industry. *RAND J. Econom.* **24**(2) 248–270.

Holland, J. 1975. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control & Artificial Intelligence*. University of Michigan Press, Ann Arbor.

Keller, G., S. Rady. 1999. Optimal experimentation in a changing environment. *Rev. Econom. Stud.* **66**(3) 475–507.

Lant, T., S. Mezias. 1992. An organizational learning model of convergence and reorientation. *Organ. Sci.* **3**(1) 47–72.

Le Mens, G., J. Denrell. 2011. Rational learning and information sampling: On the naivety assumption in sampling explanations of judgment biases. *Psychol. Rev.* **18**(2) 379–392.

Luce, R. 1959. *Individual Choice Behavior: A Theoretical Analysis.* Wiley, New York.

March, J. G. 1991. Exploration and exploitation in organizational learning. *Organ. Sci.* **2**(1) 71–87.

March, J. G. 1996. Learning to be risk averse. *Psychol. Rev.* **103**(2) 309–319.

March, J. G. 2003. Understanding organizational adaptation. *Soc. Econom.* **25**(1) 1–10.

March, J. G. 2010. *The Ambiguities of Experience.* Cornell University Press, Ithaca, NY.

McNamara, P., C. Baden-Fuller. 1999. Lessons from the Celltech case: Balancing knowledge exploration and exploitation in organizational renewal. *British J. Management* **10**(4) 291–307.

Miller, D. 1987. The structural and environmental correlates of business strategy. *Strategic Management J.* **8**(1) 55–76.

Nelson, R. 2008. Bounded rationality, cognitive maps, and trial and error learning. *J. Econom. Behav. Organ.* **67**(1) 78–89.

Nelson, R., S. G. Winter. 1982. *An Evolutionary Theory of Economic Change.* Belknap Press, Cambridge, MA.

Peters, T. J., R. H. Waterman. 1982. *In Search of Excellence: Lessons from America's Best-Run Companies.* Harper & Row, New York.

Radner, R., M. Rothschild. 1975. On the allocation of effort. *J. Economic Theory* **10**(3) 358–376.

Robbins, H. 1952. Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc.* **58**(5) 527–535.

Schmalensee, R. 1975. Alternative models of bandit selection. *J. Econom. Theory* **10**(3) 333–342.

Sorenson, J., T. E. Stuart. 2000. Aging, obsolescence, and organizational innovation. *Admin. Sci. Quart.* **45**(1) 81–112.

Sutton, R. S., A. G. Barto. 1998. *Reinforcement Learning: An Introduction.* MIT Press, Cambridge, MA.

Teece, D. J., G. Pisano, A. M. Y. Shuen. 1997. Dynamic capabilities and strategic management. *Strategic Management J.* **18**(7) 509–533.

Tushman, M., E. Romanelli. 1985. Organizational evolution: A metamorphosis model of convergence and reorientation. L. L. Cummings, B. M. Staw, eds. *Research in Organizational Behavior.* JAI Press, Greenwich, CT, 171–222.

Valdes-Dapena, P. 2009. Ford: No Volt for us. *CNNMoney* (January 13), http://money.cnn.com/2009/01/12/autos/ford _electric_plans/index.htm.

Vermorel, J., M. Mohri. 2005. Multi-armed bandit algorithms and empirical evaluation. J. Gama, R. Camacho, P. Brazdil, A. Jorge, L. Torgo, eds. *Machine Learn.: ECML 2005*, LNAI 3720. Springer-Verlag, Berlin, 437–448.

Weber, E., S. Shafir, A.-R. Blais. 2004. Predicting risk sensitivity in humans and lower animals: Risk as variance or coefficient of variation. *Psych. Rev.* **111**(2) 430–445.

Weick, K. 1969. *The Social Psychology of Organizing.* Addison-Wesley, Reading, MA.

Welch, D. 2008. GM: Live green or die. *Bloomberg Businessweek* (May 15), http://www.businessweek.com/print/magazine/ content/08_21/b4085036665789.htm.

Wholey, D., J. Brittain. 1989. Characterizing environmental variation. *Acad. Management J.* **32**(4) 867–882.

Whittle, P. 1988. Restless bandits: Activity allocation in a changing world. *J. Appl. Probability* **25** 287–298.