Received: 28 January 2021 | Revised: 9 June 2024 | Accepted: 5 July 2024

DOI: 10.1002/smj.3646

RESEARCH ARTICLE

SMS | Strategic Management Journal

WILEY

Organizational adaptation in dynamic environments: Disentangling the effects of how much to explore versus where to explore

Kannan Srikanth[1] | **Tiberiu Ungureanu**[2]

[1]Department of Management and Human Resources, Max M. Fisher College of Business, The Ohio State University, Columbus, Ohio, USA

[2]Department of Management, Walker College of Business, Appalachian State University, Boone, North Carolina, USA

Correspondence
Kannan Srikanth, Department of Management and Human Resources, Max M. Fisher College of Business, The Ohio State University, 2100 Niel Ave, Columbus, OH 43202, USA.
Email: srikanth.18@osu.edu

Abstract

Research Summary: There is considerable debate about how firms should adapt to environmental dynamism. Theoretically, some scholars suggest that with increasing dynamism, firms should explore more, whereas others argue that firms should explore less. Empirical evidence remains mixed. We attempt to reconcile these mixed findings by (a) distinguishing between two facets of exploration—exploration propensity versus exploration breadth, and (b) recognizing that firms may make these two decisions using different decision-making processes. Using a computational model we show that with increasing environmental dynamism, for high performance, (a) firms' exploration propensity may increase, decrease, or stay the same depending on their decision-making process, but (b) firms' exploration breadth always increases. Our results help explain the mixed findings in this domain and have implications for future empirical work.

Managerial Summary: Responding to dynamic environments is challenging for managers. There is limited support for the intuition that firms should explore more in more dynamic environments. We recognize that exploration decisions in firms are temporally and hierarchically separated—senior managers first decide how much to explore and middle managers then decide which projects to fund. In this research, we use a computational model to unpack how these two facets of exploration may change in dynamic environments for firms to maintain high performance. We find that as dynamism increases, how much firms explore depends on how sensitive their decision-making process is to the perceived attractiveness of the different options, but when they explore, they should always choose options further away from their status-quo.

KEYWORDS

computational model, decision-making, environmental turbulence, exploration-exploitation, organizational adaptation

1 | INTRODUCTION

In dynamic environments, firm performance critically hinges upon finding an appropriate balance between the pursuit of new knowledge—*exploration*—versus the use of existing knowledge—*exploitation* (Levinthal & March, 1993; March, 1991; O'Reilly & Tushman, 2013). Understanding how firms should balance the need for exploration versus exploitation in more dynamic environments is of increasing practical importance, and CEOs are keenly aware that their firms are competing in rapidly evolving industries. In response, academics have generated an avalanche of research on how firms should adapt to environmental dynamism (turbulence). However, currently, there is no consensus on how they should respond to increasing environmental turbulence.

Some scholars theoretically argue that exploration should increase with turbulence (Eisenhardt & Martin, 2000; Kim & Rhee, 2009; Siggelkow & Rivkin, 2005), whereas others argue that it should decrease (Keller & Rady, 1999; Posen & Levinthal, 2012; see also population ecology scholars Hannan & Freeman, 1984). Empirical evidence about the relative importance of exploration versus exploitation as environmental change increases is also mixed. Some studies have found that with increasing dynamism, firms should *increase* exploration (Jansen et al., 2006; Nadkarni & Narayanan, 2007; Wang & Li, 2008), while other research has suggested that they should *decrease* exploration (Amburgey et al., 1993; Barnett, 1994; Barnett & Freeman, 2001; Marino et al., 2015; Ruef, 1997; Schilke, 2014).

We believe current conceptual frameworks are ill-equipped to explain this discrepancy in empirical work because they employ a unitary actor perspective that does not reflect the empirical reality that explore-exploit decisions in most organizations are made hierarchically. Empirical work has identified two distinct facets to exploration (Billinger et al., 2021): (a) *exploration* *propensity* measuring whether to search (a little, a lot) and (b) *exploration breadth*, measuring where to search (narrowly, adjacent to current activities, or broadly pursuing new technical or market domains). Practitioners and recent scholarly work not only distinguish between these two facets of exploration but also suggest that they are temporally and organizationally separated (Criscuolo et al., 2017; Rigby & Zook, 2002).

In addition, if exploration decisions are made by different people at different times, it is also likely that they follow different decision-making processes; they may be *evaluative*, that is, choosing alternatives based on a careful evaluation of their relative merit, or *nonevaluative*, that is, choosing alternatives in a more random, lottery type process. Although the firm's decision-making process will likely influence its performance in stable versus turbulent regimes, this aspect has remained uninvestigated.

Modeling organizational explore-exploit decisions as hierarchically made using different decision processes is thus the next step in building our cumulative knowledge on how firms should respond to environmental dynamism. We add to prior work by Posen and Levinthal (2012) and build a computational model that separates these two facets of exploration—exploration propensity and exploration breadth. In addition, we model two different decision processes that firms may use to make these decisions. Drawing on Billinger et al. (2021), in our model, organizations first decide how much to search and then decide how widely to search. In each stage, the organization could either make this decision using an evaluative or non-evaluative decision process.

We find that to improve performance in increasingly dynamic environments, (1) firm's exploration propensity may increase, decrease, or stay the same, depending on the decision-making process (i.e., whether it is evaluative or nonevaluative), and (2) firm's exploration breadth always increases, although the extent to which it increases depends on the decision process. In our model, increasing exploration breadth increases knowledge about which alternatives are attractive. Since in dynamic environments, alternatives' attractiveness can change unpredictably, broad exploration will help increase knowledge. However, such broad exploration simultaneously increases the firm's exploration cost, as it samples and discards many bad alternatives. Exploration propensity controls how often the firm explores and thus directly controls cost, and is sensitive to the decision-making process managers use. These findings partially explain current mixed findings from prior work as special cases of our more general framework and have significant implications for future empirical work.

2 | LITERATURE REVIEW

2.1 | Empirical results are conflicting and inconclusive

Empirical work related to organizational adaptation and learning in dynamic environments has thus far produced conflicting results, suggesting that exploration should increase, decrease, or even stay the same as environmental dynamism increases. These studies have used varied measures of exploration and dynamism and used differing logics to argue their propositions.

Several studies have suggested that firms should explore more when the environment is dynamic. Empirical work has shown that in dynamic environments, proxies for greater exploration, such as exploring new knowledge domains (Wang & Li, 2008), new products and markets (Jansen et al., 2006), increasing top management attention to innovation and considering a wider variety of information (Eisenhardt, 1989; Garg et al., 2003), strategic variety and flexibility

(Larrañeta et al., 2014; Nadkarni & Narayanan, 2007), and introducing a wider range of product features (Klingebiel & Joseph, 2016) are associated with positive financial performance, and to related measures such as innovation and faster product development (Eisenhardt & Tabrizi, 1995; Sidhu et al., 2007). These scholars often argue that the value of exploration increases in dynamic environments because new opportunities can be discovered (Bourgeois & Eisenhardt, 1988; Kim & Rhee, 2009; Siggelkow & Rivkin, 2005). The theoretical basis for their empirical work is that when new opportunities are fleeting, increasing exploration allows firms to seize them as they arise (Jansen et al., 2006).

In contrast, a few studies have found that firms that explore less in dynamic environments are better off. Schilke (2014) finds that the value of dynamic capabilities reduces in dynamic environments, while Markovitch and O'Brien (2021) found that increases in R&D intensity are associated with a lower Tobin's q. In a similar vein, studies have shown that increasing product changes or new product introductions backfires in more dynamic environments (Amburgey et al., 1993; Barnett & Freeman, 2001; Marino et al., 2015). These scholars suggest that environmental dynamism increases the quantity and complexity of information processing for making good decisions (Karim et al., 2016), by which time the environment has changed again (Posen & Levinthal, 2012; also see arguments by population ecology scholars, Hannan & Freeman, 1984; Ruef, 1997; Wholey & Brittain, 1989). In other words, as environmental shocks become too frequent, the useful life of new information is too short to cover the cost of obtaining it (Posen & Levinthal, 2012). Accordingly, exploration is only effective if the frequency of change is not too high (Keller & Rady, 1999).

2.2 | Empirical results appear to be sensitive to measurement

Lavie et al. (2010) have argued that considerable heterogeneity exists in the literature on how scholars define and measure exploration (and exploitation). They further suggest that these heterogeneous measures have significantly impeded knowledge cumulation in this area and note that "results are highly sensitive to particular operationalization" (p. 118).

Many scholars have assumed that a centralized unitary actor makes exploration decisions and have captured how often or how much resources were devoted to exploration vis-à-vis exploitation activities. For example, several studies have measured exploration as R&D spending, whereas others measure exploration as the extent of new products or activities (Greve, 2003; Gaba & Greve, 2019; Lee & Meyer-Doyle, 2017; Schilke, 2014). These scholars, however, only capture whether the firm does something new, not how novel it is, although the assumption that increased spending increases novelty is implicit in most work. For example, increasing R&D budgets may result in the pursuit of projects that are close to or distant from the firm's existing projects; new products may be incremental improvements of existing ones or may pursue radically different technology or market opportunities. To capture this dimension, other scholars have measured exploration explicitly as search breadth or scope (Jansen et al., 2005; Larrañeta et al., 2014; Sidhu et al., 2007; Wang & Li, 2008).

Thus, prior research has considered at least two distinct decision dimensions of exploration—how much to search (not at all, a lot) versus how broadly to search (locally, adjacent to current activities, or distally, in new domains). Recognizing this, scholars have measured exploration using scales that combine these two dimensions (Cao et al., 2009; Heij et al., 2014; Jansen et al., 2012; Lavie et al., 2011; Stettner & Lavie, 2014). However, in line with

Billinger et al. (2021), we expect that treating these two dimensions as distinct, yet interdependent decisions may yield superior insights.

2.3 | Different measures may reflect different decision processes

Implicit in much of this work is the considerable heterogeneity in assumptions regarding firms' decision-making process about their exploration choices that is reflected in the different measures. Qualitative and anecdotal evidence suggests that managers do distinguish between these two aspects of exploration, and that they are temporally and organizationally separated (Rigby & Zook, 2002). Recognizing this distinction also suggests that firms may make these two decisions using different processes.[1]

Prior work that mainly adopts the unitary actor view of firm exploration decisions already implicitly suggests different decision processes. One implicit process is that managers pursue something new (explore) because they value novelty, often because their existing options produce results below their aspiration levels or because they have slack (Greve, 2003, 2007). In contrast, qualitative work in strategic decision-making in turbulent environments often describes a more deliberative process where managers constantly evaluate their current products and strategies against alternatives and explore only when they believe that these alternatives may be superior to status quo (Bhardwaj et al., 2006; Brown & Eisenhardt, 1997; Eisenhardt, 1989; Tripsas & Gavetti, 2000). These two kinds of decision-making may be applied to one or both facets of the exploration decision. Scholarly and managerial work that takes a more hierarchical view of how exploration decisions are made in firms describes alternative ways by which these firms decide on the two facets of exploration. We detail these below.

2.3.1 | Standard decision process

A common process involves a higher-level manager such as the CEO deciding on exploration propensity, such as R&D budgets (as opposed to say, the budget for operational excellence; cf. Benner & Tushman, 2003), but delegates the decision on which projects are funded to a second decision maker such as an R&D committee (Criscuolo et al., 2017). For example, during COVID, Pfizer's senior leadership not only increased its R&D budget, but also tasked R&D leaders to increase their search breadth and evaluate more candidates (Bourla, 2021). Similarly, government departments such as the Department of Energy, NIH, or NSF set aside funding for research by legislation, and an expert committee evaluates proposals and decides whether to fund well-understood or novel projects. We believe this process is very common and label it the *standard process* going forward.

It is important to note that in the *standard process*, often the increase in R&D budgets is reflexive to increasing turbulence, and not necessarily because any potential opportunity is superior to the firms' current best practice. For example, Gates said Microsoft "would continue to boost spending to invest in future technology and pour as much as $6.9 billion on R&D during its current fiscal year,"[2] while Bourla said the first thing he did after becoming CEO of Pfizer was to increase R&D budget from $6 billion to $11 billion to build infrastructure that will

[1]We thank an anonymous reviewer for suggesting we model the different decision processes.
[2]https://www.hpcwire.com/2003/08/01/microsoft-to-increase-rd-budget/, retrieved June 6, 2024.

help the company weather disruptions.[3] Note that these decisions to increase R&D budgets were not made because these firms were doing something suboptimal but with the desire to explore and gain knowledge about other alternatives. *Thus, in the standard process, the first stage (propensity) decision is nonevaluative, whereas the second stage (breadth) decision is evaluative.*

2.3.2 | Evaluative decision process

Other managers are very deliberative about whether to spend money on exploration activities, constantly comparing their current strategies or product offerings to other potential alternatives. For example, Eisenhardt (1989) described a process where the CEO and their close advisors evaluated alternatives and set direction, which other executives tried to achieve. In some of the cases she describes, the decision was to stick to their current strategy, and enhance their core product offering rather than develop new ones (also see Bourgeois & Eisenhardt, 1988). Similarly, Klingebiel and Joseph (2016) document an exploration process in the mobile phone industry, where senior executives make portfolio wide decisions on breadth and speed of feature introduction by comparing their current portfolio with that of competitors, and middle managers execute those decisions by evaluating different alternatives. The more managers believe their current strategy is superior, such as strong belief about the superiority of the consumables business model in Polaroid (Tripsas & Gavetti, 2000), the less likely they are to explore. For example, Steve Jobs implemented such an evaluative process when he turned around Apple in 2000, when he said "I am going to wait for the next big thing" (quoted in Rumelt, 2011; p14), and until then pursued a strategy of cost reduction rather than innovation. We henceforth call this the *evaluative process*. Thus, in the evaluative process, the first stage (propensity) decision as well as the second stage (breadth) decision are evaluative. Recall that such evaluation is absent in the first stage of the *standard process*. Incentives, leadership, and cultural factors can influence how much a firm explores even with identical evaluations.

2.3.3 | Lottery decision process

In both the standard and evaluative processes, the second decision on exploration breadth is evaluative, where a committee evaluates the relative attractiveness of proposals and funds them accordingly. Scholars and practitioners have suggested that this practice tends to favor familiar ideas over more novel ones (Kahneman, 2003; March, 1994). More familiar proposals that are closer to the organizations' current processes often appear more attractive at first look compared to more distant alternatives for three reasons. First, due to path dependence (Nelson & Winter, 1982) and myopia of learning (Levinthal & March, 1993), adjacent technologies often appear more attractive than novel ones even though the latter may have better long term potential. Second, since distant technologies are less understood, loss-averse individuals may discount them more (Tversky & Kahneman, 1973). Finally, decision makers may be experts in more familiar domains and may fundamentally over-value their expertise in a self-enhancement process (Katz & Allen, 1982). Thus, scholars have long suggested that funding committees err on

[3]https://fortune.com/2022/04/13/pfizer-ceo-albert-bourla-covid-vaccine-development-big-pharma-public-partnership/, retrieved June 6, 2024.

the side of overinvesting in familiar technologies even when given the mandate to explore broadly (Criscuolo et al., 2017).

It is in this context that scholars and practitioners have suggested that the organization may fund truly novel projects by allocating at least a portion of its resources randomly by a lottery with no evaluation (Azoulay & Li, 2020; The Economist, 2022). One such example of firms funding novelty is the funding of central research labs, such as AT&T Bell Labs, Xerox PARC, Microsoft Research Labs or DuPont Central Research. These labs' budgets were often funded as a percentage of sales (or profits), decided by senior management, and were given a mandate to pursue basic science. The funding model utilized a portfolio approach, since it was unclear which, if any of the projects pursued will prove to be profitable for the company.

We model the *lottery process* to capture this model for funding innovation, where *both the first stage* (*propensity*) *decision and the second stage* (*breadth*) *decisions are nonevaluative*. Thus, in the *lottery process*, the funding committee uses most of its budget to fund the best available alternative (i.e., to the traditional industrial research) but uses a portion of the budget to fund a project at random (i.e., to basic science, which is akin to a lottery). Thus, the key decision is what fraction of the budget to devote to the lottery.

In sum, when we consider that exploration decisions are multi-faceted and organizationally rooted, different decision processes come into play. Thus, any advice regarding how firms make exploration decisions in turbulent environments should account for the decision-making process for it to be useful. While the three decision processes we describe are not exhaustive, we believe the first two are fairly common, and the lottery process an interesting alternative, that was adopted historically by major firms, and is again increasingly discussed in the policy space.

2.4 | Research gap

Explore-exploit decisions in organizations are often made hierarchically and different decision processes may be involved in making these contingent decisions. Conceptual work thus far however does not incorporate the hierarchical nature of organizational decision-making when they consider how firms should respond to environmental turbulence. We can understand neither the influence of these distinct facets nor the mechanism behind how these different facets change in different kinds of environmental turbulence by simply examining existing empirical work more carefully. Since exploration-exploitation research aims to explicate how firms should adapt their strategies in the face of environmental dynamism, lack of understanding of how different facets of exploration, or the different decision processes affect firm performance represents an important gap in our knowledge.

3 | MODEL

To answer our research question, we follow prior theoretical research and propose to examine these contingencies using a computational model. In our simulation, we use a modified version of the canonical N-armed bandit model that has been widely used in prior research to understand the tradeoff between exploration and exploitation in dynamic environments (Lee & Puranam, 2016; March, 1996; Posen & Levinthal, 2012; Stieglitz et al., 2016). In the N-armed bandit model, the agent chooses between uncertain options to maximize their outcomes, and the decision-making process is modeled as driven by reinforcement learning (Daw et al., 2006; Sutton & Barto, 1998).

Reinforcement learning comprises two interrelated processes: (1) the *choice* process of deciding between exploiting what is currently believed to be the best-performing option for immediate reward versus exploring other options currently believed to be inferior to gain additional information; and (2) the *adaptive learning* process of translating feedback from accumulated experience into beliefs about the relative attractiveness of the available options (*henceforth beliefs*) (Denrell & March, 2001; March, 1996; Sutton & Barto, 1998). These two processes are closely intertwined, given that choices determine what feedback is received and that beliefs shaped by learning determine subsequent choices (Li et al., 2009). This tradeoff between gathering and using the acquired information is especially pertinent in changing environments, where learning from prior experience may be less useful for making choices about future actions (Daw et al., 2006; Posen & Levinthal, 2012).

3.1 | The task environment

The bandit model consists of N discrete choices or arms. Each arm i at time t provides a normal distributed payoff with mean $\mu_{i,t}$ and variance 0.05, where each arm's mean $\left(\mu_{i,t}\right)$ is drawn from $N(0.5, 0.05)$ distribution. We chose this distribution since it has the same mean and variance as the beta (2, 2) distribution used in prior studies (i.e., Lee & Puranam, 2016; Posen & Levinthal, 2012), but in addition has desirable properties that allow us to manipulate amplitude of environmental turbulence while preserving the distribution of the arm means.[4]

3.2 | Changes to the task environment (or environmental turbulence)

The task environment is subject to exogenous shocks. We modeled environmental shocks as changes in the mean of the payoff distribution (μ_{it}) for each arm, similar to prior work. In our main model, we change the task environment on only one dimension, the frequency (η) of shocks. In sensitivity analysis, we also vary the amplitude (λ) of shocks. In our model, we followed the approach of Posen and Levinthal (2012) by suggesting that shocks probabilistically alter the payoffs for the arms. When a shock occurs, the payoff probability (μ_{it}) for each arm is independently reset by an independent draw from the same underlying distribution with a probability of 0.5.[5]

[4]Preserving the underlying distribution of arm means is particularly important when we attempt to compare exploration strategy between different regimens of turbulence amplitude and frequency. If the distribution of the underlying arm means changes, the same exploration strategy will result in a different proportion of explore to exploit choices. Therefore, when the distribution changes, we cannot interpret changing values of exploration strategy as increase or decrease in exploration. This is a problem in some prior models, where the underlying distribution changes with turbulence, making it hard to interpret their results.

[5]In this formulation, multiple arms are likely to change their means at the same time. However, for each arm, the new mean is an independent draw from the underlying distribution.

3.3 | Choice process

Since the relative attractiveness of the arms is unknown to agents, and the outcomes to choices are probabilistic, decision-makers need to sample the arms repeatedly to identify the superior choices. This reinforcement learning process involves choosing arms based on current beliefs and updating beliefs based on the observed payoffs to previous choices. Prior work has strongly suggested that different firms may have different choice processes, which in turn have important consequences for firm adaptation.

The simplest choice process used in bandit models is the ϵ-*greedy* model (Sutton & Barto, 1998), where the agent's exploration strategy is denoted by the parameter "ϵ." In this model, the decision maker exploits, that is, chooses the best believed alternative with probability 1-ϵ, and explores, that is, chooses the alternative believed to be suboptimal with probability ϵ. When the agent explores, it chooses among the available alternatives randomly. As Posen and Levinthal (2012, p. 590) point out, a key disadvantage of the ϵ-*greedy* model is that agents are simultaneously too value-sensitive, and completely value insensitive. For example, if the three arms have values [0.11, 0.10, and 0.01], the process always chooses the first arm when it exploits and is indifferent between the second and third arms when it explores.

To overcome this issue, a second method involves the decision maker considering the relative attractiveness of available alternatives in making their exploration choices. This decision process is widely deployed in prior bandit models and is represented by the *softmax* algorithm (Camerer & Ho, 1999; Luce, 1959; Sutton & Barto, 1998) as shown below:

$$\pi_{i,t} = \exp\left(\frac{q_{i,t}}{\tau}\right) / \sum_{j=1}^{N} \exp\left(\frac{q_{j,t}}{\tau}\right) \tag{1}$$

in which $\pi_{i,t}$ is the probability of choosing the ith arm among the total number of available arms, N, in round t, $q_{i,t}$ is the agent's belief about the mean payoff of arm i at time t, and $\tau>0$ is the parameter that governs the exploration strategy. As before, when the agent chooses the option that is currently believed to be the best, it exploits; otherwise, it explores. When τ is high, the agent simultaneously has a higher propensity to explore, that is, choose the alternative that is currently not the best—and to explore broadly, that is, choose options with little regard to its attractiveness. When τ is low, the agent is less likely to explore, and when it does explore, explores more narrowly.

The τ parameter translates the agent's beliefs about the available alternatives into action choices (Posen & Levinthal, 2012). One way to interpret this parameter in an organizational setting is to consider this as the innovation focus or culture of the top management team (Garg et al., 2003); the more the innovation focus, the more likely the firm to try something novel, after considering the attractiveness of the currently best performing choice relative to other choices. As a widely used choice process (Posen & Levinthal, 2012 used this to model this research question), with significant empirical traction at the individual level (Daw et al., 2006), this also serves as a second control process, against which we can compare the other models of decision-making.

3.4 | Modeling the two parameter decision processes

As we argued previously, organizations may decouple the decisions on how much to explore from how widely to explore. In this case, the decision of how much to explore is antecedent to

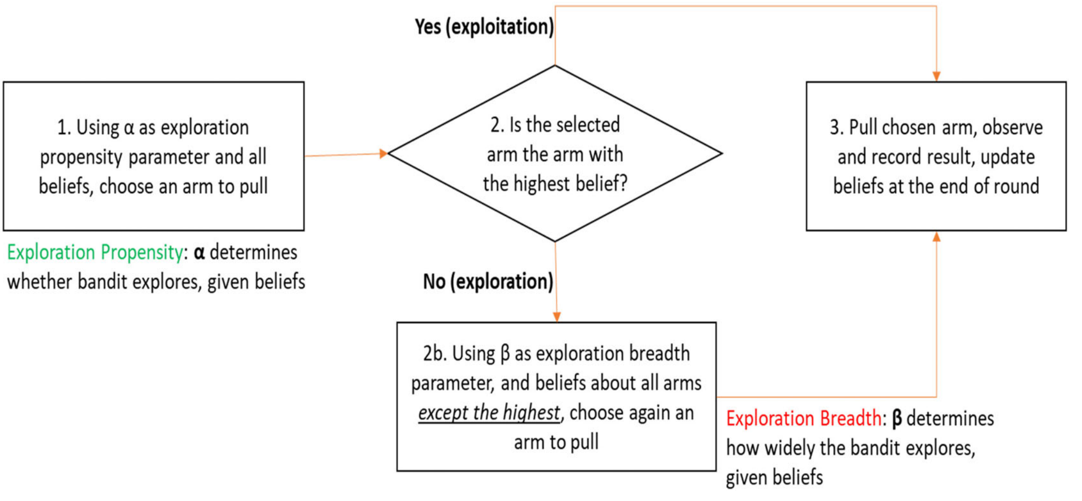

FIGURE 1 Two-parameter exploration/exploitation decision-making process. In step 1, in the *standard process* and *lottery process*, use α instead of ϵ as the choice parameter; in the *evaluative process*, use α instead of τ in Equation (1). Note that in this step, all 10 arms in the bandit are considered when making the choice. In Step 2b, in the standard and evaluative process, replace τ with β in Equation (1); in the lottery process, replace ϵ with β. Note that in this step, we exclude the currently best believed arm from being an allowable choice; thus this bandit only has nine arms. In the baseline implementation, the agent has a budget of 10 in each round. In this case, in step 1, make 10 choices. Out of these 10, for the number allotted in exploration, step 2b comes into play. Then in step 3, the payoffs are updated, and beliefs recalculated for all the choices made.

the decision on how widely to explore (cf. Billinger et al., 2021). We implemented the two-stage process as shown in Figure 1.[6]

In our model, the agent first determines whether to explore, by using the specified choice process (ϵ-*greedy* or *softmax*) over all the arms governed by the exploration propensity parameter α. If at this stage, the agent chooses the currently most attractive option, then the agent exploits (by choosing the best arm), observes the outcome, and updates their beliefs (the top pathway in Figure 1). The higher the α, the more likely the agent explores, capturing exploration propensity. If in the first stage the agent chooses an arm that is *not* the best arm, then the agent explores in this round (see the bottom pathway in Figure 1). At this point, the agent chooses among the exploration options (i.e., among all the arms *excluding* the arm that is currently believed to be the most attractive), using a *softmax* (or ϵ-*greedy*) choice process with exploration breadth parameter β.[7]

In this way, we can decouple decisions about *how much to explore* (α) from decisions regarding *how widely to explore* (β). As α increases, the agent tends to explore more often, and when it does explore, the larger the β, the more frequent the choice of an arm currently believed to be less valuable. We model the three decision processes outlined earlier as follows:

3.4.1 | The evaluative process

We model this process as one where both stages involve the *softmax* rule, where the first stage decision on whether to explore is based on managers' current understanding of how superior

[6]The simulation code used (python) is available for download at https://github.com/tiberiuu/DualBandit.

[7]Thus, at the second step of the choice process the agent is facing a (N-1)-arm bandit.

the currently best-performing option is relative to other available opportunities. In the second stage, again, a committee decides which projects to fund based on their understanding of their relative attractiveness. This process expands on prior research (Lee & Puranam, 2016; Posen & Levinthal, 2012; Stieglitz et al., 2016), which modeled an evaluative process based on a single-stage *softmax* choice.

3.4.2 | The standard process

In this process, the first-stage managers use an ϵ-*greedy* rule to determine whether they explore, followed by a second stage *softmax* rule, where they determine how widely they explore. We made these choices because in the *standard process* the first stage decision is nonevaluative, but the second stage decision is evaluative.

3.4.3 | The lottery process

In this process, both stages employ the ϵ-*greedy* rule. Here, the larger the α the more the organization tends to explore, and the larger the β, the less the agent chooses the second most attractive arm, instead choosing randomly from the remaining N-2 choices.

3.5 | Belief updating process

The agent updates its beliefs about the attractiveness (i.e., mean value) of the arm every period after it samples one of the arms.[8] Following the averaging learning rule from prior work (Lee & Puranam, 2016; Posen & Levinthal, 2012; Stieglitz et al., 2016), the belief $q_{i,t+1}$ about the estimated payoff of the ith arm at time $t+1$ is calculated as a function of the previous belief, observed outcome, and the (discounted) number of times the arm has been sampled, as shown below:

$$q_{i,t+1} = q_{i,t} + \frac{1}{k_i+1}\left(\sigma_{i,t} - q_{i,t}\right) \tag{2}$$

Here, $\sigma_{i,t}$ represents the outcome of pulling arm i at time t, and k_i represents the number of times arm i has been pulled at time t. We chose this as the baseline since in the behavioral theories, behavioral plausibility is of fundamental concern, and arguably averaging is behaviorally the simplest and more straight-forward assumption (Posen & Levinthal, 2012). However, as LiCalzi and Marchiori (2013) pointed out, in highly turbulent environments, it is unlikely that managers simply average information they know to be stale. Therefore, in sensitivity analyses, we also implemented a process where managers discount old information.

[8]When the agent makes multiple choices in the same period, the beliefs are updated at the end of the period, considering both the number of times arms were chosen and each individual feedback the arms have returned.

3.6 | Model parameter choices

To make our results comparable with those of prior work, our choice of parameters closely follows those of Posen and Levinthal (2012). Thus, we modeled a bandit with 10 arms, and for each combination of strategy parameters (α and β), we seeded our simulation with 100,000 firms, each facing a unique environment.[9]

Every period the firm allocates 10 resources, each of which being allocated to either exploration or exploitation, based on the firm's beliefs and selected decision-making process. We made this choice (instead of the traditional one resource per period used in prior work) to resemble the allocation process in firms more closely, where multiple projects are funded simultaneously, and show sensitivity of our results to allocating only one resource per round.[10]

We ran the simulation for 500 periods and reported the cumulative performance averaged over all firms at the end of the simulation. The initial beliefs across alternatives are set to be homogenous and set to the value of 0.5, which is the actual mean value of the arms' payoff distribution. We used pilot tests to identify the range of our parameters, where we observed theoretically meaningful changes in behavior.[11] In line with our theory, we set α and β independently, resulting in 420 α/β combinations at each level of turbulence. Since our model has more parameters and our search space more fine-grained, we need to use more simulations to identify robust results compared to single parameter models (North & Macal, 2007).

As in prior research, we also assumed that exploration strategy is exogenous, but behavior (i.e., number of exploration pulls) is endogenous to the task environment. Thus, for each level of turbulence (η), we reported the best-performing (henceforth referred to as *optimal strategies*; performance is measured as cumulative earnings) combination of exploration strategies (α and β) at each environmental condition.

4 | ANALYSIS

4.1 | Strategies in dynamic environments—Single parameter models

4.1.1 | Evaluative process

Before we begin to analyze our results, it is useful to briefly recall findings from prior related models. Posen and Levinthal (2012) found that as frequency of environmental change increases, the optimal exploration *strategy* (τ)—that is, the exploration strategy that leads to highest performance at a given turbulence level—has an inverse-*U*-shaped effect. We replicate these results (see Figure 2).

[9]We see the general pattern in our results occurring when we seed our simulation with 5000 firms.

[10]In the choice-making process, the agent chooses which arms to pull at the beginning of each period, using the same set of beliefs it had when the period started. For example, if the agent has 10 resources it first makes a choice with parameter α 10 times. Then, assuming the agent chose to exploit six of these 10 times, it then makes a choice with the parameter β four times to pick the arms she is going to explore from the N-1 arms in the exploration set. After all the choices for the period are made, the agent pulls the arms as dictated and observes the outcome of her choices. At the end of the period, after all choices are made and outcomes observed, the agent updates her beliefs.

[11]More specifically, in the evaluative model, α varied between .0175 and .0675, whereas β varied between .015 and .30 in increments of .015. In the standard model, α varied between .01 and .1, and β between .01 and .40 in increments of .01.

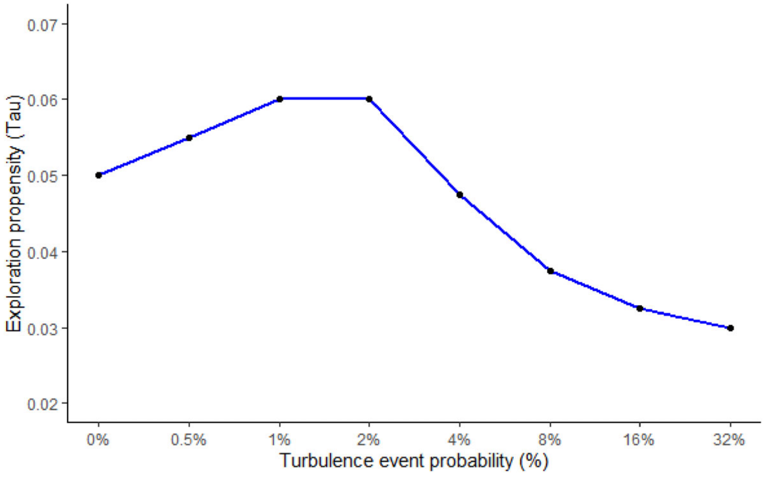

FIGURE 2 Exploration strategy in dynamic environments (using softmax rule).

The fundamental intuition behind this result is that the value of exploration degrades with increasing turbulence while its cost increases. Recall, the agent does not know whether and when a turbulence event occurred; and they gain knowledge from experience.[12] Increasing τ increases knowledge but at a decreasing rate. In highly turbulent environments, new knowledge unearthed by increasing exploration is less useful for two reasons: (a) if the agent forms beliefs by averaging prior experience, their beliefs are likely to be inaccurate at any given point in time, since they are the average of many different values the arm has taken over time; and (b) by the time experience overcomes this problem to form accurate beliefs, the environment has changed again. In sum, the slope of the knowledge curve with increasing τ flattens as turbulence increases.

At the same time, the cost of exploration—the cost the agent incurs to gather more knowledge by choosing an option they believe is not the best—increases for two reasons.[13] First, with increasing τ, the agent chooses more or less randomly between valuable and less valuable choices, thus incurring a penalty. Second, with increasing τ, the agent explores more often. In highly turbulent environments, the agent explores much more often due to belief flattening (Posen & Levinthal, 2012). As knowledge degrades (reason [a] above) the beliefs formed by averaging prior experience about all the arms regress toward the mean. The closer the beliefs about arms, the greater the exploration, that is, the more the agent explores at any given strategy (τ), and exploration cost skyrockets. In sum, the slope of the cost curve with increasing τ steepens as turbulence increases.

In sum, increasing exploration (increasing τ) increases the bandit's knowledge, but at a decreasing rate. As the environment becomes more turbulent, the rate of knowledge increase

[12]We draw on Lee and Puranam (2016) and formally define (lack of) knowledge as the difference in the true value of the actual best arm and the best-believed-arm, capturing how "wrong" the agent is in its beliefs. Each period we subtract this measure from 1 so that increasing values of this measure capture increasing knowledge. In our figures, we show cumulative knowledge at the end of the simulation by adding this up for every period.

[13]We define per-period exploration cost as the difference in the value of the best believed arm and the arm that is chosen, since this is what the agent gives up by not exploiting in that round. The agent minimizes "regret" over time when they maximize difference between cumulative knowledge and cumulative exploration cost, thus maximizing cumulative performance (Agrawal & Goyal, 2012; Bubeck & Cesa-Bianchi, 2012; Lattimore, 2016; Lu et al., 2020).

with increasing exploration reduces significantly. At the same time, when the agent explores, it incurs exploration cost. With increasing turbulence, this cost cumulates rapidly. This leads to the inverse-U-shaped curve result that Posen and Levinthal (2012) show (please see Section 1.3 in the Online Supplement for a more detailed explanation).

It is worthwhile noting two issues: (a) it is unclear whether both mechanisms of degrading knowledge and belief flattening are necessary for this inverse-U-shaped effect to occur, or whether the former is sufficient; and (b) this trade-off between knowledge and exploration cost is sensitive to how beliefs are formed, and choices made, and may change the shape of Figure 2 (cf. LiCalzi & Marchiori, 2013). We discuss this in more detail in the Online Supplement, Section 1.4.[14][15]

4.1.2 | Nonevaluative process

Our second control process utilizes the ϵ-*greedy* choice rule, which we interpret as a single decision maker controlling exploration budgets, with exploratory project selection occurring by lottery. The results in Figure 3 suggest that when the environment is subject to more frequent changes, increasing ϵ leads to better performance, unlike the *softmax* process that resulted in an inverted-U-shaped pattern. The difference between the two decision processes is that the decision to explore is sensitive to the difference in beliefs between the best versus other arms in *softmax*, but insensitive in ϵ-*greedy*. Therefore, the *softmax* process is highly sensitive to belief flattening leading to endogenous exploration, which is counteracted by reducing the strategy (τ). The ϵ-*greedy* process is less sensitive to belief flattening. This leads to increasing knowledge (from exploitation learning. since the best believed arm changes often with belief flattening), but simultaneously limits the increase exploration cost (since number of exploration pulls is controlled by ϵ regardless of beliefs).[16] Therefore, increasing ϵ leads to better performance.

4.2 | Strategies in dynamic environments—Distinguishing the two facets of exploration

Next, we present results for our two-parameter models, which is our novel contribution to the literature. Here, we distinguish between two facets of exploration—how much to explore or *exploration propensity* (α) from the region of exploration or *exploration breadth* (β). As discussed in the modeling section, here we use different choice rules (ϵ-*greedy* vs. *softmax*) for both parameters (α and β) to distinguish between different decision processes in firms.

[14]In the Online Supplement, Section 1.4, we try to independently manipulate knowledge degradation and belief flattening to understand whether both these mechanisms are necessary or only the former is sufficient to explain this curve. Our experiments suggest that both mechanisms are necessary. When we "turn off" belief flattening by reducing the agent's reliance on old feedback, the inverse-U-shaped effect no longer occurs.

[15]Although our explanation is plausible, it is not "proof." We thank an anonymous reviewer for drawing our attention to the imprecision in the mechanism underlying this result.

[16]Unlike the softmax process that is sensitive to the relative difference in attractiveness between the alternatives, here exploration increases because different alternatives are considered the most attractive at different time periods. Thus, in this process, the increase in exploration as beliefs become closer is much smaller. A consequence is that this process can "settle" quickly on one (above-average) choice, which may not be useful in dynamic environments.

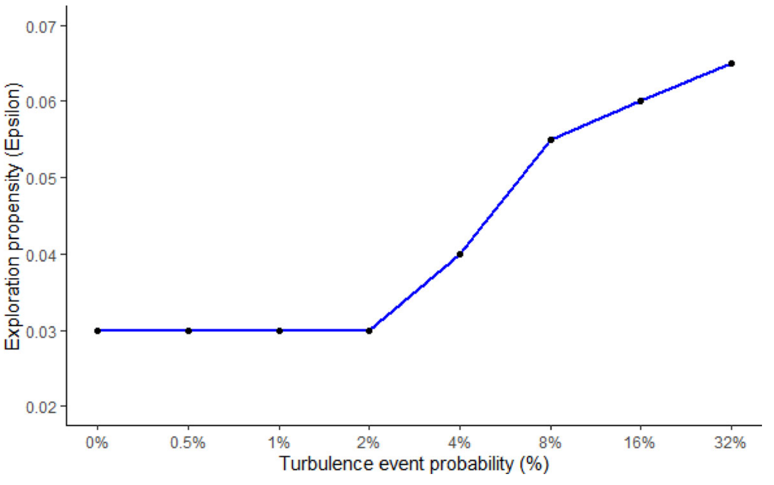

FIGURE 3 Exploration strategy in dynamic environments (using ϵ-greedy rule).

Figure 4, panels a & b, c & d, and e & f, plots the α and β values that lead to maximum performance at each level of environmental dynamism for the *evaluative*, *standard*, *and lottery* decision processes, respectively.[17] The left panels (a, c, and e) show how the optimal exploration propensity (α) changes, whereas right panels (b, d, and f) show changes in optimal exploration breadth (β) (recall, we label the strategy at which performance is highest at a given turbulence level the "optimal" strategy at that turbulence level).

We find that how the optimal *exploration propensity* (α) changes with increasing turbulence critically depends on the decision process. In the *evaluative* decision-process, which is the model closest to prior work (i.e., Lee & Puranam, 2016; Posen & Levinthal, 2012), we see that with increasing turbulence optimal α *decreases* (see Figure 4, panel a). By contrast, in the *standard* decision process the optimal α remains *constant* (see Figure 4, panel c), whereas in the *lottery* decision process the optimal α *increases* (see Figure 4, panel e). In contrast, *exploration breadth* (β) *increases* with increasing turbulence in all our models, please see panels b, d, and f in Figure 4. This figure provides evidence for our thesis that how firms explore as the environment becomes more turbulent is dependent on the decision process, that is, whether it is hierarchical and whether it is evaluative. Below, we provide some intuition into the underlying mechanisms, although in such a complicated model, the mechanism may not be fully explained.[18]

4.2.1 | Evaluative model

We start here since it is the direct extension to Posen and Levinthal (2012). From Figure 4, panels a and b, we note that optimal α decreases whereas optimal β increases in the two-parameter model with increasing η, when compared to the inverse-U effect for τ in the single parameter model. To understand the mechanism behind this result, let us first consider what happens when one of the two parameters is held constant, and the other one changes. In this

[17]We omit describing the *softmax-ϵ-greedy* combination since we could not easily find an empirical parallel.

[18]We thank an anonymous reviewer for pointing out the difficulty in explaining the result in the single-parameter model and suggesting that it may be very difficult to prove the complete mechanism in a two-parameter model.

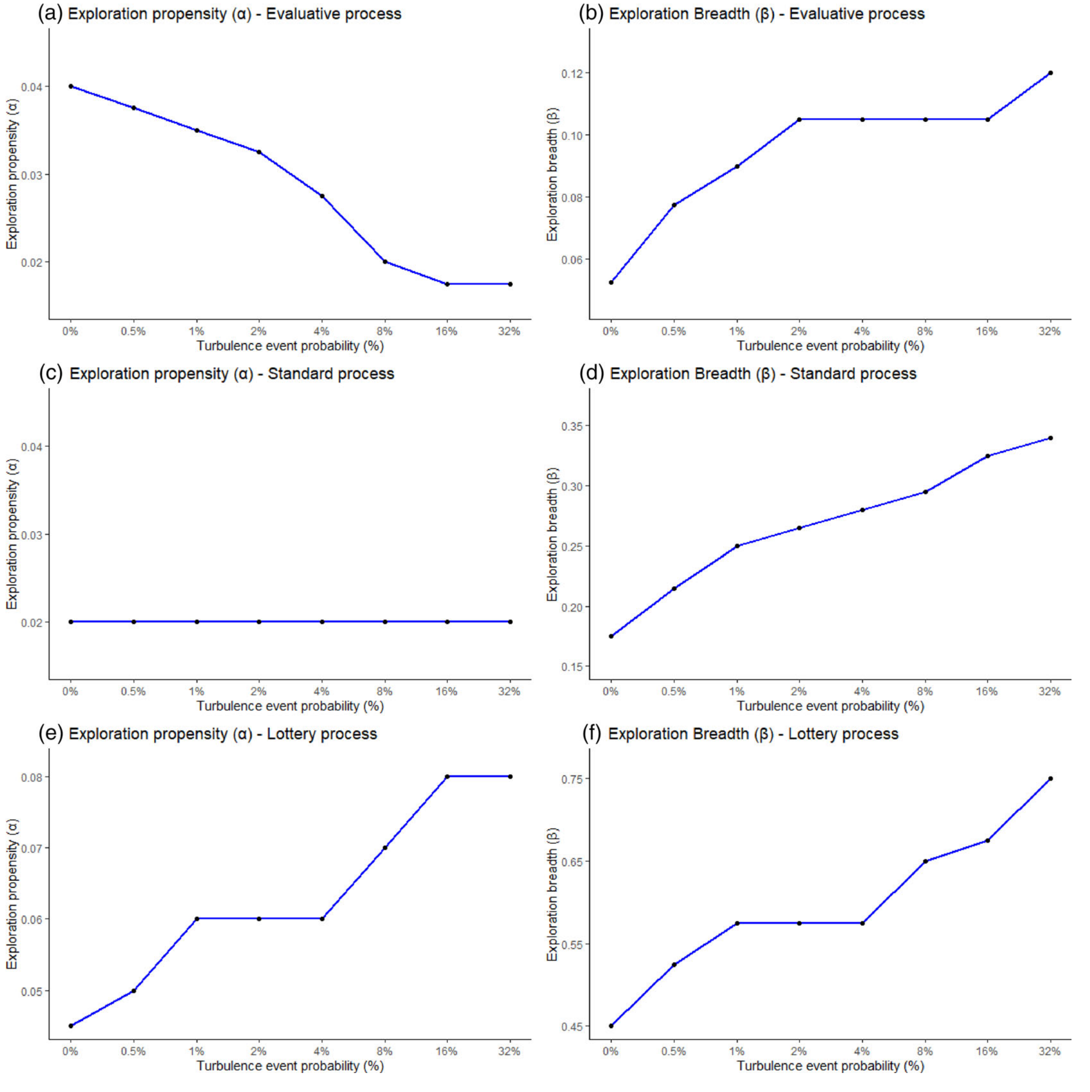

FIGURE 4 Exploration propensity and exploration breadth at best performance. This figure depicts optimal exploration propensity (α) for the *evaluative*, *standard*, and *lottery* decision processes in panels a, c, and e, respectively, and optimal exploration breadth (β) in panels b, d, and f, respectively. Each point on this figure represents the combination of exploration propensity and breadth that results in the highest performance, given the turbulence frequency. Note that they cannot be plotted on the same graph since the different functions employed mean the Y-axis values across plots are not readily comparable.

case, we should expect the inverse-U-shaped effect we saw from the single parameter model (see Figure 5, panels a and b). This is because both α and β affect the knowledge and exploration cost curves; therefore, when one parameter is held constant, the optimal value of the other parameter occurs when the increase in knowledge from additional exploration is matched by the increase in exploration cost.

To understand how the two parameters jointly determine performance, we first start with the case when α is held constant and β varies. As β increases, the agent is more likely to choose arms believed to be less attractive, increasing the agent's knowledge. With increasing

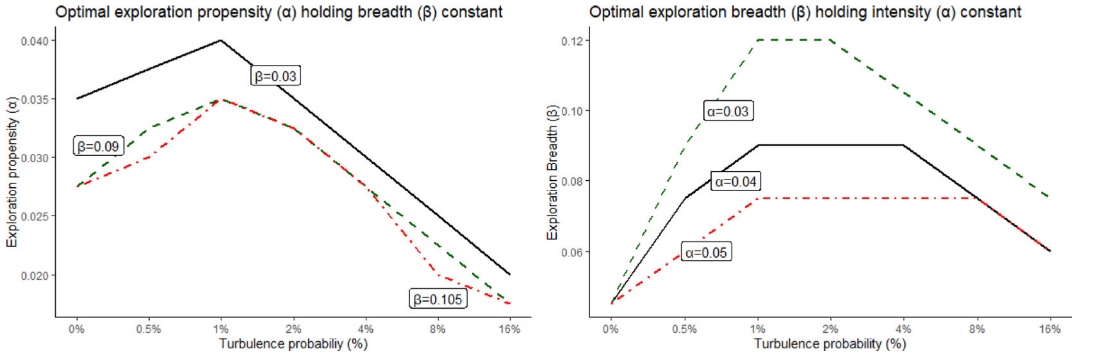

FIGURE 5 Optimal exploration propensity and breadth when holding the other parameter constant. For the left panel, for each η we vary α and we measure the cumulative performance of the agent at the end of the simulation (holding β constant at $\beta \in \{.03, .09, .105\}$). The α value for which maximum performance is achieved is plotted in this graph. For the right panel, similarly, we vary β and we measure the cumulative performance of the agent at the end of the simulation (holding α constant at $\alpha \in \{.03, .04, .05\}$). The β value for which maximum performance is achieved is plotted in this graph.

turbulence, an additional increment in β yields less knowledge (as in the single parameter model, turbulence degrades the value of exploration). Thus, to maintain knowledge levels, β needs to increase more with increasing turbulence.

However, increasing β simultaneously increases exploration cost, and it increases more as turbulence increases. Recall that in the single parameter case, τ reduces with increasing turbulence to counteract the increase in exploration costs from increasing endogenous exploration (and we see the same pattern with β when α is held constant, see Figure 5, panel b). However, in the two-parameter case, the increase in exploration cost can be arrested by tuning exploration propensity (α), since β explores only when α allows it to explore. As we discussed in the single parameter case, the increase in endogenous exploration is the prime culprit in increasing exploration costs as turbulence increases. This endogenous exploration can be arrested by lowering exploration propensity (α).[19][20] When α is low, it curtails how effective β is in increasing knowledge. In Figure 5, panel b, we see that when α is held constant at a higher value, the corresponding optimal β is smaller than when α is held constant at a lower value (similarly when β is held constant; Figure 5, panel a). Therefore, for the same level of knowledge gains from exploration, when α is lowered, β needs to increase even more. We further flesh out this intuition in the Online Supplement, Section 2.3.

4.2.2 | Standard process

In the *standard process*, the first stage decision (α) is performed in a nonevaluative manner and is implemented using a ϵ-*greedy* function whereas in the *evaluative process* we use a *softmax*

[19]Please see a more complete explanation in Section 1 of the online supplement for the single parameter model, which applies here to how α changes as well.

[20]Reducing β to control exploration costs while increasing α is not efficient. This is because, a low β only chooses the second best-believed arm, which may or may not be very good. Therefore, it neither controls per-pull cost nor cumulative exploration cost in frequently changing environments. In addition, it does not increase knowledge much.

function. Recall that the principal difference between them is that the former is insensitive to the relative attractiveness of the options, whereas the latter is very sensitive to it. This change has an important effect on how the agent explores.

The evaluative agent explores a lot initially, when all options appear equally attractive; as this agent's knowledge increases, they explore less and less, and settle on sampling only the best few arms. Increasing turbulence resets this behavior and gets the agent to explore more without changing strategy. The nonevaluative agent, in contrast, explores at a constant rate, regardless of how close in value the options are, and therefore, their exploration is comparatively much less sensitive to their knowledge. Therefore, turbulence has a markedly smaller effect on how much they explore (see Table 1 for an illustration, when comparing $\eta = 0$ with $\eta = 2$, the number of arm changes increases 87% for the evaluative process, but only 8.5% for the nonevaluative process). In other words, the increase in exploration is much lower for the agents following the *standard process* compared to those following the *evaluative process*. This change therefore has an important effect on how their optimal strategy changes with more frequent turbulence.

When turbulence increases, the agent needs to update their knowledge, accomplished by sampling arms they believe have lower attractiveness. Recall, in the *standard process*, the second stage is evaluative (same as the *evaluative process*). As β increases, it samples arms widely, leading to increasing exploration cost. As beliefs readjust to shocks, it makes arms look more equal (even if only temporarily in infrequent shock conditions). Since the *softmax* choice process is sensitive to the relative values of arms, at any value of α, the agent explores more (chooses more exploration pulls) in the *evaluative process*. This increase in exploration pulls further increases exploration cost in the *evaluative process*, and as we saw earlier, optimal α reduces to reduce this cost. In the *standard process*, since the first-stage is value insensitive, this "feedback loop" is much smaller. Therefore, exploration cost does not increase as much, and optimal α (as ϵ) need not reduce.

In sum, knowledge gains from increased β compensate for the increased exploration cost; thus, optimal α is likely to stay constant or rise very slightly in the *standard process* as

TABLE 1 Exploration pulls and arms changed in single parameter bandit.

	Softmax			ϵ-Greedy		
Turbulence %	**Tau**	**Exploration pulls**	**Arms changed**	**Epsilon**	**Exploration pulls**	**Arms changed**
0	0.04	324.19	467.70	0.05	250.06	394.03
1	0.04	484.71	699.24	0.05	250.00	411.51
2	0.04	608.72	871.55	0.05	250.04	488.27
4	0.04	781.21	1104.91	0.05	250.00	456.20
8	0.04	1000.94	1391.98	0.05	259.17	506.54
16	0.04	1224.44	1674.58	0.05	250.01	602.72

Note: This table depicts total exploration pulls and arm changes as a result of changes in beliefs in the single parameter 10-arm normal bandit model using softmax (evaluative model) or ϵ-greedy (standard model). The true mean of the arms is drawn from the Normal (0.5, 0.05) distribution, and for each draw, the variance of the arm is 0.05. The table presents cumulative results at the end of a 500-period simulation, with 10 resources allocated per period. For each resource, an individual exploration/exploitation choice was made, using the agent's beliefs at the beginning of the period. The exploration pulls column records how many total exploratory pulls the agent chose to make. The arms changed column records a sum of the exploratory pulls and the number of times the best believed arm has changed during the simulation.

turbulence becomes more frequent. At extreme turbulence, we expect α to rise because the second stage needs to explore more to generate adequate knowledge, which it cannot do unless α (ϵ) increases.[21][22]

4.2.3 | Lottery process

The mechanism explained so far suggests that both optimal α and optimal β should increase with more frequent turbulence in the lottery process. In this process, both stages of the decision process follow an ϵ-greedy process, which means both stages are value insensitive. The second stage being value-insensitive does not explore as much to generate knowledge about the different arms (unlike the standard process) when turbulence increases. Therefore, for the downstream agent to generate adequate knowledge, they need to be allowed to explore more often, that is, α needs to increase.

4.2.4 | Performance consequences

Which choice rule results in the highest performance? From Figure 6, we note that the two-stage models outperform the one-stage models, and the *evaluative* process outperforms the *standard* process across all levels of turbulence. This happens due to two reasons. First, having the ability to de-couple the exploration propensity and exploration breadth give a decisive advantage to two-stage models over simpler models. This is because now each decision can be fine-tuned in the two stage models, rather than a single decision trading-off the two competing requirements. We note that the advantage is larger as turbulence increases, that is, when the trade-off between the two decisions is most critical. Second, the *evaluative* process performs better than the *standard* process, since it is able to use information more efficiently, to explore more when the choices appear similar, but explore less when the agent believes some choices are clearly superior to others. However, the evaluative process is also cognitively more demanding, and as we saw, can lead to over-exploration (see the Online Supplement, Section 3.2 for more detail).

Finally, we examined how deviations from optimal strategies influence firm performance. In the Online Supplement (see Section 3.3) we show that at low η it is more important to correctly tune β, while ensuring α is within a range of acceptable values, while at high η the importance of the parameters reverses, especially for the evaluative process.

[21]This explanation suggests that endogenous exploration from the homogenization of beliefs is an important part of the explanation for the evaluative process. This also suggests that the optimal exploration with more frequent turbulence for the evaluative process when managers discount prior knowledge heavily should look more like the baseline standard process. At extreme discounting, it should look like the baseline lottery process. We check this in the robustness section and find that our conjecture holds true.

[22]We note that in our design, the firm invests 10 resources every round and thus receives 10 data points every round. Since the exploration decision is made for each resource invested, it is more likely the firm using a softmax process receives new data every round, sufficient to make good investment decisions. We show in the sensitivity analysis that when the firm only invests one resource per round (instead of 10), the optimal α increases at higher levels of turbulence also for the *standard* decision process.

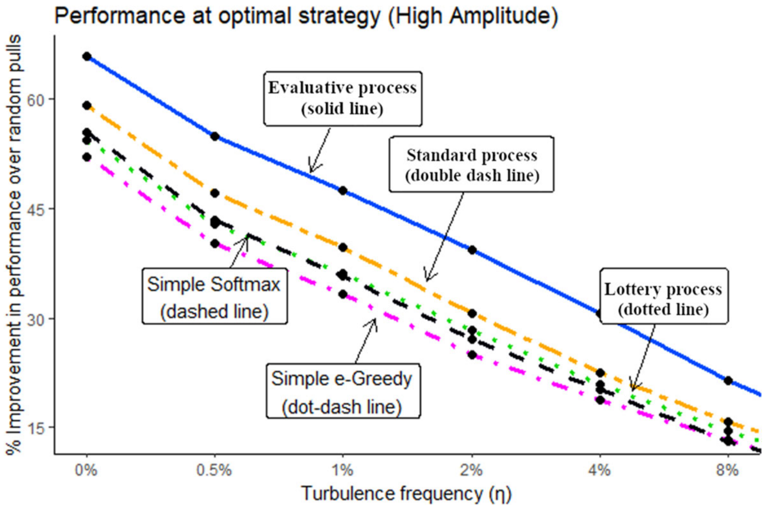

FIGURE 6 Relative performance of decision-making processes in dynamic environments. This figure depicts percentage improvement in performance relative to a firm that choses alternatives completely random. As Posen and Levinthal (2012) noted that performance is invariant to strategy at high levels of turbulence, we note that it is also invariant to decision-making processes at turbulence levels above 8%.

5 | ROBUSTNESS CHECKS

We checked the sensitivity of our results to different modeling choices. Especially important assumptions in our model include choices on (a) the nature of the task environment, (b) the belief formation process, especially how soon managers discard old information (c) how many resources managers allocate per period, and (d) the nature of turbulence, esp. the amplitude of changes that accompany a shock. Therefore, we checked robustness of our main result—optimal β increases with more frequent turbulence, but the change to optimal α is sensitive to the decision-making process—to these different assumptions. While we briefly present some results here, we provide more detail in the Online Supplement.

5.1 | Changing the task environment

We altered the distribution that the underlying value of the arm is drawn from. Instead of drawing the arms from a normal distribution (N[0.5, 0.05]), we drew the underlying value from a Beta(2,2) distribution, which has been widely used in bandit models (i.e., Lee & Puranam, 2016; Posen & Levinthal, 2012). In this case, feedback for each pull results in a Bernoulli draw, with the probability of success equal to the true value of the arm. Since the normal distribution provides more continuous feedback compared to the Bernoulli distribution, the feedback in our baseline specification is closer to the arm means and therefore more conservative. Our results are qualitatively similar to what we find in our baseline specification.

5.2 | Changing the relative importance of recent versus older feedback

An important implicit assumption of our models is that managers are blind to environmental change and thus equally value recent and old experience, similar to prior work (Posen & Levinthal, 2012). How much managers are prone to inertia is hotly debated theoretically. On the modeling side, predictions of bandit models with infinite memory are sensitive to the duration of the experiment (Denrell, 2005), and overweight stale information in dynamic environments (LiCalzi & Marchiori, 2013). Thus, it is important to examine how robust our findings are to different assumptions about managerial inertia modeled as how much they discount memory. It is also important to examine in our context since one of the key mechanisms at play both in our model and the Posen and Levinthal (2012) model, belief flattening, is critically influenced by this assumption.[23] To evaluate robustness, we discount experience k_i in Equation (2). Following Christensen et al. (2021), we calculate the value of k_i based on the discount parameter γ:

$$k_{i,t} = 1_{\{i\}}(I_t) + \gamma * k_{i,t-1} \tag{3}$$

where $1_{\{i\}}(I_t)$ takes the value 1 when arm i was pulled at time t, and 0 otherwise (see also Kocsis & Szepesvári, 2006). For alternatives that are not selected in a given period $q_{i,t+1} = q_{i,t}$.[24] In the baseline model, we set $\gamma = 1$, that is, classic averaging. In subsequent experiments, we used discount factors $\gamma \in \{0.99, 0.95\}$, to better understand the value of forgetting. This is an important check since model dynamics is likely to be very sensitive to the belief formation process as described previously using the parameter γ in Equation (3). The higher the discounting, the more the managers weigh recent feedback as more important than old experiences. Increased discounting creates sharper beliefs and therefore reduces endogenous exploration.

Figure 7 indicates how optimal α (panels a, c, and e) and β (panels b, d, and f) for the three decision processes change as firms discount old experience (for $\gamma = 0.99$ and $\gamma = 0.95$), compared to the model with "no discounting" ($\gamma = 1$). Our first observation is that the shape of the curve, for α and β are robust to discounting old experience. In the *evaluative process*, α decreases more slowly, the higher the memory decay. In the *standard* and the *lottery* processes, optimal α increases with higher memory decay. In all three processes, optimal β increases with increasing memory discounting. We explain these results in Section 2.4 of the Online Supplement.

[23]In Section 1.4 of the Online Supplement, we examine the robustness of Posen and Levinthal (2012) model to this assumption and use it to gain greater clarity on the mechanisms underlying their finding. This exercise then extends to our model as well in Section 2.4.

[24]Even when the arm i is not selected in period t, we discount the value of k_i using the parameter γ. Consider, for example, when an arm is selected three times, in periods 1, 2, and 75. At the end of period 75, k_i will be 3 when $\gamma = 1$, 1.955 when $\gamma = 0.99$, and 1.046 when $\gamma = 0.95$. As a result, the value of the fraction $\frac{1}{k_i+1}$ which represents how much, which represents how much weight this third feedback will have on the updated belief of the agent, respectively.

5.3 | Changing the resource allocation format

Instead of allocating 10 resources every period, we repeat our analysis by allocating just one resource per period.[25] When 10 resources are allocated per period, the feedback received is aggregated over multiple trials and therefore more accurate. Thus, the underlying mechanism should exacerbate changes in the parameters when only one resource is allocated per period. We observe this pattern for both the α and β parameters (please see Sections 2.3 and 3.1 of the Online Supplement).

5.4 | Turbulence amplitude

Next we discuss how the decisions regarding how much to explore and where to explore change with different levels of turbulence amplitude. The lower the amplitude of change, the less abrupt the change in customer preferences. We observe the same pattern in the change of our parameters α and β, only at a lower slope; that is, α decreases more gently, and β increases more gently. This happens because when the shock preserves a portion of the prior value of the arm, it destroys knowledge to a lesser extent. We replicate Figures 4 and 7 in Section 3.1 of the Online Supplement at lower values of λ ($\lambda = 0.5$ and $\lambda = 0.2$; baseline $\lambda = 1$).

5.5 | Changing initial beliefs

In the next test, we checked robustness to initial beliefs. Instead of assigning homogenous beliefs about all arms, we checked robustness to (a) accurate beliefs and (b) random beliefs.[26] In the first case, the agents' initial beliefs for the arms are set to the true mean values of their payoff distribution; that is, managers knew their choices perfectly before turbulence occurs. In the second case, the agents' initial beliefs for the arms are set to a random value. In both cases, the shape of the curve is similar for α and β to the baseline. All results remain robust (see Online Supplement, Section 3.4).

5.6 | Changing the number of parameters

Finally, while each of our experiments outperforms the same experiment with a single exploration parameter, it is plausible that adding more parameters to describe exploration may further improve performance. We did not find a benefit of adding a third parameter. We explain this in more detail in the Online Supplement (Section 3.5).

[25]When allocating one resource per period, we extended the time horizon of the simulation to 1500 periods (thus allocating 1500 total resources, compared with 5000 total resources allocated when the per period allocation was 10 resources). To ensure results are not biased by the total resource allocation, we also repeat the simulation with shorter time horizons (50, 100, 150 periods) for the model with 10 resources allocated per period. Results are qualitatively the same and are available from the authors upon request.
[26]We thank an anonymous reviewer for suggesting this robustness check.

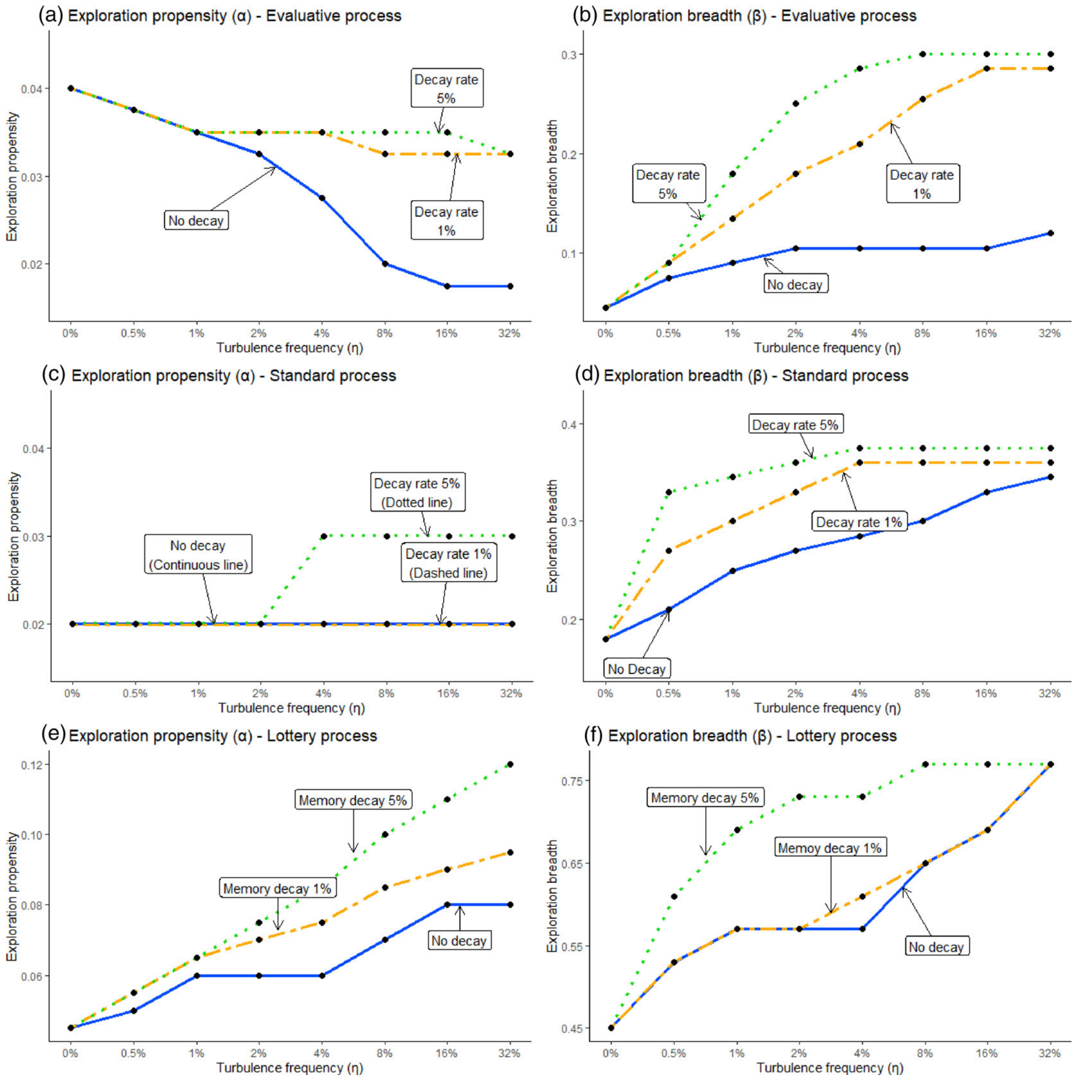

FIGURE 7 Exploration propensity and exploration breadth at best performance for firms with discounting older experience. This figure depicts optimal exploration propensity (α) for the *evaluative*, *standard*, and *lottery* decision processes in panels a, c, and e, respectively, and optimal exploration breadth (β) in panels b, d, and f, respectively. Each point on this figure represents the combination of α and β that results in the highest performance, given the turbulence frequency(η) and memory discounting level (γ), at high turbulence amplitude ($\lambda = 1$).

5.7 | Limits and extensions

As with all research, our model has various limitations. Interdependencies across arms may influence how exploration strategies develop. We recognize that the presence of such links may introduce additional relevant complications our model cannot address. A possibly fruitful extension could endogenize exploration strategies, given that managerial behavior may be shaped by organizational slack, or performance aspirations. Another extension of this research could consider how exploration intensity and breadth strategies may vary in different hierarchies, which may be more or less centralized, and where information may be summarized more or less when it crosses hierarchical boundaries. In addition, in our model, cumulative performance difference between the non-optimal and optimal parameters is small in absolute terms, although sizeable in relative terms when compared with the baseline of random choice. However, "effect sizes" in models are not always indicative of "effect sizes" in empirical research. Current empirical work suggests that changes in exploration breadth meaningfully alters firm performance, although we do not know how it changes in conjunction with changes in exploration propensity.

Our second limitation concerns the mechanism. Although we have explained the intuition behind the mechanism—that the two-parameter model is able to regulate the tradeoff between exploration cost and knowledge gains more precisely compared to the single parameter model, this may not be the only mechanism in play. This is a complex model, and we note that the precise mechanism underlying the single-parameter model results in Posen and Levinthal (2012) that we build on is imprecisely understood. Thus, although we propose an explanation, it may need to be supplemented with other mechanisms for a complete explanation of the patterns we observe. Despite this limitation, we note that the model advances research in two important ways. First, it is more indicative of hierarchical decision-making in organizations, and suggest this process is more efficient at adaptation than what the unitary actor model would suggest. Second, it suggests that how the firm adapts to environmental turbulence is heavily dependent on its decision-making process, clarifying the importance for controlling for the process in future empirical and theoretical work in this domain.

6 | DISCUSSION AND CONCLUSIONS

In this article, we examine how optimal strategies for exploration change as the environment becomes more turbulent (or dynamic). Despite important advances in our knowledge regarding how firms balance exploration and exploitation, little consensus exists on how much firms should explore (vs. exploit) under different levels of environmental change. Prior work has attempted to reconcile these mixed predictions and findings by understanding how different facets of environmental dynamism influence optimal exploration strategies (Stieglitz et al., 2016).

We departed from prior work by unpacking the unitary actor model. We argued that explore-exploit decisions in firms are separated both temporally and hierarchically. Thus, we consider how different facets of exploration—exploration propensity versus exploration breadth—need to change as the environment becomes more turbulent. Disentangling these different facets brings to the forefront the idea that decisions to explore versus exploit are organizational—different members of the organization undertake them, and perhaps follow different decision processes that are black boxed by the unitary actor model. In line with this objective, we examine three different decision processes in how firms make explore-exploit decisions and understand how they need to change for superior performance in turbulent environments.

The results from our model suggest that as turbulence increases, firms need to reduce, keep constant, or increase their exploration propensity, depending on their decision process. However, in all cases firms simultaneously need to increase their exploration breadth strategy. These baseline findings were remarkably stable to different assumptions. This secular trend becomes much more pronounced when the environment changed both frequently and unpredictably.

Surprisingly, these trends were also stable to differing assumptions about managers' inertia in decision-making, modeled as how much they discount historical feedback as the environment becomes more dynamic, and to how well they know their environments before turbulence occurred.

Our model also suggested some nuanced understanding of these baseline trends. We predicted that when the environment is changing frequently, it is more important to identify the correct level of exploration propensity, with exploration breadth set at a fairly high level. This complexity may be mitigated for firms utilizing the standard process since a simple but accurate "gatekeeper" for firm exploration improves search performance.[27] In contrast, when the environment changes only infrequently, it is more important to pay attention to how widely the firm is exploring, since too narrow a search can fail to identify good opportunities but too broad a search can lead to wasted effort that discounts the firm's knowledge base. In this case, identifying the optimal exploration propensity takes somewhat of a back seat. Thus, our enhanced model provides clearer and more actionable predictions compared to prior work.

These results offer an additional pathway to reconcile the mixed findings in prior work on how firms should adapt to environmental change, as well as make new predictions that we can test in future empirical work. We reconcile predictions from prior theoretical work by reconciling the different assumptions they make about environmental turbulence. For example, Eisenhardt and Martin (2000) and Keller and Rady (1999) assume "high velocity" environments that change fast and in unpredictable ways, and caution firms against over-exploring in these environments, but do not compare them to how firms should respond when the environment changes only infrequently but can be competence enhancing or destroying. Kim and Rhee (2009) focus on exploration breath, assuming the firm always explores, and find that increased breadth is beneficial in high turbulent environments. These results can be thought of "special cases" of our more "general" treatment, which allows us to make further predictions about optimal strategies in different environments.

Turning to empirical work, many studies that suggest the importance of increasing exploration with higher turbulence have measured exploration using a breadth or variety measure (e.g., Garg et al., 2003; Klingebiel & Joseph, 2016; Nadkarni & Narayanan, 2007). Schilke (2014) suggests an inverse-U-shaped effect and Barnett and Freeman (2001) found a negative effect using measures more related to the "how much to explore" dimension. Interestingly, we find almost no published papers using traditional measures of exploration, such as R&D budgets or R&D intensity on performance in more versus less turbulent environments. Our finding is that this relationship is very sensitive to the decision-making process, which is often not controlled for. Thus, this lack of studies is perhaps not surprising given well-known publishing biases such as not publishing null results and the file-drawer problem (Ioannidis, 2005; Rosenthal, 1979).

In addition, our model speaks to changes in exploration strategy, but not exploration behavior. Exploration behavior, as in number of exploration choices, may increase even when exploration strategy decreases (see Posen & Levinthal, 2012 for a detailed discussion of these effects). Thus, relationship between performance and measures of behavior such as new product introduction may differ from measures of strategy such as R&D intensity or TMT incentive contracts for exploration. However, there is a lack of empirical work that explicitly measures how facets of exploration strategy, such as incentive contracts or introducing a culture of innovation, needs to change in turbulent environments. Thus, although our theoretical work exploring the contingent effect of the type of exploration has important implications for future empirical work in

[27]We thank an anonymous referee for pointing this simple way of communicating our result.

better understanding the relationship between exploration and performance, we caution against over-interpreting our results directly against current empirical work, especially our results on exploration propensity (our results suggest exploration breadth always increases, making it a little less problematic to equate strategy and behavior when interpreting empirical work).

In addition, our findings also shed some light on the mechanisms underlying why the structural mode of achieving ambidexterity may be superior to the contextual mode, as speculated by O'Reilly and Tushman (2013). They argue that the contextual mode often fails when the environment radically changes because employees rarely have the authority to significantly deviate from their mandated activities, but senior management who hold the reins can make that possible in the structural mode (see also Benner, 2009).

Our model suggests another explanation: cognitively, individuals may not be able to balance the dueling effects of exploration. By distinguishing between the two facets of exploration, our model suggests that one of these facets decreases or remains constant simultaneously when the other facet increases, in the more common decision situations. Recent empirical work in psychology shows that a single parameter (the exploration temperature in the softmax equation) fits individual behavior very well (Daw et al., 2006). The upshot of this finding is that when individuals explore more, they simultaneously explore more broadly, which hampers firms that rely on individual adaptation efforts. Organizations, however, can transcend this limitation by structurally separating exploration activities from exploitation activities; either by conducting them in different units (Tushman & O'Reilly, 1996) or even separating them across firm boundaries (Lavie et al., 2011; Stettner & Lavie, 2014). Thus, they can control propensity separately from breadth. Examples of this approach include open innovation efforts such as the P&G Connect and Develop program and the acquisition-reliant product development at Cisco.

Our model also offers the additional insight that such approaches are valuable even when the environment is not changing radically but changes often. Frequent shocks, even though incremental, change the environment significantly over a period of time (McKendrick & Wade, 2009). Thus, our results suggest that contextual ambidexterity is most appropriate when the environment undergoes only infrequent changes and these changes do not significantly alter the relative attractiveness of firms' capabilities; in other words, when they are not competence-destroying (Tushman & Anderson, 1986). Our model therefore offers an explanation for why innovation approaches such as the 80/20 rule—one means of achieving contextual ambidexterity—provide benefits for some firms such as 3M (Govindarajan & Srinivas, 2013) but appear to backfire for others like Google (D'Onfro, 2015; Solomon, 2016; Towsend, 2013).

Our model offers insights on the core question of how organizations should adapt in dynamic environments and specifically points out the important role of pursuing broad innovation strategies. Our findings are in line with significant practitioner concerns about encouraging radical innovation in their organizations (Criscuolo et al., 2017; Govindarajan & Srinivas, 2013) and further argues that optimal exploration propensities for firms are dependent on the decision-making processes they employ. The contrasting empirical results and the different advice offered in the business press on how managers' should tackle the turbulence problem has resulted in some confusion about how best to tackle this challenge. Our model offers a theoretical basis for reconciling these ideas and provides a framework for future empirical research.

ACKNOWLEDGEMENTS

We would like to thank Oliver Baumann, Jerker Denrell, Thorbjorn Knudsen, Dan Levinthal, Hart Posen, Phanish Puranam, George Zheng, and seminar audiences at the University of Southern Denmark and the TOM conference for many constructive comments on previous versions of this paper. We would also like to thank the associate editor Prof. Martin Ganco, and two anonymous reviewers for constructive comments that elevated this paper over two rounds of review. All errors remain our own. The authors contributed equally and are listed alphabetically.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Kannan Srikanth https://orcid.org/0000-0002-5900-4503
Tiberiu Ungureanu https://orcid.org/0000-0001-6820-9298

REFERENCES

Agrawal, S., & Goyal, N. (2012). Analysis of Thompson sampling for the multi-armed bandit problem. In *Conference on learning theory*. JMLR Workshop and Conference Proceedings.

Amburgey, T. L., Kelly, D., & Barnett, W. P. (1993). Resetting the clock: The dynamics of organizational change and failure. *Administrative Science Quarterly*, *38*(1), 51.

Azoulay, P., & Li, D. (2020). Scientific grant funding. In A. Goolsbee & B. F. Jones (Eds.), *Innovation and public policy* (pp. 117–150). University of Chicago Press.

Barnett, W. P. (1994). The liability of collective action: Growth and change among early telephone companies. In J. V. Singh & J. A. Baum (Eds.), *Evolutionary dynamics of organizations* (pp. 337–354). Oxford University Press.

Barnett, W. P., & Freeman, J. (2001). Too much of a good thing? Product proliferation and organizational failure. *Organization Science*, *12*(5), 539–558.

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

Benner, M. J., & Tushman, M. L. (2003). Exploitation, exploration, and process management: The productivity dilemma revisited. *Academy of Management Review*, *28*(2), 238–256.

Bhardwaj, G., Camillus, J. C., & Hounshell, D. A. (2006). Continual corporate entrepreneurial search for long-term growth. *Management Science*, *52*(2), 248–261.

Billinger, S., Srikanth, K., Stieglitz, N., & Schumacher, T. R. (2021). Exploration and exploitation in complex search tasks: How feedback influences whether and where human agents search. *Strategic Management Journal*, *42*(2), 361–385.

Bourgeois, L. J., & Eisenhardt, K. M. (1988). Strategic decision processes in high velocity environments: Four cases in the microcomputer industry. *Management Science*, *34*(7), 816–835.

Bourla, A. (2021). The CEO of Pfizer on developing a vaccine in record time. *Harvard Business Review*, *3*(99), 34–39.

Brown, S. L., & Eisenhardt, K. M. (1997). The art of continuous change: Linking complexity theory and time-paced evolution in relentlessly shifting organizations. *Administrative Science Quarterly*, *42*(1), 1.

Bubeck, S., & Cesa-Bianchi, N. (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, *5*(1), 1–122.

Camerer, C., & Ho, T. H. (1999). Experience-weighted attraction learning in normal form games. *Econometrica*, *67*(4), 827–874. https://doi.org/10.1111/1468-0262.00054

Christensen, M., Sørensen, S. K., & Knudsen, T. (2021). Tax bandits: Learning in an interactive environment. University of Southern Denmark Working Paper.

Criscuolo, P., Dahlander, L., Grohsjean, T., & Salter, A. (2017). Evaluating novelty: The role of panels in the selection of R&D projects. *Academy of Management Journal*, *60*(2), 433–460.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876–879.

Denrell, J. (2005). Why most people disapprove of me: Experience sampling in impression formation. *Psychological Review*, *112*(4), 951–978.

Denrell, J., & March, J. G. (2001). Adaptation as information restriction: The hot stove effect. *Organization Science*, *12*(5), 523–538.

D'Onfro, J. (2015). The truth about Google's famous '20% time' policy. Business Insider. Retrieved from https://www.businessinsider.com/google-20-percent-time-policy-2015-4

Eisenhardt, K. M. (1989). Making fast strategic decisions in high-velocity environments. *Academy of Management Journal*, *32*(3), 543–576.

Eisenhardt, K. M., & Martin, J. A. (2000). Dynamic capabilities: What are they? *Strategic Management Journal*, *21*(10–11), 1105–1121.

Eisenhardt, K. M., & Tabrizi, B. N. (1995). Accelerating adaptive processes: Product innovation in the global computer industry. *Administrative Science Quarterly*, *40*(1), 84.

Gaba, V., & Greve, H. R. (2019). Safe or profitable? The pursuit of conflicting goals. *Organization Science*, *30*(4), 647–667.

Garg, V. K., Walters, B. A., & Priem, R. L. (2003). Chief executive scanning emphases, environmental dynamism, and manufacturing firm performance. *Strategic Management Journal*, *24*(8), 725–744.

Govindarajan, V., & Srinivas, S. (2013). The innovation mindset in action: 3M corporation. *Harvard Business Review*. Retrieved from https://hbr.org/2013/08/the-innovation-mindset-in-acti-3

Greve, H. R. (2003). A behavioral theory of R&D expenditures and innovations: Evidence from shipbuilding. *Academy of Management Journal*, *46*(6), 685–702.

Greve, H. R. (2007). Exploration and exploitation in product innovation. *Industrial and Corporate Change*, *16*(5), 945–975.

Hannan, M. T., & Freeman, J. (1984). Structural inertia and organizational change. *American Sociological Review*, *49*(2), 149. https://doi.org/10.2307/2095567

Heij, C. V., Volberda, H. W., & Van Den Bosch, F. A. J. (2014). How does business model innovation influence firm performance: The effect of environmental dynamism. *Academy of Management Proceedings*, *2014*(1), 16500.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124.

Jansen, J., van Bosch, F., & Volberda, H. (2005). Managing potential and realized absorptive capacity: How do organizational antecedents matter? *Academy of Management Journal*, *48*(6), 999–1015.

Jansen, J. J., Simsek, Z., & Cao, Q. (2012). Ambidexterity and performance in multiunit contexts: Cross-level moderating effects of structural and resource attributes. *Strategic Management Journal*, *33*(11), 1286–1303.

Jansen, J. J. P., Van Den Bosch, F. A. J., & Volberda, H. W. (2006). Exploratory innovation, exploitative innovation, and performance: Effects of organizational antecedents and environmental moderators. *Management Science*, *52*(11), 1661–1674.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, *93*(5), 1449–1475.

Karim, S., Carroll, T. N., & Long, C. P. (2016). Delaying change: Examining how industry and managerial turbulence impact structural realignment. *Academy of Management Journal*, *59*(3), 791–817.

Katz, R., & Allen, T. J. (1982). Investigating the not invented here (NIH) syndrome: A look at the performance, tenure, and communication patterns of 50 R & D project groups. *R&D Management*, *12*(1), 7–20.

Keller, G., & Rady, S. (1999). Optimal experimentation in a changing environment. *Review of Economic Studies*, *66*(3), 475–507.

Kim, T., & Rhee, M. (2009). Exploration and exploitation: Internal variety and environmental dynamism. *Strategic Organization*, *7*(1), 11–41.

Klingebiel, R., & Joseph, J. (2016). Entry timing and innovation strategy in feature phones. *Strategic Management Journal*, *37*(6), 1002–1020.

Kocsis, L., & Szepesvári, C. (2006). Machine learning: ECML 2006, 17th European Conference on Machine Learning Berlin, Germany, September 18-22, 2006 Proceedings. *Lecture Notes in Computer Science*, 282–293.

Larrañeta, B., Zahra, S. A., & González, J. L. G. (2014). Strategic repertoire variety and new venture growth: The moderating effects of origin and industry dynamism. *Strategic Management Journal*, *35*(5), 761–772.

Lattimore, T. (2016). Regret analysis of the finite-horizon Gittins index strategy for multi-armed bandits. In *Conference on learning theory* (pp. 1214–1245). PMLR.

Lavie, D., Kang, J., & Rosenkopf, L. (2011). Balance within and across domains: The performance implications of exploration and exploitation in alliances. *Organization Science*, *22*(6), 1517–1538.

Lavie, D., Stettner, U., & Tushman, M. L. (2010). Exploration and exploitation within and across organizations. *The Academy of Management Annals*, *4*(1), 109–155.

Lee, E., & Puranam, P. (2016). The implementation imperative: Why one should implement even imperfect strategies perfectly. *Strategic Management Journal*, *37*(8), 1529–1546.

Lee, S, Meyer-Doyle, P. (2017). How performance incentives shape individual exploration and exploitation: Evidencefrom microdata. *Organization Science*, *28*(1), 19–38. https://doi.org/10.1287/orsc.2016.1104

Levinthal, D. A., & March, J. G. (1993). The myopia of learning. *Strategic Management Journal*, *14*(S2), 95–112.

Li, S., Mayhew, S. D., & Kourtzi, Z. (2009). Learning shapes the representation of behavioral choice in the human brain. *Neuron*, *62*(3), 441–452.

LiCalzi, M., & Marchiori, D. (2013). *Artificial economics and self organization: Agent-based approaches to economics and social systems* (pp. 205–216). Springer International Publishing

Lu, Y., Meisami, A., Tewari, A., & Yan, W. (2020). Regret analysis of bandit problems with causal background knowledge. In *Conference on Uncertainty in Artificial Intelligence* (pp. 141-150).

Luce, R. D. (1959). On the possible psychophysical laws. *Psychological Review*, *66*(2), 81–95.

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

March, J. G. (1994). *Primer on decision making: How decisions happen*. Simon and Schuster.

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

Marino, A., Aversa, P., Mesquita, L., & Anand, J. (2015). Driving performance via exploration in changing environments: Evidence from formula one racing. *Organization Science*, *26*(4), 1079–1100.

Markovitch, D. G., & O'Brien, J. (2021). When capital markets discount R&D expenditures: The problemistic search effect. *Journal of Strategy and Management*, *14*(4), 444–460.

McKendrick, D. G., & Wade, J. B. (2009). Frequent incremental change, organizational size, and mortality in high-technology competition. *Industrial and Corporate Change*, *19*(3), 613–639.

Nadkarni, S., & Narayanan, V. K. (2007). Strategic schemas, strategic flexibility, and firm performance: The moderating role of industry clockspeed. *Strategic Management Journal*, *28*(3), 243–270.

Nelson, R. R., & Winter, S. G. (1982). *An evolutionary theory of economic change*. Harvard University Press.

North, M. J., & Macal, C. M. (2007). *Managing business complexity: Discovering strategic solutions with agent-based modeling and simulation*. Oxford University Press.

O'Reilly, C. A., & Tushman, M. L. (2013). Organizational ambidexterity: Past, present, and future. *Academy of Management Perspectives*, *27*(4), 324–338.

Posen, H. E., & Levinthal, D. A. (2012). Chasing a moving target: Exploitation and exploration in dynamic environments. *Management Science*, *58*(3), 587–601.

Rigby, D., & Zook, C. (2002). Open-market innovation. *Harvard Business Review*, *80*(10), 80–129.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641.

Ruef, M. (1997). Assessing organizational fitness on a dynamic landscape: An empirical test of the relative inertia thesis. *Strategic Management Journal*, *18*(11), 837–853.

Rumelt, R. (2011). *Good strategy bad strategy: The difference and why it matters*. Profile Books.

Schilke, O. (2014). On the contingent value of dynamic capabilities for competitive advantage: The nonlinear moderating effect of environmental dynamism. *Strategic Management Journal*, *35*(2), 179–203.

Sidhu, J. S., Commandeur, H. R., & Volberda, H. W. (2007). The multifaceted nature of exploration and exploitation: Value of supply, demand, and spatial search for innovation. *Organization Science*, *18*(1), 20–38.

Siggelkow, N., & Rivkin, J. W. (2005). Speed and search: Designing organizations for turbulence and complexity. *Organization Science*, *16*(2), 101–122.

Solomon, Y. (2016). Why Google's '20 percent' rule is actually useless. INC. Magazine. Retrieved from https://www.inc.com/yoram-solomon/20-of-my-time-will-not-make-me-more-creative.html

Stettner, U., & Lavie, D. (2014). Ambidexterity under scrutiny: Exploration and exploitation via internal organization, alliances, and acquisitions. *Strategic Management Journal*, *35*(13), 1903–1929.

Stieglitz, N., Knudsen, T., & Becker, M. C. (2016). Adaptation and inertia in dynamic environments. *Strategic Management Journal*, *37*(9), 1854–1864.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. MIT.

The Economist. (2022). *How to escape scientific stagnation*. The Economist. Retrieved from https://www.economist.com/finance-and-economics/2022/10/26/how-to-escape-scientific-stagnation

Towsend, C. (2013). 'Innovate or die': Why Google's 80/20 rule is a red herring. Customer Think online magazine. URL: https://customerthink.com/innovate_or_die_why_google_s_8020_rule_is_a_red_herring/ (Retrieved 7/15/2024)

Tripsas, M., & Gavetti, G. (2000). Capabilities, cognition, and inertia: Evidence from digital imaging. *Strategic Management Journal*, *21*(1011), 1147–1161.

Tushman, M. L., & Anderson, P. (1986). Technological discontinuities and organizational environments. *Administrative Science Quarterly*, *31*, 439.

Tushman, M. L., & O'Reilly, C. A. (1996). Ambidextrous organizations: Managing evolutionary and revolutionary change. *California Management Review*, *38*(4), 8–29.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232.

Wang, H., & Li, J. (2008). Untangling the effects of overexploration and overexploitation on organizational performance: The moderating role of environmental dynamism. *Journal of Management*, *34*(5), 925–951.

Wholey, D. R., & Brittain, J. (1989). Research notes: Characterizing environmental variation. *Academy of Management Journal*, *32*(4), 867–882.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Srikanth, K., & Ungureanu, T. (2025). Organizational adaptation in dynamic environments: Disentangling the effects of how much to explore versus where to explore. *Strategic Management Journal*, *46*(1), 19–48. https://doi.org/10.1002/smj.3646