7-2020

Learning and adaptation under uncertainty and ambiguity

Lei ZHENG
Singapore Management University

Learning and Adaptation under Uncertainty and Ambiguity

by

Zheng Lei (George)

Submitted to Lee Kong Chian School of Business in partial fulfillment
of the requirements for the Degree of PhD in Business

Thesis Committee:

Adam Tatarynowicz (Supervisor / Chair)
Associate Professor of Strategic Management
Singapore Management University

Geng Xuesong
Associate Professor of Strategic Management
Singapore Management University

Daniel Mack
Assistant Professor of Strategic Management
Singapore Management University

Kannan Srikanth
Associate Professor of Strategy
Ohio State University

Singapore Management University

2020

ABSTRACT

The business environment is more uncertain and ambiguous than ever. Learning and adaptation in such environment is critical for organizational survival and prosperity. This dissertation, with three essays, investigates how organizations can effectively learn and adapt under uncertainty and ambiguity.

The first essay addresses the question about how high and moderate aspiration levels compare in terms of affecting the decision making and reinforcement learning in an uncertain environment. After developing a thought experiment and a computational model, I used lab experiments to test the model's predictions: a high (moderate) aspiration level reduces (increases) feedback ambiguity about the relative attractiveness of different options, thus increases the exploitation (exploration) tendency of the decision maker. The behavioural difference suggests that high aspirations lead to better performance in stable environments, but worse performance after disruptive shocks.

The second essay investigates whether organizations should commit more (or less) to exploration in response to an increased environmental dynamism. Using a computational model, I address the literature contradictions by disentangle exploration intensity and width. I demonstrate that the phenomenon of "chasing a moving target" (Posen & Levinthal, 2012) – the decreasing optimal exploration level under increased environmental dynamism – is caused by the entanglement of exploration intensity and width.

The third essay addresses the question about how ambiguous performance feedback across organizational levels affects resource allocation. Attribution theory suggests organizations and organizational members will attribute success internally while

attributing failures externally, resulting different learning and response patterns following organizational success and failure. Using professional basketball data, I demonstrate the resources (minutes) allocated to players are subject to the players prior performance. Team performance (game win) positively moderates the relationship between allocated resource and a player's performance. The moderating effect is the weakest when the team experience a loss with large point-deficit.

TABLE OF CONTENT

ACKNOWLEDGEMENT

I would like to thank my dissertation committee members, Prof. Tatarynowicz, Prof. Geng, Prof. Mack and Prof. Srikanth for their patient feedback and kind encouragement throughout the journey. I would like to thank the university, the school and the department for providing research funding and creating a learning environment. I would like to thank my fellow doctoral students for their help, support and friendship. I would like to thank audiences from AOM and TOM conferences for providing constructive feedback to my research, especially Prof. Puranam, Prof. Denrell and Prof. Raveendran. Last but not least, I would like to thank my parents and my wife Yanyan for their love, support and understanding.

ESSAY A: Aspiration levels, and exploration: an adaptive learning approach

Introduction

Since March (1991) introduced the exploration–exploitation trade-off to organizational theory and strategy, a central concern in the literature on organizational search has been how companies can strike a balance between *exploiting* existing competencies and *exploring* new opportunities. In the last two decades, a substantial theoretical and empirical literature has studied how organizations balance the conflicting demands of exploration and exploitation – for example, in the context of ambidexterity (for a review, see O'Reilly and Tushman, 2013; for a meta-analysis, see Junni *et al.*, 2013) or problemistic search and organizational adaptation (see reviews by Baumann, Schmidt, and Stieglitz, 2019; Posen *et al.*, 2018). One focus in this literature is on understanding how the aspiration-performance gap affects an organization's choice between these two activities (for reviews, see Bromiley and Rau, 2019; Greve and Gaba, 2017; Shinkle, 2012).

At the organization level of analysis, there is broad agreement that performance below aspirations triggers search away from the status quo, or *exploration*. In contrast, performance at or above the aspiration level triggers narrow search that reinforces the status quo – that is, *exploitation* (Cyert and March, 1963). However, March (1991) framed the trade-off differently: as a process of allocating resources among uncertain alternatives. Viewed in this way, there is no specific status quo for the manager to abandon; hence the trade-off between exploration and exploitation is modeled as a problem of *learning* (under uncertainty) about the

relative attractiveness of options and not as a problem of broad versus narrow search (see Billinger *et al.*, 2019; Denrell, 2003; Denrell and Fang, 2010; Denrell and March, 2001). we know much less about how managers' aspiration levels affect exploration–exploitation choices in this "learning *and* choice" scenario than in the traditional search scenario.

Learning under uncertainty is a ubiquitous problem in organizations. At the corporate level, managers need to decide how to reallocate resources from cash cow businesses among relatively uncertain options. Even if industries are known to be more or less attractive, the expense and uncertainty involved in successfully redeploying resources and capabilities makes this a problem where managers need to learn from feedback to update their priors on suitable opportunities. Business managers and entrepreneurs need to iterate entire business models as decision bundles that contain market segments and product offerings under different degrees of competitive pressure until they find that sweet spot. In all these instances, individual managers (or a small team) make exploration-exploitation decisions that then get reflected at the organizational level. I expect that the aspiration levels managers adopt for the organization or unit performance influence this process.

Reflecting this process, an emerging insight from the literature is that organizational balance depends critically on individual behavior – in particular, on how employees or managers make choices between exploration and exploitation and how those choices are conditioned by the organizational context (Eisenhardt *et al.*, 2010; Mom *et al.*, 2009; Von Hippel, 1988). Studying the micro-foundations underlying the organizational exploration–exploitation activities is of particular

relevance in the behavioral theory of the firm, where behavioral plausibility is a chief consideration when seeking to establish process validity and ground formal models (Cyert and March, 1963; Gavetti *et al.*, 2012; Harrison *et al.*, 2007; Tracy *et al.*, 2017). In this essay, I examine how aspiration levels affect choices between exploration and exploitation at the individual level.

The exploration–exploitation trade-off has, in its own right, been studied extensively at the individual level. Here scholars define the trade-off relative to the decision maker's beliefs: *exploitation* is choosing the option that the decision maker currently believes to be the best alternative available; in contrast, *exploration* amounts to choosing an option that is currently not believed to be the best available (Daw *et al.*, 2006). It is this definition, at the individual level of analysis, that I will use throughout the paper. Because individuals are uncertain about the relative attractiveness of the options available to them, they explore for the purpose of gathering more information (Cohen, McClure, and Yu, 2007). This trade-off 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 (Posen and Levinthal, 2012).

Prior work at the individual level models the trade-off between exploration and exploitation in terms of reinforcement learning, which 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, for additional information; and (2) the *adaptive learning* process of translating feedback from accumulated experience into

representations of the task environment (Cohen, McClure, and Yu, 2007; Daw *et al.*, 2006; Denrell and March, 2001; Li, Mayhew, and Kourtzi, 2009; March, 1996; Sutton and 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, Mayhew, and Kourtzi, 2009).

However, this work on adaptive learning has not yet considered how aspiration levels – a core construct in the behavioral theory of the firm – affect these choice and learning processes. Aspiration levels are fundamental to "satisficing" (Simon, 1955, 1997), and they have a strong effect on search and decision making (Cyert and March, 1963). However, we do not know how aspiration levels and related satisficing processes affect reinforcement learning and, thereby, choice. Hence I am motivated to study how aspiration levels influence the exploration–exploitation trade-off in uncertain environments.

To highlight the importance of studying these two aspects jointly, consider the following example. An executive at a pharmaceutical company must decide how to allocate resources among several alternative new technologies – such as gene therapy, engineering proteins, and monoclonal antibodies. This firm performs experiments by funding projects on each of these technologies. The executive realizes that the success or failure of a particular project to meet objectives (e.g., meeting internal patenting or rate-of-return requirements) is at best a noisy indicator of how attractive a technology is; as a result, the executive must learn about their relative attractiveness, experientially, from feedback. Although such choice problems are endemic at both the individual and organizational levels (one can view many resource allocation

problems as essentially choice problems under conditions of uncertainty and experiential learning), hardly any research has been devoted to exploring exactly how aspiration levels (or reference points) affect feedback-driven learning.

I argue that different aspiration levels will affect learning from feedback, and thus the choice between exploration and exploitation, because aspiration levels affect how feedback is *interpreted* by the decision maker. I shall attend solely to exogenous aspiration levels, or those agreed upon or imposed from the outside, since they are the most relevant to the organizational context (Fiegenbaum and Thomas, 1995; Greve, 2002). Simon (1955) argued that aspirations perform an encoding function for the decision maker – in effect, categorizing the feedback space into successes and failures (cf. prospect theory). This encoding of continuous outcomes into categorical feedback can alter an individuals' perception of the relative attractiveness of the choices, so it can have a notable impact on learning from feedback in ways not readily explained when one considers only the objective payoffs.

To illustrate this idea, consider the following example. An individual may invest in two mutual funds, A and B, and obtain returns of (respectively) 9% and 11% after the first year. If this investor is targeting an overall 10% return as her aspiration level, then fund A will be classified as a failed investment and fund B as a successful investment. The investor will thereafter be more likely to invest in fund B, and in amounts that are disproportionate to the actual difference in payoffs. Suppose the investor continued to invest in both funds. If fund A provided returns between 8% and 10% and if fund B provided returns between 9% and 11%, then the investor would become much more confident that fund B is superior to fund A because she

receives unambiguous feedback from her experience: fund A provides "successes" (relative to her aspiration level of 10%) half the time, but it is rare for fund B to generate a success. In this way, even small differences between actual values – provided they are distributed around an aspiration level – can reduce ambiguity in the decision maker's mind about which option is superior while increasing her confidence about its superiority. Thus I build on Posen and Levinthal's (2012) insight that a decision maker's confidence about the relative attractiveness of different choices (i.e., her *strength of opinion*) may endogenously influence her exploration behavior even when her strategy remains constant. Our paper takes this intuition one step further and suggests that strength of opinion is itself a function of the decision maker's aspiration level, which in turn affects what I call *feedback ambiguity*.

Feedback ambiguity captures the extent to which the possible payoffs from different options overlap. The more the possible payoffs (i.e., their distributions) overlap, the greater the feedback ambiguity experienced by decision makers when choosing either option. At the two extremes, feedback ambiguity will be zero if the possible payoffs from the different options have no overlap at all whereas feedback ambiguity will be complete if the possible payoffs from all choices are identical. The theory I develop shows that, as a decision maker learns from experience, aspiration levels can help (or hinder) him by reducing (or increasing) feedback ambiguity. Asymmetry in feedback ambiguity, which can arise among decision makers with different aspiration levels, affects their strength of opinion, that is the perceived relative attractiveness of options, and hence the decisions made about exploration and exploitation.

To investigate these interrelationships, I conduct a behavioral experiment. I employ a simple agent-based model to develop the theory and use the model's underlying multi-armed bandit task to test our hypotheses in a behavioral lab experiment. I find that high aspiration levels in stable environments can reduce feedback ambiguity, which results in greater strength of opinion and higher performance; both of these are driven not only by more extensive exploitation but also by narrower exploration. When a disruptive environmental shock alters the relative attractiveness of different payoffs, I find that the reduction in feedback ambiguity – which is beneficial in stable environments – can become a liability in unstable environments. Thus I document that the propensity of decision makers to exploit, both early and often, has a deleterious effect: it delays their adaptation to the exogeneous shock and so their performance recovers more slowly than does that of decision makers who have lower strength of opinion and who therefore engage in relatively more exploration. In addition, I test the mechanism's boundary conditions by examining different types of payoff structures and environmental shocks.

This study contributes to the literatures on adaptive learning and aspiration levels as well as (more broadly) to research on the exploration–exploitation trade-off. Previous work on that trade-off at the individual level has viewed the problem either through the lens of reinforcement learning or as a purely organization-level choice problem when there is a gap between performance and aspiration levels. I bring these two strands together by examining how individuals of varying aspiration levels learn about choices under uncertainty, thereby integrating two pillars of behavioral theories in the Carnegie tradition (Gavetti *et al.*, 2012; see also Denrell, 2008). In this way, I

contribute to the micro foundations of how managers balance the need for exploration and exploitation in organizations. Thus I enrich the extensive literature on the behavioral theory of the firm, especially with regard to the organization's response to an aspiration–performance gap (Gary *et al.*, 2017; Greve, 2018; Greve and Gaba, 2017; Posen *et al.*, 2018; Shinkle, 2012).

Literature Review

In behavioral theories of the firm, the extent of exploration and exploitation depends not only on the environment but also on the organization's performance relative to its aspiration level (Cyert and March, 1963). How performance relative to aspirations affects organizational decision making is the focus of an active stream of research that draws on several theoretical perspectives, including the behavioral theory of the firm and strategic reference point theory (for reviews, see Bromiley, 2010; Bromiley and Rau, 2019; Greve, 2003; Greve and Gaba, 2017; Posen *et al.*, 2018; Shinkle, 2012).

Cyert and March (1963) argued that organizations whose performance falls short of aspirations will engage in problemistic search, or explore; if performance exceeds aspirations then they continue to execute and refine current routines and policies, or exploit. Subsequent research has added considerably to our understanding of this basic premise; thus studies have addressed the effects of different levels of the performance–aspiration gap in positive or negative domains (cf. Eggers and Kaul, 2018; Greve, 2003), whether performance provides ambiguous or instead unambiguous feedback with respect to multiple dimensions of aspirations (Greve and Gaba, 2019), and the hierarchical role of the decision maker (Joseph, Klingebiel, and

Wilson, 2016). However, all these studies build on Cyert and March's (1963) premise that lower performance relative to aspirations necessarily increases search (i.e., exploration). A corollary is that the higher the aspiration level, the more likely the firm explores – that is, because it encounters failure more frequently (Greve, 2017, 2018).[1]

In their comprehensive review of the problematic search literature, Posen *et al.* (2018) identified 53 studies that test this basic proposition. The authors were surprised to find that only about half of the studies found empirical support for it. The others found either non-significant or mixed results, and some even found the opposite relationship. So as the review indicates, much scholarly effort has been devoted to identifying moderators that can characterize the contingencies under which this basic hypothesis is *not* supported; examples include "slack" search and survival concerns. In a provocative study, Denrell (2008) questioned whether increased risk taking is actually the underlying mechanism, suggesting that the mechanism may instead be feedback-driven adaptation to an uncertain environment.

As I pointed out in the Introduction, these alternative perspectives may arise under different circumstances. Conceptualizing exploration as search away from the status quo may be most apt when managers change existing routines based on performance feedback. Modelling the trade-off between exploration and exploitation as a consequence of reinforcement learning may be more applicable in a context

[1] In this theory, aspirations themselves change with feedback (March, 1988) – in contrast to strategic reference point theory, which suggests that aspirations are often externally imposed on organizations (Fiegenbaum and Thomas, 1995; Greve, 2002). Greve and Gaba (2017) and Shinkle (2012) remarked that few studies at the firm level actually measure aspirations or observe whether organizational aspirations do in fact change depending on achieved performance.

where managers are learning about unknown alternatives in an uncertain environment, where an unambiguous status quo may not exist. History plays a different role in these approaches: in behavioral theories, it affects aspirations (March, 1988); in reinforcement learning, it changes beliefs about the relative attractiveness of choices (Daw *et al.*, 2006). Our aim is to shed more light on how these two aspects – aspiration levels and adaptive learning – interact in guiding the decision maker's choice between exploration and exploitation.

Under the micro-foundational approach, it is ultimately managers' decisions that are aggregated to generate organizational choices regarding exploration and exploitation. Scholars have called for a renewed look at the micro foundations of individual and group behavior as a building block in efforts to reconcile mixed empirical findings and to build a more robust theory of organizations. By adopting a micro-foundational view of the aspiration–performance link, I aim to uncover what structural and/or cognitive mechanisms drive the choice between exploration and exploitation.

At the micro level, the goal-setting literature presents contrasting predictions about how high versus moderate levels of goals (which are closely related to aspirations) affect individual behavior (Locke and Latham, 2009; Ordóñez *et al.*, 2009a, 2009b). Locke and Latham undertook a comprehensive review of the goal-setting literature and reported that, to the extent that individuals are committed to goals and are able to attain them, "there is a positive linear relationship between goal difficulty and task performance" (2006, p. 265) because goals that are challenging increase effort and persistence. It is worth noting that these authors articulate the

relationship in terms of performance – and not with reference to whether higher aspirations influence exploration versus exploitation – although they do suggest that challenging goals motivate the search for high-performance strategies (similar to exploration) as well as the subsequent exploitation of those strategies (cf. Earley, Connolly, and Ekegren, 1989).

In contrast, Ordóñez *et al.* (2009a, 2009b) argued that individuals who face challenging goals – rather than moderate ones – make more risky decisions (see also Heath, Larrick, and Wu, 1999; Larrick, Heath, and Wu, 2009). These studies suggest also that individuals tend to choose options with a high variance in payoffs; therefore increased aspiration levels do not improve overall performance. Yet most of the empirical evidence for these claims comes from laboratory studies, where (a) subjects choose once between a relatively safe bet and a risky bet and (b) the safe bet's payoff is usually below the high aspiration level. The few multi-period studies that examine how goals affect self-regulatory processes (e.g., Cervone, Jiwani, and Wood, 1991; Gary *et al.*, 2017; Kanfer and Ackerman, 1989; Wood, Bandura, and Bailey, 1990) fail to capture the learning–performance link. Scholars have similarly argued that, in reality, individuals and organizations are seldom presented either with plausible choices or their outcome distributions; the latter must be learned from experience (Denrell and Le Mens, 2011; Gigerenzer, 2008; Gigerenzer and Gaissmaier, 2011; Le Mens and Denrell, 2011; see also Bromiley and Rau, 2019), which is subject to such pathologies as the "hot stove" effect (Denrell and March, 2001) and time-period effects (Song, Bnaya and Ma, 2019).

A considerable empirical literature in computer science, neuroscience, and psychology considers how individuals make explore–exploit decisions under uncertainty. However, that research does not consider how aspirations (high versus low) affect those decisions. This is a major gap in the literature when one considers that so much individual behavior is goal directed, especially in organizations, and involves decision making under uncertainty.

In order to address this gap and examine more carefully how aspiration levels affect the choice between exploration and exploitation in uncertain environments, I build a simple adaptive learning model to develop hypotheses. I then test our hypothesized predictions in a series of behavioral laboratory studies.

Adaptive Learning Model

In order to inform our theory building, I develop a model for how adaptive learning affects the exploration–exploitation choice under different aspiration levels. To explain the basic intuition underlying the model, I first describe the following thought experiment.

A Simple Thought Experiment

Consider the two-armed bandit task illustrated in Panel (a) of Figure 1. There are two choices to select from, "arm 1" and "arm 2". These arms could represent investment options or research and development (R&D) projects. Whenever a decision maker selects or "pulls" an arm, she receives a payoff (e.g., return on investment). In our example, arm 1 returns a payoff uniformly distributed between 0 and 1 (i.e., any payoff received by the decision maker for choosing arm 1 will be a value between 0

and 1 with equal probability); arm 2 returns a payoff uniformly distributed between 0.5 and 1.5. This example features a sizeable overlap in the two arms' respective payoffs, which implies that identifying the better arm (arm 2) requires learning from experience over a number of trials.[2]

[[INSERT **Figure 1** ABOUT HERE]]

Suppose there are two decision makers (DMs): one with a high aspiration level of 1, which is equal to the mean of the better arm's payoff (this is DM1 in Figure 1); and one with a moderate aspiration level of 0.75, which is equal to the average payoff of both arms (DM2).[3] If decision makers knew the true payoff distributions for both of these arms, then choosing between them would be a trivial task: each decision maker would always choose (i.e., exploit) arm 2, the better arm. There would be no need for reinforcement learning because the payoff distributions would be known and the decision makers' aspiration levels would not affect their choices – both would choose the better arm every time.

However, decision makers are seldom aware of the arms' payoff distributions; and notwithstanding any awareness they do have, decision makers' beliefs about each arm's possible payoffs will be updated through learning from new experiences. For example, DM2 may first choose arm 1 (the worse one), receiving a payoff of 0.9, and then choose arm 2 (the better one), receiving a payoff of 0.7. If DM2's judgment is

[2] Unlike the games described in many goal-setting studies, in this game the two arms have equal risk (i.e., their variance in payoffs is identical). In this way I isolate the exploration problem from risk taking (or preferences for risk).

[3] I could envision a third decision maker, DM3, with a low aspiration level equal to the average payoff of the worse arm; DM3 would face the mirror image of DM1's conditions. Hence I limit our thought experiment to the minimum number of decision makers and aspiration levels necessary to convey the intuition.

based only on these two trials, then he would judge arm 2 to be worse than arm 1. Yet over successive trials – assuming DM2 chooses both options repeatedly – he will develop a more accurate sense of the two arms' relative payoff distribution. Thus, for instance, sampling both arms repeatedly may yield payoffs of less than 0.5 for arm 1 and of more than 1 for arm 2 but *not* vice versa. I follow the literature prior literature and say that the decision maker *exploits* if he chooses the option that he believes gives the highest payoff and *explores* if he selects an option that he believes does not give the highest payoff. As he continues to select different options, the decision maker updates his beliefs about which option will return the better outcomes. This simple example underscores that learning figures prominently in a decision maker's search for better options.

How does the decision maker's aspiration level affect this learning process? According to Simon (1955, p. 105, Fig. 1), aspirations perform an encoding function that reduces a complex environment into a smaller number of states. He argues that this encoding function is an essential purpose of aspirations, partitioning the payoff space into successes and failures (see also Heath *et al.*, 1999). If that process is applied to our thought experiment in Figure 1b, then DM1 – with a high aspiration level of 1 – will encode all received payoffs above 1 as "successes" and all received payoffs below 1 as "failures". Yet DM2, with a moderate aspiration level of 0.75, will encode all received payoffs above 0.75 as "successes" and below 0.75 as "failures." This simple difference in decision makers' interpretation of the received payoffs fundamentally alters their learning and choice processes.

Decision makers with different aspiration levels likewise encounter different degrees of feedback ambiguity. Recall that feedback ambiguity captures the extent to which the possible payoffs from different options overlap. In our thought experiment, the feedback ambiguity experienced by DM1 is much lower than the ambiguity that DM2 experiences. More specifically, DM1 will encode *almost all* payoffs received from arm 1 as "failures" because the worse arm's highest possible payoff is 1 – the same as her aspiration level. Given the small overlap between arm 1's payoff distribution and her aspiration level, the feedback ambiguity is quite low (yet it is nonzero because she has a 50% chance of receiving payoffs of less than 1 from arm 2). Thus DM1, who has a high aspiration level, will probably categorize most feedback from the low-payoff option as "failures." Hence this reduction in feedback ambiguity increases the high-payoff option's relative attractiveness and, thereby, the decision maker's confidence about the two arms' relative attractiveness, or her strength of opinion (Posen and Levinthal, 2012). In contrast, DM2 (whose aspiration level is a moderate 0.75) has a 75% chance of encoding any given payoff from arm 2 as a "success" (vs. a 50% chance by DM1). Yet DM2 interprets a greater proportion of arm 1's payoffs as successful as well, with a 25% probability of encoding the payoffs received from the worse arm as "successes". The result is a higher level of feedback ambiguity for DM2 because the greater overlap in "success" payoffs across the two options reduces her strength of opinion about the relative attractiveness of the two arms.

When aspiration levels are high (DM1 in Figure 1), the worse arm consistently returns payoff feedback that is categorized as a failure. Even with a

limited number of trials, these repeated failures unambiguously signal to the decision maker that this arm should be avoided. Thus, his high aspiration level – and his resulting categorization of feedback into successes and failures – has the effect of reducing feedback ambiguity. I therefore posit that aspiration levels may either increase or decrease the feedback ambiguity experienced by a decision maker, thus reducing or amplifying (respectively) his strength of opinion. In this sense, feedback ambiguity is a structural mechanism that affects a cognitive factor: the decision maker's strength of opinion.

In contrast to the case of high aspiration levels, a decision maker with only moderate aspirations (here, DM2) experiences successes and failures from both arms. When the number of trials is more limited, these mixed successes and failures send a more ambiguous signal about both arms; hence the decision maker will be less certain (than if she had high aspirations) about which arm is worse, and which arm is better, and so faces relatively greater *feedback ambiguity*. In our thought experiment, then, an agent with high aspirations is more confident about identifying the inferior arm (and therefore the superior arm) – has greater "strength of opinion" – than an agent with moderate aspirations.[4]

Posen and Levinthal (2012) theorized that a reduction in strength of opinion endogenously increases exploration behavior even if the decision strategy is constant. I augment their account by suggesting that strength of opinion may itself be a function of the decision maker's aspiration level that affects the feedback ambiguity that he experiences. Thus I suggest that feedback ambiguity is the mechanism via

[4] I provide a more detailed explanation of this process in the Online Supplement.

which aspiration levels influence adaptive learning and, ultimately, the choice between exploration and exploitation. These considerations lead to the following proposition.

> *Aspiration levels can generate different degrees of feedback ambiguity, which in turn affect the decision makers' choice of exploration versus exploitation. To the extent that higher aspiration levels are more likely to reduce feedback ambiguity about the inferiority of the inferior options, , they are also more likely to encourage more exploitation (resp., less exploration) of superior (resp., inferior) options.*

For the sake of simplicity, our thought experiment uses a two-armed bandit task in a stable environment. However, choices regarding exploration and exploitation often arise in more complex decision tasks that may involve changing environments and multiple options (arms).

First, environmental changes reduce the usefulness of prior learning through experience and require that decision makers re-sample the arms to learn how their payoffs are distributed after any change. For instance, the previously worse arm may now provide better payoffs on average. Precisely because the feedback ambiguity mechanism and strength of opinion both rely on differences in the beliefs formed through adaptive learning, I can demonstrate this mechanism more convincingly in an environment where such opinion strength can be a liability – as in, for example, an unstable environment (Posen and Levinthal, 2012).

Second, a task involving multiple choices, constitutes a more complex decision-making environment. In order to test whether the feedback ambiguity generated (or reduced) by aspiration levels affects the decision-maker's strength of opinion, the payoff environment needs to be sufficiently complex. In particular, aspiration levels can change feedback ambiguity differentially across available choices. A *three*-armed bandit task is therefore the minimal set-up in which I can explore the mechanism of feedback ambiguity as well as how it affects the level of exploitation under changing environments.

However, a pure thought experiment with three options in a changing environment quickly becomes intractable. I therefore develop our hypotheses with the aid of a simple agent-based simulation designed to explore the effect of aspirational levels and feedback ambiguity on the level of exploitation and performance. Then I use the same task set-up to test those hypotheses in a behavioral laboratory experiment.

Agent-Based Simulation

In order to understand how adaptive learning and aspiration levels jointly affect the exploration–exploitation trade-off with environmental shocks, I set up a three-armed bandit task whose arms are differentially attractive yet equally risky (i.e., their payoffs have different means but the same standard deviation) and for which the payoff distribution is unknown to the agents. More specifically, I create a three-armed bandit task for which the arms' payoffs are uniformly distributed with means of [25, 20, 15] for arms 1, 2, 3 (respectively), each with constant interval of ±10. This setup is identical to the investment game in our laboratory experiment. The agents can

obtain information about the arms only by sampling and adaptive learning. Thus agents learn by updating their beliefs about the different options' payoffs as they sample them across the rounds t of the simulation.

In line with our thought experiment, I set the high aspiration level equal to the best arm's mean payoff (25) and the moderate aspiration level equal to the average payoff (20) of all three arms. Then the middle arm's payoff is distributed around the moderate aspiration whereas the best arm's payoff is distributed around the high aspiration (as described in the thought experiment). This set-up of the arms and aspiration levels results in asymmetric feedback ambiguity: the worst arm provides feedback that is unambiguously bad for an agent with high aspirations yet is ambiguous for agents of moderate aspiration levels. Thus our set-up replicates the differential feedback ambiguity across arms discussed previously. At the same time, our set-up captures the common conceptualizations of high and moderate aspirations: high (also labelled "challenging" or "stretch") aspiration is difficult to reach, yet still feasible, corresponding to the industry leader's performance; moderate aspiration is relatively easy to reach, reflecting industry average performance (Gary *et al.*, 2017; Locke and Latham, 1990).

If I set the high aspiration level even higher, agents with high aspiration levels would converge on the best arm even faster; if I set the moderate aspiration to the average of the lowest arm, then feedback ambiguity (between the low and middle arm) would be reduced for those low–aspiration level agents by an amount equivalent to the reduction in feedback ambiguity for the high–aspiration level agents. Setting the aspiration at high and moderate levels allows us to maximize the difference in

feedback ambiguity between the two aspiration level conditions studied *without* adding needless complexity to the model.[5]

Adaptive learning models of a bandit tasks are commonly modelled with both a belief updating rule and a choice rule (Sutton and Barto, 1998, Puranam et al, 2015, Posen and Levinthal, 2012). In addition to these standard components, I incorporate a feedback interpreting process as commonly implemented by the reference point literature (Tversky and Kahneman, 1992; Larrick & Wu, 1999) to model the effect of aspiration levels.[6] Here, I describe each of these in turn.

Feedback interpreting process: To incorporate the effect of aspiration levels in the adaptive learning model, I build on the reference point literature (Tversky & Kahneman, 1992; al-Nowaihi, Bradley & Dhami, 2007; Heath, Larrick & Wu, 1999) and interpret aspirations as reference points to interpret feedback. I therefore model feedback interpretation as

$$U_{i,t} = \begin{cases} \left(R_{i,t} - A\right)^{\alpha}, if\ A \leq R_{i,t} \\ -\left(A - R_{i,t}\right)^{\alpha}, if\ A > R_{i,t} \end{cases}$$

In this formulation, $U_{i,t}$ represents the utility that an agent attaches to the payoff received from arm i in round t, i.e. the utility based on interpreted feedback. $R_{i,t}$ denotes the received payoff (or reward) from arm i in round t, and A captures the aspiration level (or reference point). Therefore, the agent's feedback interpretation in

[5] If the high aspiration is set so high that all arms result in failure, or if the moderate aspiration level is set so low that all arms return successes, then the feedback ambiguity mechanism is no longer in play.
[6] Our results are robust to a wide range of τ-values (or temperatures): from 0.002 to about 1. This range is comparable to that used in most modeling work based on a Bernoulli bandit (Posen and Levinthal, 2012).

round t is positive if the aspiration level A is greater or equal than the received payoff $R_{i,t}$ and negative if the the aspiration level is below the received payoff. This feedback interpretation is affected by $\alpha \in [0,1]$, the diminishing sensitivity parameter.[7] $\alpha = 0$ is a special case when feedback was strictly dichotomized into "successes" or "failures"

Belief updating rule: I follow the standard textbook (Sutton and Barto, 2018, p. 32) and use an incremental value updating rule (i.e., temporal difference rule) to update the agent's belief (also see Denrell 2007; Stieglitz, Knudsen and Becker, 2016).

$$q_{i,t+1} = q_{i,t} + \theta\left(U_{i,t} - q_{i,t}\right)$$

here $q_{i,t}$ is the estimated representation of the agent's belief about arm i arm at time t (and $q_{i,t+1}$ denotes the updated representation at time $(t+1)$). The $U_{i,t}$ term inserts the feedback interpreting process into the belief updating rule and $\theta \in (0,1)$ is the stepwise parameter governing the speed of learning – the greater θ is, the stronger the recent feedback is weighted in the representation.

Choice rule: I continue to follow the classic setup (Sutton and Barto, 1998) for the learning rule. Here, the probability of a given agent choosing a particular arm i is based on his beliefs $q_{i,t}$, following a Softmax rule:

[7] For simplicity, I assume preference-homogeneity: the diminishing sensitivity parameter is identical for feedback above and below aspiration (al-Nowaihi, Bradley & Dhami, 2007).

$$\pi_i = \frac{\exp\left(\frac{q_{i,t}}{\tau}\right)}{\sum_{j=1}^{M} \exp\left(\frac{q_{j,t}}{\tau}\right)}$$

in which π_i is the probability of choosing the ith arm among the total number of available arms M, and $\tau \geq 0$ is the parameter that governs exploration strategy. The higher the τ is, the lower the agent's sensitivity to the current beliefs and the greater his likelihood to choose the arm with inferior beliefs. As τ approaches zero, the agent only chooses the arm with the highest belief, i.e. only exploits.

The simulation runs for 50 periods. For each aspiration level, I simulate 500,000 agents, with parametrization $\alpha = 0.5, \theta = 0.2$ and $\tau = 0.2$. I "seed" the simulation by assuming that agents will pull each arm once before the simulation starts, giving them a value for the first term in the belief updating rule in period 1[8]. This approach ensures that the simulation results are not biased by our assignment of an arbitrary value for $q_{i,0}$. Thereafter, agents choose an option in every round based on the choice rule described. As agents with high versus moderate aspiration levels learn, the simulation reveals several interesting interactions among learning, aspiration levels, and environmental shock and also shows how these interactions affect exploration–exploitation behavior. These outcomes form the basis of our hypotheses.

In line with our proposition advanced in the previous section, the model shows agents with high aspiration levels are less (more) likely to explore (exploit), that is, not pulling the arm with the highest belief $q_{i,t}$ (shown in Figure 2a). The

[8] Robustness of the parameterization can be found in the Appendix, where I systematically vary these parameters. I also vary the seeding of initial belief.

model also shows that agents with high aspiration levels settle on the best arm more quickly than do moderate–aspiration level agents. Figure 2b plots the likelihood of choosing each arm as agents learn over time. The results highlight that feedback ambiguity remains even for high–aspiration level agents: they do explore (i.e., choose the option they believe *not* to be the best arm) and they neither immediately nor entirely settle on the best option. Yet when compared with agents who have a moderate aspiration level, high–aspiration level agents experience less feedback ambiguity and are therefore more likely to choose the arm they believe is the best (i.e., to exploit), and are also more likely to be correct (i.e., actually choose the truly best arm).

[[INSERT **Figure 2** ABOUT HERE]]

To highlight the mechanism by which feedback ambiguity affects learning and choices, I examine the impact of feedback ambiguity on agents' formation of beliefs and hence on their strength of opinion. Figure 3 illustrates the standard deviation of the *difference* in the beliefs of agents with high versus moderate levels of aspiration; that is, the feedback from payoffs as interpreted through the lens of their respective aspiration levels. The three panels in Figure 3 illustrate this difference for every combination of arms (within-panel comparisons). In this figure, the *lower* the standard deviation of differences in belief, the *higher* the ambiguity about which arm is better.

[[INSERT **Figure 3** ABOUT HERE]]

Overall, a high aspiration level leads to a lower standard deviation of the beliefs about arms' payoffs and hence to less ambiguity about which arm is better.

Thus agents with high aspirations pick the best arm across each pair with greater certainty and, as a result, are less likely to explore other arms. In particular, Figures 3(a) and 3(b) establish that high-aspiration agents are more certain (than their moderate-aspiration counterparts) that the best arm is superior to both the middle arm and the worse arm. This dynamic explains why, in Figure 2, high–aspiration level agents are more likely to exploit the best arm; it also emphasizes that agents with a high aspiration level sample the worst arm less often, which further increases their certainty about the best arm's payoffs. Thus I are led to our first hypothesis, as follow.

> **Hypothesis 1** (H1)**.** *High–aspiration level agents choose the option that they believe is* ***best*** *(exploit)* ***more*** *frequently than do agents of moderate aspiration level.*

In a stable task environment, where the payoff distributions are not changing, exploitation of the best arm leads to better performance regardless of aspiration level. High aspiration levels facilitate finding the best arm – as well as a strategic focus on exploitation – by reducing ambiguity in the feedback from alternatives, which leads to finding the best-performing arm sooner and also to higher cumulative performance (as shown in Figure 2b). Thus H1 implies that high–aspiration level agents explore less often than do agents with moderate aspiration levels. High–aspiration level agents exploit more and explore less; hence they will probably perform better at tasks, such as this one, for which the arms' payoffs (means and distributions) remain constant. Formally, I have the following hypotheses.

Hypothesis 2a (H2a)**.** *High–aspiration level agents choose the truly best option **more** frequently in a **stable** environment than do agents with moderate aspiration levels.*

Hypothesis 2b (H2b)**.** *High–aspiration level agents exhibit **better** cumulative performance in a **stable** environment than do agents of moderate aspiration level.*

When the task environment is changing, however, a strategy that focuses on exploitation hinders adaptation to new payoff structures. If agents use adaptive learning to form their beliefs about the relative attractiveness of choices, then those with stronger beliefs are more likely to persist with them – and for a longer period – under some kinds of environmental turbulence (Posen and Levinthal, 2012). Suppose, for instance, that the payoffs of previously unattractive choices improve while those of previously attractive choices decline. Because agents with stronger beliefs are less likely to explore unattractive choices, they are also less likely to observe this particular change as it occurs. Since a higher aspiration level can lead to stronger belief formation, it follows that agents with higher aspiration levels adapt less quickly to the new environment following a disruptive shock.

I undertake another simulation to assess this intuition. Expanding the previous set-up, I introduce a disruptive shock – after period 30 – that changes the mean payoff of arms 1, 2, and 3 from [25, 20, 15] to [20, 15, 25]. That is, the worst arm becomes the best arm and the other arms shift downward in value.[9]

[9] I shift these payoffs – that is, rather than merely increasing the worst arm's payoff above the other two arms – so as to maintain the relevance of high (and moderate) aspiration levels. If the worst arm's payoff increases beyond the payoffs of the other arms, then "high" aspiration loses its meaning.

[[INSERT **Figure 4** ABOUT HERE]]

I find that, although the strong beliefs developed by a high-aspiration agent serve her well in a stable environment, those beliefs prevent her from adapting to a *changed* payoff structure in a timely fashion. Figure 4 plots the difference between high and low aspiration levels in the choice between arms 1 and 3 following the shock. This figure reveals that, before the shock, agents with high aspirations exploit the best arm and avoid exploring the worst arm. However, this strong preference for the best arm *before* a shock delays the agent's identification of the best arm *after* a shock. There are clear performance implications of high–aspiration level agents' slower adaptation to the changed payoff, and our simulation motivates the following hypotheses.

> **Hypothesis 3a** (H3a)**.** *High–aspiration level agents choose the newly changed best option **less** frequently **after** a disruptive environmental shock than do agents with moderate aspirations.*
>
> **Hypothesis 3b** (H3b)**.** *High–aspiration level agents exhibit **worse** performance directly **after** a disruptive environmental shock than do agents of moderate aspiration level.*

In sum, I have used this agent-based simulation of a multi-armed bandit task in a changing environment to refine the insights derived from our previous thought experiment. Perhaps more importantly, the model itself makes explicit our assumptions about how aspiration levels affect learning from feedback. Doing so enables direct translation into the lab setting, to which I turn next.

Methodology

To test the hypotheses, I design a behavioral lab experiment in which I manipulate the aspiration level and then observe participants' exploration and exploitation choices as well as their performance. The hypotheses are tested in the three main studies, and boundary conditions are explored in an additional five studies. Each experimental participant plays a single-player investment game that uses the same multi-armed bandit task as the adaptive learning model already described. This task has been used extensively to study exploration–exploitation behavior in studies both theoretical (Denrell and March, 2001; Lee and Puranam, 2016; Posen and Levinthal, 2012; Sutton and Barto, 1998; for a review, see Puranam *et al.*, 2015) and experimental (Cohen *et al.*, 2007; Daw *et al.*, 2006; Gans, Knox, and Croson, 2007; Knox *et al.*, 2012). As Simon (1947) stressed, exploration–exploration decisions are inherently behavioral because prior probabilities are unknown and must be learned in a trial-and-error process. And because choice and learning processes are intertwined in these problems (March, 1991; Sutton and Barto, 1998), experimental studies can be used to make valuable inferences by controlling the information available (Edmonds, 2001; Schunk, 2009; Sterman, 1989).

As described in the discussion of our thought experiment, a multi-armed bandit task gives the decision-maker a choice between multiple options, or "arms". Each option has an uncertain payoff, which implies that: (1) the decision maker does not know ex ante which of the different options is better or worse; and (2) the feedback received from choosing an option is noisy. In our set-up, all three options are uniformly distributed and have the same variance but different means. Panel (a) in

Figure 5 illustrates the three arms' respective payoff distributions for all studies. All studies start with the same three arms: the payoffs are uniformly distributed with means [25, 20, 15] for arms 1, 2, and 3, respectively, and with a constant interval of ±10. These payoffs were chosen so that there is not only sufficient overlap for feedback ambiguity to exist but also sufficient unique feedback for decision makers to discern payoff differences over time.

[[INSERT **Figure 5** ABOUT HERE]]

In order to enable learning from experience, decision makers play the investment game over 50 (or 80) rounds. In each round, the decision maker chooses whatever option he wants and receives a payoff. Over time, the decision maker learns which, of the options he explores, yields higher or lower payoffs (on average). The decision maker's payoffs accumulate over the rounds of the game, and his goal is to maximize the end-of-game payoff. In each of the studies, participants are randomly allocated to a high- or moderate-aspiration condition. In our baseline (Study 1) the environment remains constant, whereas participants experience an environmental shock in Studies 2 and 3; this shock changes the means of the different investment options yet leaves the overall payoff landscape otherwise unchanged (as shown in Figure 5). I also perform robustness checks with an added control condition (viz., *no* manipulation of the aspiration level) and explore the boundary conditions of our findings with different incentive structures and alternative shock characteristics (Studies 4, 5a, 5b, 6a, and 6b).

Experimental Procedure

The lab experiment was set up as a between-subject design across three main studies. In each study, participants were randomly assigned to one of two conditions: a high or moderate level of aspiration; I also re-ran all three studies with an added control ("no–aspiration level" condition). I ensured that each participant participated in only one of the studies.[10]

The three main studies exposed participants to different task environments. In two of them, the arms were subject to a shock (after round 30) that re-ordered the arms' relative attractiveness. It is the adjustments that participants make to these shocks that help us understand their exploration behavior. In particular: the more a participant explores, the more likely she is to shift choices in response to an abrupt shock. Study 1 reflects the baseline (no-shock) environment, while Studies 2 and 3 incorporate an identical disruptive shock (but with different post-shock periods: Study 2 has 20 post-shock periods to keep the total number of periods at 50; Study 3 has 50 post-shock periods for a total of 80 rounds). When studying the boundary conditions of our findings, I explore the effects of a positive shock in Study 4. Figure 5(a) summarizes the different study conditions used across all studies. The task set-up ensured that time pressure was not a factor: I did not limit participants' available time, and the time needed to play the game averaged just under 1.5 minutes (excluding the introduction and briefing of the game; 2 minutes for games with extended periods).

[10] The verbatim instructions and procedures for this lab experiment can be found in the Online Supplement. I include screenshots for the investment game and the balloon analogue risk task. Additional information on the underlying model can be found here: .

Task Description. In line with our theoretical intent, I framed the task as one of investment choice under uncertainty. Each participant adopts the role of an R&D manager for a digital business firm and, in each round, decides on a product platform in which to invest. Participants are informed that the payoffs are uncertain and subject to market turbulence. The task description also points out that random environmental shocks could alter the relative attractiveness of the arms. These instructions are given regardless of whether the Study contained a shock. The Online Supplement includes screenshots of the experimental set-up and highlights the (few) instructional lines that differ across conditions and studies.

In each period, participants select one of the three available choices (arms) for investment. Upon making an investment, a participant immediately receives a payoff that is displayed as points earned on the given arm. The feedback is clearly tied to the choice just made (see Online Supplement), and it also updates the cumulative payoffs (at the top of the screen) as well as the current round and progress. Participants can see – at any time – their cumulative payoff, the total number of trials played, the number of times each choice was sampled, and the average payoff received from each choice.

Treatment Manipulation. Participants were randomly assigned to the high or moderate aspiration level condition. In all studies, participants were informed that they should attempt to achieve a particular performance goal. I implemented the manipulation for high [resp. moderate] aspiration levels by instructing the participants as follows: "The previous manager had achieved total earnings of 1,250 [1,000]

points over their tenure of 50 rounds, that is, 25 [20] points per round. You should aim to earn at least this amount."[11]

I set the high aspiration level equivalent to the value of the best arm's mean payoff, so it serves as the benchmark of the best possible alternative in this game. I selected this value (in the upper 10th percentile) in line with goal-setting theory on establishing goals that are challenging yet achievable (Locke and Latham, 2006). The moderate aspiration level is equivalent to the mean payoff of all three arms and is relatively easy to achieve, since pulling the arms at random would ensure a 50% chance of receiving a payoff higher than the moderate aspiration level.

Feedback ambiguity captures the extent of possible payoff overlap across different options. The degree to which the decision maker experiences feedback ambiguity, in turn, depends on the decision maker's aspiration level. Panels (b) and (c) of Figure 5 plot the feedback ambiguity experienced by participants with (respectively) high and moderate aspiration levels. These panels illustrate how feedback ambiguity experienced differs in our setup, due to payoff distributions across choices and aspiration levels. I anticipate that the feedback ambiguity experienced by decision makers accounts for the strength of their opinions about the relative attractiveness of different options, which in turn affects their exploration–exploitation behavior (and hence performance).

Participant Recruitment. For Studies 1 and 3 (as well as the boundary condition studies 5a, 5b, 6a, and 6b), participants were recruited through Amazon's

[11] I also ran all three studies again while using an added control condition without a specified aspiration level but with a "do your best" instruction.

Mechanical Turk (MTurk). For Study 2, I recruited participants from a public university in the United States and the experiment was conducted in a lab setting. In Study 4, students from an undergraduate course at a Singapore university participated in the experiment as part of a class exercise. Because adaptive learning and search behavior are fundamental human behaviors, I decided to run our laboratory studies in different countries so as to increase (howsoever slightly) the external validity of our findings.[12] Table 1 summarizes our data collection as well as participants' demographics.[13] I used our first study to inform our power analysis and estimated that a sample size of 72 (resp. 98) would give us statistical power of 0.8 (resp. 0.9). The power analysis is reported in the Online Supplement.

[[INSERT **Table 1** ABOUT HERE]]

Across the different studies, MTurk participants were similar in terms of the basic demographic dimensions that I were able to collect. On average, about 32% identified as female, and reported an age of 35 on average (with a wide range between 19 and 70 years old). The majority of MTurk participants reported English as their first language and around 66% reported at least a 2-year college degree. I found a difference in country of residence for studies run before versus during the pandemic (resp. 90% versus 71% from North America). In contrast to the MTurk studies, the two university samples were younger (21 on average), had a higher percentage of

[12] The university-based studies were conducted without this control condition and were therefore replicated on MTurk for the high-, moderate-, *and* no-aspiration condition. The results across the university and MTurk studies are consistent, and our findings from the latter are reported in the Results section (under robustness checks). Under normal circumstances I would have re-run these studies with exactly the same populations of participants. However, the Covid-19 pandemic has precluded re-running any studies except via MTurk.

female participants (46%), and fewer students with English as their first language (in the US sample, not available for Singapore). Overall, this set of study populations provides us with some degree of generalizability. The fact that search tendencies among these diverse participants were consistent, in very different exogeneous conditions of uncertainty (before and during the height of the pandemic) appears to indicate that our studies indeed capture a fundamental human behaviour.

Incentive Structures. The incentives designed for this study clearly matter because they make choices consequential (Smith, 1976). In Studies 2 and 4, subjects were rewarded with course participation credits. I incentivized the participants to perform well by offering a $10 gift certificate to each of the top five performers. For Studies 1 and 3, I replaced course credit with a fixed payment for participation but otherwise maintained the same performance-based incentives. In setting up this first incentive structure, I followed the goal-setting literature in using "mere goals" – in other words, goals that simply establish a reference point (Heath *et al.*, 1999; Larrick *et al.*, 2009) – and then observing whether such a simple manipulation results in behavioral differences.

However, research has also shown that winner-takes-all incentives (the type described above) may boost risk taking and hence exploration (Ederer and Manso, 2013; Manso, 2011). If the incentive design applied to the first three studies does indeed encourage more risk taking and exploration, then the set-up will provide a conservative test of hypotheses that instead predict more exploitation. Yet because incentive design has a pronounced effect on learning and aspiration levels, I ran additional studies to examine the effect of different incentive structures on the

relationship under study. In particular, I ran the Studies 1 and 3 (since Study 2 is just a shorter version of the latter) with two alternative incentive structures. One structure ties performance directly to the aspiration levels by specifying a fixed bonus for achieving a clear goal expressed in accumulated payoff points. The other structure is a "pay for performance" set-up based on a stipulated exchange rate between payoffs earned and US dollar (USD) values. I integrated the latter into our manipulation of the aspiration level by way of the examples included in the design description. I expand on these incentive structures when discussing the results for these boundary condition studies.

Measures and Analyses

Our study's objective is to explain how exploration–exploitation behavior changes with different aspiration levels. I manipulate and measure *aspiration levels* by setting them and their payoff distributions, as described previously.

To test Hypotheses 1, I must capture exploitation frequency. An effective test of this hypothesis requires that exploration and exploitation behavior reflect the decision maker's perspective. I therefore estimate decision makers' beliefs as in Daw *et al.* (2006). Thus I estimate the decision makers' beliefs about which option is the best or worst in any given round by fitting participants' choices and payoffs into a *temporal difference learning algorithm* and a *softmax choice algorithm* (see Daw *et al.* 2006). The best-fitted model then yields estimates of the subjects' beliefs about each arm in each period, which enables our encoding of the decision maker's exploration and exploitation behavior over the 50 (or 80) rounds of the study. As described above, I define *Exploitation* as choosing the option that the focal decision

maker believes to be the best; exploration then amounts to choosing an options that he believes is *not* the best (exploration measured this way is the exact inverse of *Exploitation*).

Performance is measured in two ways. To test Hypotheses 2a and 3a, I measure how frequently a decision maker chooses the truly best option among the three choices (known to us but not known with certainty to the experiment's decision makers). In the stable environment (H2a), this option does not change over the course of the study; in the environmental shock condition (H3a), I measure how frequently decision makers choose the truly best *new* option after the shock in round 30. To assess Hypotheses 2b and 3b, I take the accumulated number of points earned by a decision maker over the 50 (or 80) rounds in order to compare relative performance (across aspiration levels).

Analyses. I analyze the data using *t*-tests and analysis of variation (ANOVA) for basic comparisons in addition to generalized linear mixed models for our hierarchical round–subject analyses (rounds are nested within participants); the Online Supplement presents basic logit and ordinary least-squares (OLS) regressions as robustness checks. In the round-level analyses, I follow suggestions in prior work (e.g., Billinger *et al.*, 2013; Levine *et al.*, 2019) and control for feedback variables, namely round number, prior round performance, prior average achieved so far, and time taken to make a choice. I also control for participant demographics such as age, gender, college major, and risk-taking propensity; the latter is measured by way of the balloon analogue risk task (BART; see Lejuez *et al.*, 2002).

Results

In discussing the results, I briefly summarize each of the study conditions and then document our findings. I hypothesized about the relative differences in exploration and exploitation choices by decision makers who face high versus moderate aspiration levels. Hence this section focuses on that comparison.

Study 1: Stable Environment

In Study 1, participants play the three-armed bandit game in a stable environment; the payoffs are uniformly distributed, with means [25, 20, 15] for arms 1, 2, and 3 (respectively) and a constant interval of ±10. The high (resp. moderate) aspiration level is set at 25 (resp. 20) points per round, accumulating to 1,250 (resp. 1,000) points over 50 rounds. Based on the decision maker's assignment of payoffs into the categories of "success" or "failure" relative to his aspiration level, the success percentages for the three arms are about [55%, 30%, 5%] for the high-aspiration condition and [80%, 55%, 30%] for the moderate-aspiration condition (this is illustrated in Figure 5).

Exploration–Exploitation Choices. I predicted that the high-aspiration decision makers will *exploit* more (H1) and hence explore less. I use ANOVA to test this hypothesis. I find that participants in the high-aspiration condition exploit the option they believe to be best significantly more often (mean $M = 29.2$, standard deviation $SD = 4.9$) than do those with a moderate aspiration level ($M = 27.4$, $SD = 4.0$; F-statistic = 8.17, p-value < 0.01) – outcomes that support Hypothesis 1. Support

for this hypothesis also implies that high-aspiration participants explore less than their moderate aspiration counterparts. The results are summarized in Table 2.

[[INSERT **Table 2** ABOUT HERE]]

Performance. I predicted that decision makers with high aspiration levels would be more likely than those with moderate aspiration to choose the truly best option (H2a) and would also exhibit better cumulative performance. I find that participants in the high-aspiration condition do choose arm 1, the objectively best arm, significantly more often (M = 29.1, SD = 5.4) than those with moderate levels of aspiration (M = 26.4, SD = 4.4; F = 15.18, $p < 0.01$) – providing support for Hypothesis 2a (see Table 2). In addition, cumulative performance is significantly higher in the high-aspiration condition (M = 1111.0, SD = 53.6) than in the moderate-aspiration one (M = 1080.6, SD = 57.8; F = 14.37, $p < 0.01$); these results support Hypothesis 2b.

Our theory suggests that the mechanism driving these outcomes is the greater strength of opinion that high–aspiration level participants develop in response to reduced feedback ambiguity. For this mechanism to be plausible, I first have to examine whether there is any indication that high-aspiration participants experience a reduction in feedback ambiguity. Recall that I manipulate feedback ambiguity by setting the two aspiration levels such that high-level participants should encode more feedback from the worst arm as failures than moderate-level ones. To illustrate that high-aspiration participants do indeed experience the feedback received differently, I use the success rate per arm as a proxy. I compute success rate for each arm by encoding the payoffs experienced above (resp. below) aspiration level as successes

(resp. failures). I then take the simple average across those experiences for each arm. For example, suppose that, prior to the focal period, a participant who chose a given arm had two experiences above aspiration and three experiences below aspiration; thus she experienced a success rate of 0.4 for that arm. Figure 6 presents the histogram of each arm's success rate as experienced by participants in the prior period (ordered by frequency of different success rates experienced, not by rounds). For high–aspiration level participants, these histograms show that the worst arm provides unambiguous feedback of "failure" (Panel (c) of the figure) whereas the best and middle arm provide feedback that is more ambiguous (see Panels (a) and (b)). For moderate–aspiration level participants, in contrast, all three arms provide ambiguous feedback.[14]

[[INSERT **Figure 6** ABOUT HERE]]

I exploit the longitudinal nature of our experiment to examine in more detail the mechanism of whether these differences in feedback ambiguity affect participants' strength of opinion which then drive the difference in exploitation behavior I find. For this purpose, I create a data set at the subject–round level so that I can analyze how prior feedback affects decisions in every round. I proxy strength of opinion by estimating participants' *belief difference between the best and the worst arm*. If high-aspiration level participants do indeed experience lower feedback ambiguity (as illustrated in Figure 5), their estimated belief difference between best and worst arm should be higher than that of moderate-aspiration level participants.

[14] Please refer to the Online Supplement for additional analyses of the success rates across arms and aspiration levels.

The reason for that is that almost every time they pull the worst arm (arm 3), they receive unambiguously negative feedback; as a result, their belief that the best (arm 1) gives different payoffs from the worst one *and* that arm 1 is superior will be quite strong. But moderate-aspiration participants receive relatively more positive feedback from arm 3 (the worst arm) *and* relatively less positive feedback from arm 1 (relative to high-aspiration participants), which makes their belief about the difference in payoffs between those two arms relatively weaker. Given the hierarchical nature of our data (i.e., rounds are nested within participants), I use a generalized linear mixed model to predict participants' exploitation of the best arm in each round.[15] In Stata, I use a Bernoulli distribution for the dependent variable "exploit" and a logit link function. I use the fixed effects model for the variables of interest and random effects per participant (since the subjects in this case were randomly sampled and allocated to conditions).

In Table 3, Model [1] replicates the ANOVA effects across periods. As expected, I observe that subjects have a higher likelihood to choose the arm they *believe* is best (i.e., exploit) in later rounds – an indication of learning – and when they have achieved high outcomes in previous rounds. I see also that participants' overall probability to exploit is lower when their propensity for risk taking (as measured by the BART score) is higher.[16] The values reported in Model [2] confirm

[15] The correlation matrix and robustness checks using OLS and logit models are given in the Online Supplement.

[16] I tested the effect of BART on the probability to exploit in all the studies. Even though the BART score was almost identical across the variety of populations I sampled, I only found a significant negative effect of the BART score on exploitation in the no-shock study and in Study 6b.

that high-aspiration participants exploit the best arm more often than do their moderate-aspiration counterparts.

[[INSERT **Table 3** ABOUT HERE]]

Finally, Models [3]–[5] in the table shed more light on the mechanism involved. Thus Model [3] shows that the focal participant's experience with different arms increases the likelihood of exploiting what he believes to be the best arm.[17] Specifically, it is the greater the difference in payoffs experienced from prior feedback about the best and worst arms' relative attractiveness that increases the participant's likelihood to exploit. Model [4] is included to show the robustness of the model, while Model [5] reveals a positive interaction effect between high aspiration level and the belief difference between the best and the worst arm. Incorporating that interaction effect into the regression renders non-significant the main effect of a high aspiration level. This finding provides further support for our hypothesized mechanism: by reducing feedback ambiguity, high aspiration levels evidently influence how feedback is interpreted, which results in greater strength of opinion which increases the likelihood of exploitation.

[[INSERT **Figure 7** ABOUT HERE]]

Overall, Study 1's baseline results indicate that high aspiration levels influence adaptive learning via a reduction in feedback ambiguity. This dynamic is illustrated in Figure 7. The graph shows that, when participants develop greater strength of opinion – i.e. greater belief difference between the best and worst arm –

[17] All participants exhibited some exploration behavior – enough so that all participants selected and received feedback about different arms.

participants can more rapidly hone in on exploiting the best arm; furthermore, this effect is stronger for participants with high aspirations. I conclude that, by reducing the ambiguity of feedback from payoffs, high (but not moderate) aspiration levels may improve cumulative performance in a stable environment.

Study 2: Unstable Environment with a Disruptive Shock

In this study I introduce a disruptive shock that reshapes the payoff landscape between periods 30 and 31. Before the shock, the three arms returned uniformly distributed payoffs with means [25, 20, 15] for arms 1, 2, and 3, respectively (just as in Study 1). After the shock, the means change to [20, 15, 25]; thus the worst pre-shock arm becomes the best post-shock arm, and the average payoff for the other two arms shifts downward (by 5 points each) so that the overall payoff landscape remains otherwise unchanged. All the arms have a constant interval of ±10 throughout the game.

Before the shock, results for Study 2 replicate those for Study 1: participants in the high–aspiration level condition exploit the arm they believe is best significantly more often (M = 27.2, SD = 9.0) than do those in the moderate–aspiration level condition (M = 23.6, SD = 7.0; F = 10.31, $p < 0.01$); so as before, H1 is supported. H2a is supported because high-aspiration participants exploit what is objectively the best arm significantly more often (M = 15.9, SD = 5.1) than do participants with moderate aspirations (M = 14.2, SD = 4.6; F = 5.94, p = 0.02); H2b is supported because cumulative performance is significantly higher in the high–aspiration level

group (M = 651.8, SD = 49.1) than in the moderate–aspiration level group (M =633.6, SD = 44.3; F = 7.3, $p < 0.01$).

Exploration-Exploitation Choices After the Shock. One implication of the adaptive learning model is that an environmental shock should have different effects on decision makers with high versus moderate aspirations. After the shock, high-aspiration participants were less likely to choose the new best option (M = 5.0, SD = 3.7) than were those with moderate aspirations (M = 7.5, SD = 4.1; F = 20.1, $p <$ 0.01) – supporting Hypothesis 3a. Cumulative performance after the shock is therefore significantly lower for participants in the high-aspiration condition (M = 406.3, SD = 33.1) than for those in the moderate-aspiration condition (M = 416.4, SD = 36.2; F = 4.31, $p < 0.05$), which supports Hypothesis 3b.[18]

Study 3: Unstable Environment with a Disruptive Shock and an Extended Post-Shock Period

In Study 2 I observe post-shock choices for only 20 periods (i.e., from period 31 to period 50). However, that duration may not be long enough to detect subjects' adaptation to the new environment. I therefore replicate Study 2 but with a longer post-shock learning period. In this study, there are 30 periods before the shock (from period 1 to 30, same as Study 2) and 50 periods after the shock (from period 31 to 80, instead of 20 periods as in Study 2).

[18] I also ran this study with greater variance (±15) from the mean, so that the worst arm in the high-aspiration condition returned a greater percentage of successes (15% instead of 5%). The results, which are qualitatively robust, have not been reported here owing to space constraints; they are available from the authors upon request.

Prior to the disruptive shock, results for Study 3 replicate those for Studies 1 and 2 (see Table 2). Recall that, according to the adaptive learning model, an environmental shock affects decision makers differently depending on their respective aspiration levels. This study helps explicate the short- and long-term effects of the shock on learning behavior and choices. Expanding the post-shock period allowed us to compare behavioral patterns – namely, choices made in the periods shortly after the shock (periods 31–50) with choices made in periods long after the shock (periods 61–80).

[[INSERT **Figure 8** ABOUT HERE]]

In Studies 1–3 I tested the hypotheses. Now, to establish boundary conditions for our major findings, I explore also the impact of different types of environments. Thus Study 4 describes the effects of a positive (rather than a neutral) shock, and Studies 5a, 5b, 6a, and 6b examine different incentive structures. I also include a discussion of robustness checks in which I the re-ran Studies 1 and 3 when a no-aspiration control condition is added.

Boundary Conditions

The sample characteristics for the studies described in this section were summarized above and can be found in Table 1. All results for these studies are summarized in Table 4.

[[INSERT **Table 4** ABOUT HERE]]

Shock Type: Unstable Environment with a Disruptive, Positive Shock. In Study 4, I introduce a *positive* shock between period 30 and 31 that changes only the worst

arm's payoff. Before the shock, the three arms returned uniformly distributed payoffs with means [25, 20, 15], but after the shock the means change to [25, 20, 30]. As in Study 2, the worst arm before the shock becomes the best arm afterwards; but unlike Study 2, the payoffs from choosing the other two arms remain unchanged. All the arms have a constant variance of ±10 throughout the game. The results mirror those reported in Study 2 and are summarized in Table 4.

In comparing Study 2 with Study 4, I expect that the latter's high–aspiration level participants will be more likely to continue choosing arm 1 (the previously best arm) than they were in Study 2. The reason is that, in Study 4, the payoff from arm 1 (the best arm pre-shock) does not decline post-shock – as it does in Study 2 – and so participants have no incentive to re-engage in exploration. Our data reveal that, after the shock, high-aspiration participants do choose arm 1 more frequently in Study 4 than in Study 2 (13.0 vs. 7.4; standard error SE = 0.72, $t = 2.88$, $p < 0.01$). (12.94 vs. 10.76, SD=5.28 vs.5.67, f=4.67, p<=0.05)

It is noteworthy also that, after the shock, Study 4's high–aspiration level participants were less likely to choose arm 3 (worst pre-shock arm but best post-shock arm) than were their Study 2 counterparts (3.3 vs. 5.0; SE = 0.411, $t = 4.13$, $p < 0.001$) (3.3 vs. 5.0, SD=3.5 vs 3.7, F=6.16, p<0.05). Thus cumulative performance post-shock is lower in Study 4 than in Study 2. In accordance with our expectations, the pre-shock choice patterns across these two studies do not exhibit any significant differences.[19]

[19] Because the assignment of participants to Study 2 versus Study 4 was not random, this comparison only uses the sample averages across studies.

Alternative Incentive Structures. Incentives play a leading role in individuals' choices between exploration and exploitation. The incentive design employed in Studies 1–3 was meant to minimize the link between aspiration levels and performance by incentivizing strong performance via a rare, global bonus. Research has shown that winner-takes-all incentives may increase risk taking, which in this context would suggest a shift toward more exploration (Ederer and Manso, 2013; Manso, 2011). Even if the incentives discussed previously did indeed encourage more risk taking and exploration, the design was actually a rather conservative test of the hypotheses that instead predicted more exploitation.

In order to explore the boundary conditions of incentive design in our set-up, I ran Studies 1 and 3 (again omitting Study 2, the shorter version of Study 3) with two alternative incentive structures each. All of these studies were run with a control condition under which participants could earn a bonus even though no aspiration level was specified. The actual aspiration levels were kept constant across incentive designs (i.e., in line with the results already reported): participants whose aspirations were high (resp. moderate) were requested to reach 25 (resp. 20) points per round or 1,250 (resp. 1,000) cumulative points.

The pandemic forced us to run these studies entirely via MTurk. Participants were paid a fixed fee for participation and were told that they could earn a performance-based bonus. The participant characteristics were broadly comparable to the initial MTurk Studies 1 and 3 as outlined above.

In Studies 5a and 5b, participants could earn a fixed bonus for reaching the given (moderate or high) aspiration level of 1,250 or 1,000 cumulative payoff points.

I implemented the manipulation for high [resp. moderate] aspiration levels by instructing the participants as follows: "If you earn more than 1,250 [1,000] points in 50 rounds, or average 25 [20] points per round, you will receive a bonus payment of $1 (and $0 bonus otherwise)." The control condition (no aspiration level) was worded thusly: "Remember, the more you earn the better off you are. If you earn more than average, you will receive a bonus payment of $1 (or $0 bonus otherwise)." In all other respects, Study 5a replicated Study 1 (a stable environment with 50 rounds) and Study 5b replicated Study 3 – an unstable environment with a disruptive shock after round 30 and a total of 80 rounds (just as in Study 3, the aspiration levels were scaled to 25 × 80 points or 20 × 80 points, respectively, for participants in the high– and moderate–aspiration level conditions).

The results from Studies 5a and 5b are consistent with those from Studies 1 and 3, respectively (see Table 4). The fixed-bonus incentive design ties performance more closely to the aspiration level and seems to reinforce the prevalence of that level in the minds of participants. Only H2b – our prediction that high–aspiration level participants would exhibit worse performance after the shock – was not supported under the fixed-bonus incentive design. This outcome may be due to the stronger performance focus in the fixed-bonus condition.

In Studies 6a and 6b, the points earned by participants were converted into US dollars, without a threshold, via an exchange rate. In this incentive structure, the preceding aspiration level manipulation was replicated and emphasized what the previous manager had earned (1,250 or 1,000 points, depending on the aspiration level); the only change was incorporating a points–USD exchange rate. I

implemented the manipulation for high [resp. moderate] aspiration levels by instructing the participants as follows: "The points you earn over your tenure will be converted to USD, with a conversion rate of 1 point = $0.001 (e.g., if you earn 1,250 [1,000] points you will get $1.25 [$1.00] bonus pay)." The control condition (no aspiration level) was worded in this way: "Remember, the more you earn the better off you are. The points you earn over your tenure will be converted to USD, with a conversion rate of 1 point = $0.001." Study 6a otherwise replicated Study 1, and Study 6b replicated Study 3.

The results in Table 4 show that this incentive design did not support our hypotheses. The only replicated effect is that participants in the high-aspiration condition tend to choose the truly best arm more frequently before the shock, but even this effect is rather weak. I interpret these results as evidence of participants' focus shifting away from aspiration levels because the strong link between performance and reward incentivizes exactly the risk taking (and hence exploration) reported by prior research (e.g., Ederer and Manso, 2013; Manso, 2011).[20] Hence one boundary condition for our findings is the need for an incentive design that supports – as in Studies 5a and 5b – rather than overrides aspiration levels.

Robustness Check: No–Aspiration Level Control Condition. Our theory development and model are based on the differences resulting from moderate versus high aspiration levels. In addition, the main variable used to assess H1 relies on fitting the participant-generated data to our model. I therefore ran the initial set-up of studies

[20] It is also possible that the aspiration level manipulation was just too weak in Studies 6a and 6b since the exchange rate between conditions did not differ.

(Studies 1, 2, and 3) without a separate control condition. So in evaluating the robustness of our aspiration level manipulation, I re-ran Studies 1 and 3 with a no–aspiration level control condition (as before, I omitted Study 2 since it is a shorter version of Study 3).

This control condition was presented as follows: "Remember, the more you earn the better off you are." Behavior under the control condition is very close, overall, to that under the moderate–aspiration level condition; this outcome confirms that feedback ambiguity in the study set-up was moderate (as it was designed to be). Analysis of the control condition alone shows that learning did take place in the absence of aspiration levels – and that it was, indeed, the high aspiration levels combined with reduced feedback ambiguity that resulted in greater exploitation, better performance in stable environments, and worse performance in post-shock, unstable environments.

Discussion

How individuals and organizations approach the trade-off between exploration and exploitation is an important and burgeoning area of scholarly inquiry. Whereas exploitation is essential for current performance, exploration is often required for long-term performance and even survival (Levinthal, 1997; Tushman and O'Reilly, 1996). Moreover, prior research suggests that exploration and exploitation activities are often in conflict (Benner and Tushman, 2003, 2015; March, 1991; Tushman and O'Reilly, 1996). Despite an abundance of scholarly research on how the exploration–exploitation trade-off affects choices and learning from experience, I still know

relatively little about how that trade-off is influenced by the decision makers' aspiration levels.

In this paper, I tackle a relatively under-studied aspect of the exploration-exploitation problem in management and organizations research, that of choice under uncertainty. Most prior work that considers how the aspiration performance gap influences exploration-exploitation decisions consider it from the vantage point of reinforcing or abandoning a reasonable well-known status-quo (Bromiley and Rau, 2019; Greve and Gaba, 2017). In contrast, empirical work on choice under uncertainty, where individuals learn about the relative attractiveness of choices from feedback, tends to ignore how goals (or aspirations) influence this choice (Cohen et al, 2007).

By combining two important (and well established) behavioral processes – learning from feedback and the subjective categorization of feedback around the aspiration level – I offer novel predictions about how aspiration levels influence exploration-exploitation decisions. Drawing from behavioral and reference point theories, I argue that aspiration levels are likely to affect how subjects interpret feedback (i.e., as success versus failure). In line with adaptive learning theory, I argue that this interpretation of feedback affects subsequent sampling behavior and thereby exacerbates any preference related to the different choices.

I then test the hypotheses arising from this model in a laboratory experiment that employs a multi-armed bandit task, which is the canonical model used for understanding how agents approach exploration and exploitation (Daw *et al.*, 2006; Gittins, 1979). I find that under certain conditions a high aspiration level reduces

feedback ambiguity about the relative attractiveness of available choices; I theorize that this reduction in feedback ambiguity allows decision makers to develop greater strength of opinion which, in turn, leads to more exploitation. I also show that such an exploitation focus is beneficial in stable environments, but detrimental in unstable environments, when a shock alters the relative attractiveness of the available options. Thus I contribute to the literature by showing how aspiration levels and adaptive learning *jointly* affect the choice between exploration and exploitation.

Our theory is especially applicable to the case where aspirations are exogenous (e.g., determined by a peer group or by superiors or external stakeholders) and where decision makers face choices whose performance consequences are not known ex ante and must therefore be learned from repeated feedback. This is the environment that March (1991) described in his seminal paper on the trade-off. Consider, as in the Introduction, a pharmaceutical company's R&D manager who makes resource allocation decisions about investing in uncertain technologies. This manager learns about the relative attractiveness of these different research programs by investing in them over some time period and then observing the outcomes. I posit that, in this case, a manager with a relatively higher aspiration level is more likely (than a manager with moderate aspirations) to exploit, which may lead to higher performance outcomes. However, such a manager is also less likely to notice technology breakthroughs that improve the payoffs of previously less attractive choices.[21] In the case of exogenously imposed – rather than personal, intrinsic –

[21] This generalization may not hold at extremely high aspiration levels (i.e., those above any currently available choices) or at aspiration levels so low that they are satisfied by *all* choices. In these instances, feedback ambiguity is not lower for one aspiration level as opposed to another aspiration level. Thus, I should not expect differences in exploration behavior across these levels.

aspiration levels, the theory and data presented here should have fairly broad business applicability.

I consider the impact of aspiration levels in an adaptive learning context, where I account not only for a single instance of choice (i.e., a single selection from a set of known options) but also for repetitive choice. The results show that increased aspiration levels actually lead to more exploitation and that the agents who do explore end up considering a narrower set of choices. The implication is that while leaders who challenge their organizations with ambitious goals may unleash exploration in some parts of the organization, they are actually making other parts of the organization more conservative. For example, a high aspiration level may encourage the engineering or marketing department to try out new ideas, yet the capital budgeting process may hinder that exploration by dictating safer alternatives. So in order for goal setting and/or aspiration levels to render exploration both desirable and fruitful, the goals must be sufficiently ambitious that they trigger search for new alternatives – that is, rather than a choice between existing ones. Nevertheless, Sitkin *et al.* (2011) and Gary *et al.* (2017) stressed that each of these responses comes with its own drawbacks.

Our findings may seem counterintuitive when one considers the prevailing view in research on the aspiration-performance link, which has largeky focused on the relative attractiveness of a well-known status quo vis a vis uncertain alternatives (Greve and Gaba, 2017; Ordóñez *et al.*, 2009a, 2009b). By instead framing this link as a problem of choice under uncertainty I were able to explore how aspiration levels may affect learning and the explore–exploit decision as well as the interpretation of

payoff feedback as either success or failure. Such a subjective interpretation affects subsequent sampling decisions, and it may reduce the rates of exploration and exploitation asymmetrically depending on whether the decision maker has high or moderate aspirations. Although our agent-based simulation and lab studies allowed us to make concrete predictions about the interplay between aspiration levels and adaptive learning about the exploration–exploitation trade-off, I believe that this paper has but scratched the surface of a promising avenue of enquiry.

List of Figures for Essay A

Figure 1. Thought experiment: Two decision makers and two arms

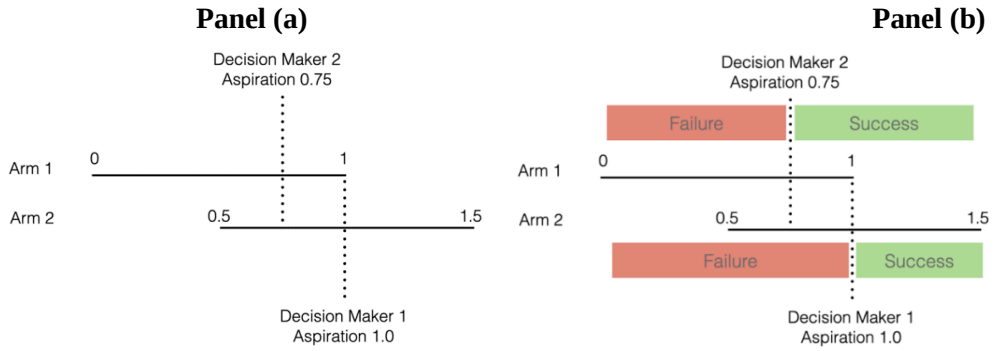
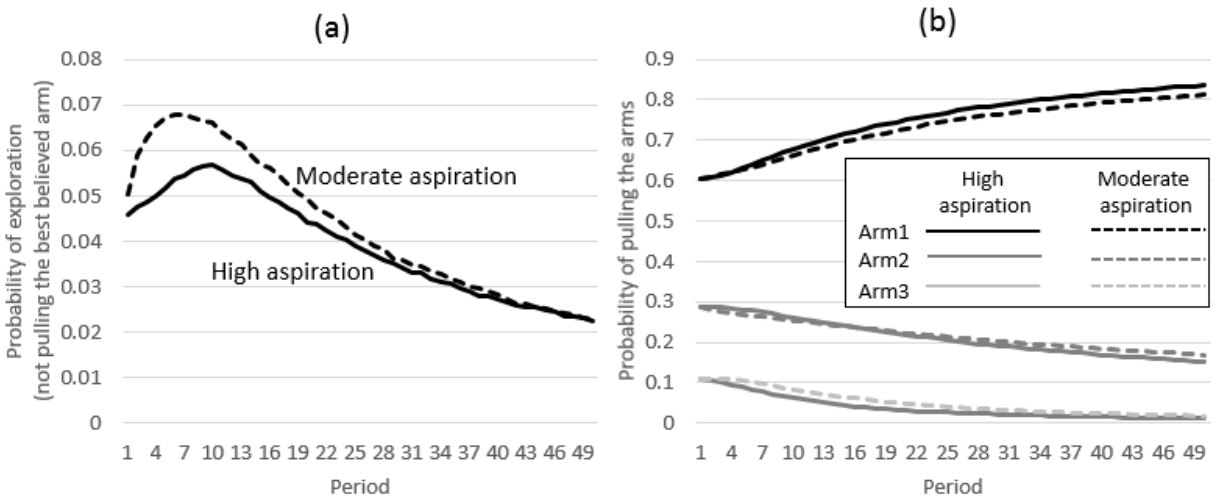
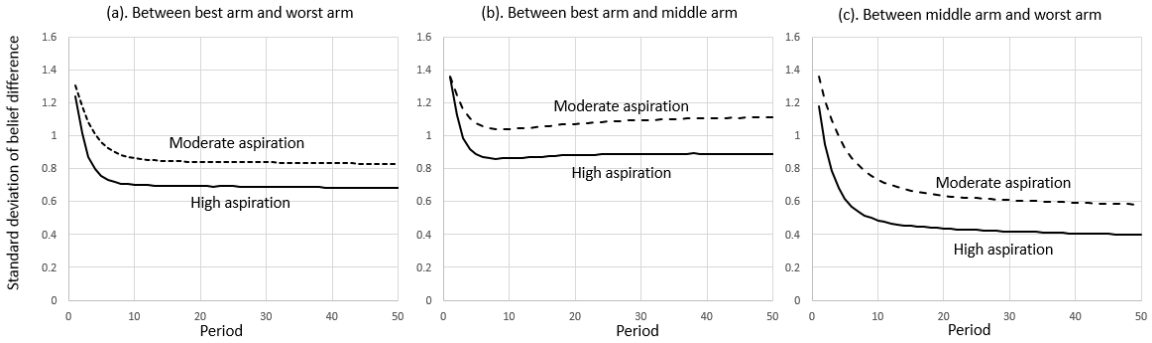


Figure 2. Probability of exploration and choosing each arm: 50 periods, no disruptive shock



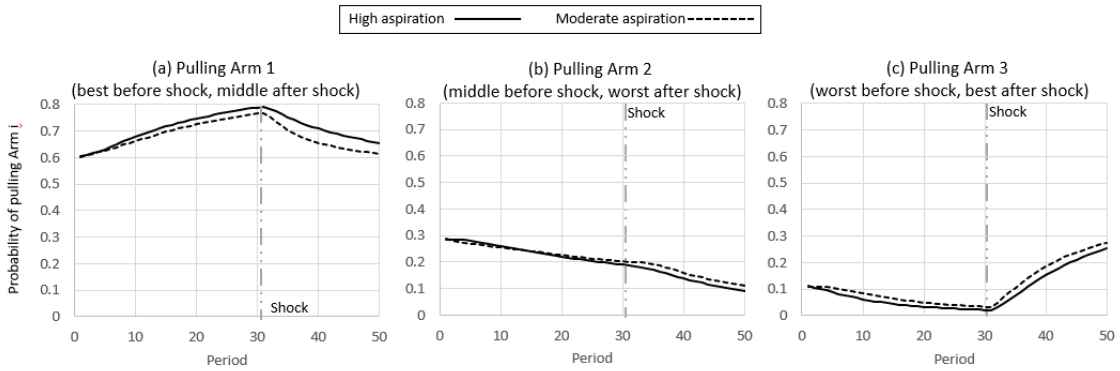
Notes: High–aspiration level agents are less likely to explore, and more likely to choose the best arm.

Figure 3. Standard deviation of difference in beliefs about arms' payoffs



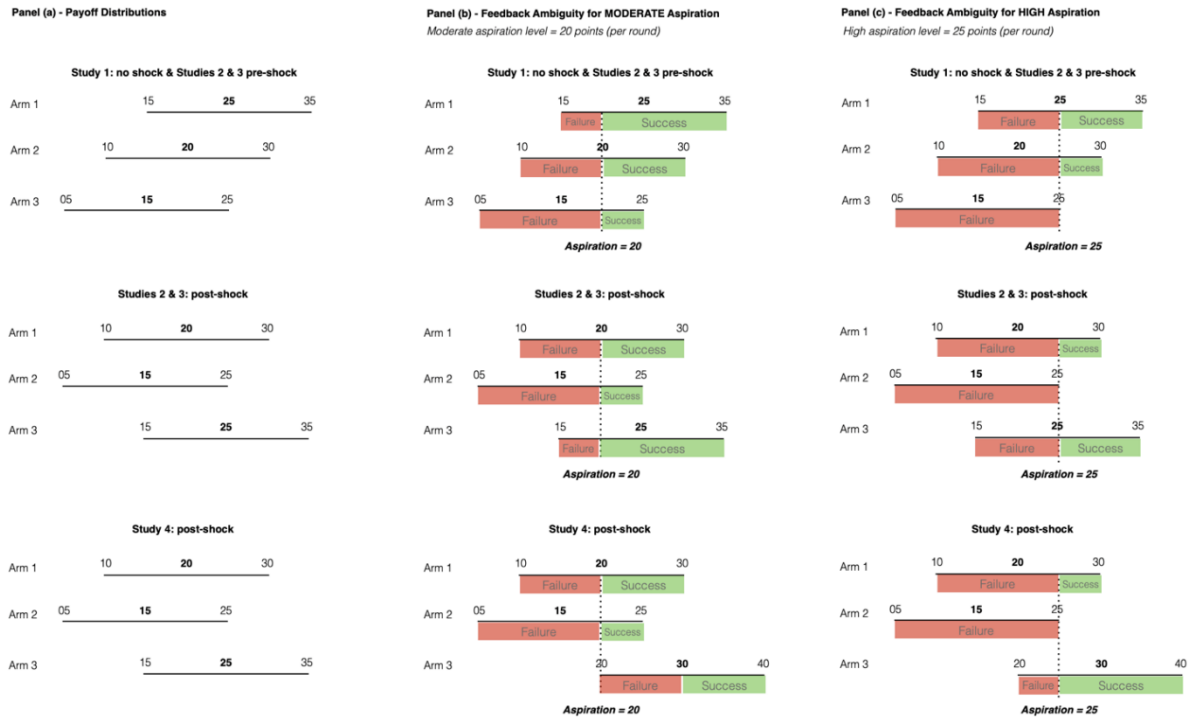
Notes: The smaller the standard deviation of belief difference between two arms, the less the agent's feedback ambiguity about whether one arm is better than the other. High aspiration levels always lead to a lower standard deviation of these differences between arms. Therefore, high-aspiration agents experience less ambiguity concerning feedback about which arm is better.

Figure 4. Probability of choosing each arm: Disruptive shock after period 30



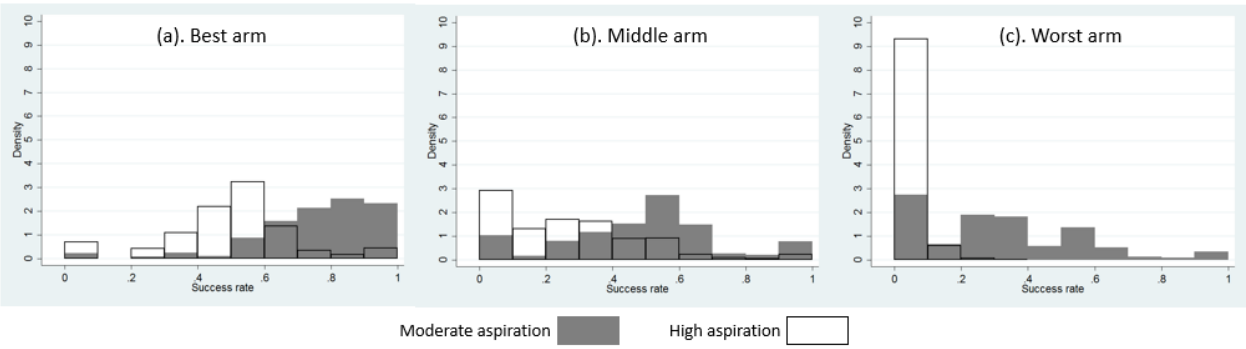
Notes: A high aspiration level leads to a strong preference for the best arm with little or no exploration of inferior alternatives. When a disruptive shock alters the payoff structure, this strong preference makes it difficult for high-aspiration agents to identify quickly the new, superior alternative.

Figure 5. Experimental setup: payoff distributions and feedback ambiguity



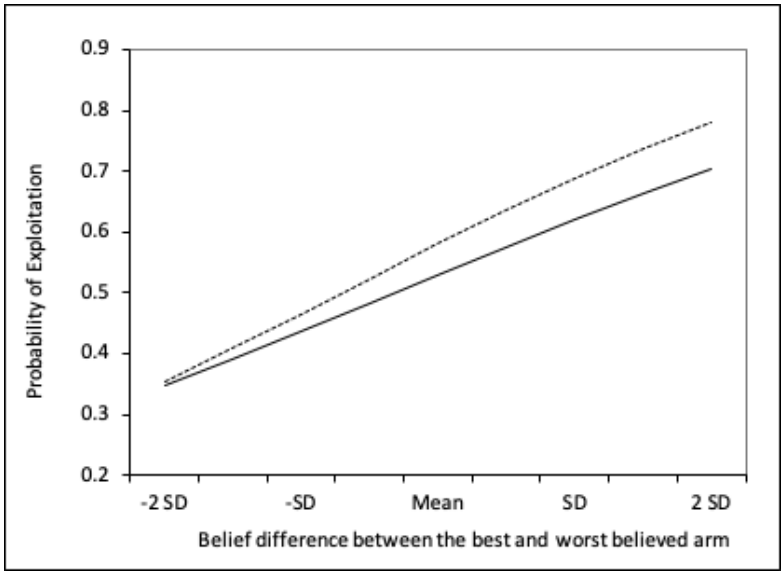
Notes: Panel (a) illustrates the payoff distributions of the three investment options that underlie Studies 1, 2, 3, and 4. Panels (b) and (c) overlay the "success" and "failure" interpretations made by the decision makers based on their respective aspiration levels. Panel (b) (resp. (c)) shows the interpretations for a moderate (resp. high) aspiration level of 20 (resp. 25) payoff points. A comparison of these two panels reveals that the feedback ambiguity for all arms is much reduced in the high relative to the moderate aspiration condition.

Figure 6. Success rate for each arm (subject–round level), Study 1



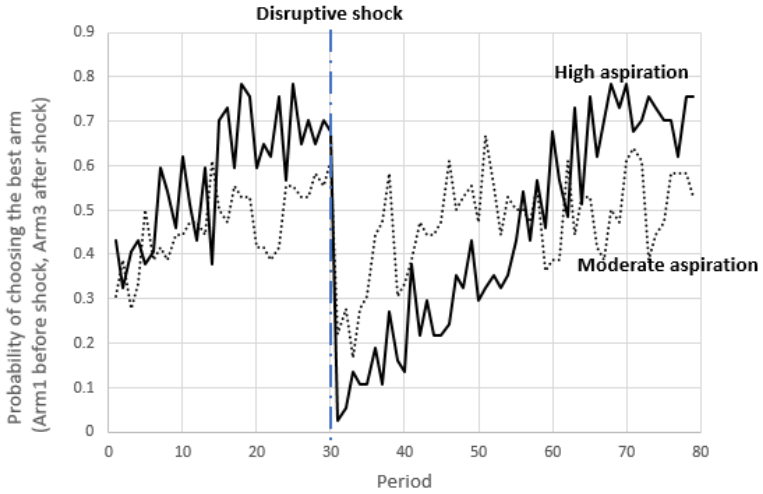
Notes: For each subject and round, the graphs plot the distribution of success rates in the prior round across the three arms. For example, a participant who had two experiences above aspiration and three experiences below aspiration – that is, from choosing an arm prior to the focal period – has experienced a success rate of 0.4 for that arm. A high aspiration level (but not a moderate one) leads to consistent feedback of failure from the worst arm (panel (c)).

Figure 7. Average feedback difference and probability of exploiting the best arm as a function of aspiration level (coefficients from Table 4, Model [5])



Note: For a given increase in belief difference between the best and the worst arm, high–aspiration level participants (dashed line) exhibit a greater increase in their probability of exploiting the best arm than do medium–aspiration level participants (solid line).

Figure 8. Study 3: Pre- and post-shock likelihood of choosing the best arm



List of Tables for Essay A

Table 1. Sample characteristics for main studies

MAIN RESULTS										
Study	**Study condition**	**Total number of participants**	**Number of participants in high aspiration condition**	**Location / Source**	**Mean age [range]**	**Female (%)**	**BART Score**	**English is first language (%)**	**Country of residence (North America %)**	**Education (College degree %)**
1	Stable environment: no shock	193 (220)	98	MTurk	35.0 [21, 70]	31.3	39.3	92.6	88.5	69.1
2	Unstable environment, disruptive shock: replicates Study 1 *with* shock	200 (220)	103	US university	20.8 [19, 31]	45.9	40.0	55.0	N/A	N/A
3	Unstable environment, disruptive shock *and* extended post-shock period: replicates Study 2 but with an additional 30 rounds post-shock	73 (80)	37	MTurk	35.4 [22, 67]	34.6	40.4	95.1	92.6	63.0
BOUNDARY CONDITIONS										
Study	**Study condition**	**Total number of participants**	**Number of participants in high aspiration condition**	**Location / Source**	**Mean age [range]**	**Female (%)**	**BART Score**	**English is first language (%)**	**Country of residence (North America %)**	**Education (College degree %)**
4	Unstable environment, disruptive *positive* shock: shifts average payoffs upward (compared with Study 2)	77 (90)	39	Singapore university	20.7 [19, 24]	45.5	40.0	N/A	66.2 (Singaporean)	85.7 Business school (vs other department)
5a	Replicates Study 1 but with fixed bonus incentive	107 (128)	42	MTurk	36 [20, 70]	38.0	39.7	82.2	70.5	72.1
5b	Replicates Study 3 but with fixed bonus incentive	98 (115)	37	MTurk	35 [20, 67]	31.4	40.0	83.1	69.5	61.0
6a	Replicates Study 1 but with exchange rate-based incentive	100 (122)	40	MTurk	36 [19, 69]	26.2	39.8	83.6	71.3	66.4
6b	Replicates Study 3 but with exchange rate-based incentive	96 (119)	39	MTurk	35.5 [22, 69]	30.7	38.7	84.9	71.4	63.9

Notes: The sample size *N* corresponds to usable data, and the total number of recruited participants is given in parentheses. The slight reduction in each study's number of subjects is due to some participants failing to fill out all parts of the questionnaire or failing to "finalize" the main game. In the former case we performed, when possible analyses on each sample (both the full and incomplete ones); no significant differences were observed. The MTurk studies were not regionally restricted; the BART (balloon analogue risk task) score captures participants' risk taking propensity (see Lejuez *et al.*, 2002).

Table 2. ANOVA results for main studies

Hypothesis	Findings	Study 1 - stable environment			Study 2 - shock			Study 3 - extended post-shock		
compared to moderate aspiration level condition		Aspiration level		Difference betw.	Aspiration level		Difference betw.	Aspiration level		Difference betw.
Across rounds		Moderate	High	conditions	Moderate	High	conditions	Moderate	High	conditions
H1	High–aspiration level agents choose the option that they believe is best (exploit) more frequently than do agents of moderate aspiration level.	27.39 (3.97)	29.23 (4.89)	$d = 1.84$ $F = 8.17$ $p = 0.0047$	23.57 (7.02)	27.24 (8.98)	$d = 3.67$ $F = 10.31$ $p = 0.0015$	40.92 (4.48)	51.70 (5.66)	$d = 10.78$ $F = 81.27$ $p < 0.001$
Pre-shock results										
H2a	High–aspiration level agents choose the truly best option more frequently in a stable environment than do agents with moderate aspiration levels.	26.35 (4.35)	29.09 (5.36)	$d = 2.74$ $F = 15.18$ $p = 0.0001$	14.22 (4.55)	15.89 (5.14)	$d = 1.67$ F = 5.94 $p = 0.0157$	14.22 (3.00)	17.51 (4.79)	$d = 3.29$ $F = 12.32$ $p = 0.0008$
H2b	High–aspiration level agents exhibit better cumulative performance in a stable environment than do agents of moderate aspiration level.	1080.57 (57.80)	1110.96 (53.55)	$d = 30.39$ $F = 14.37$ $p = 0.0002$	633.56 (44.29)	651.83 (49.10)	$d = 18.27$ $F = 7.3$ $p = 0.0064$	640.17 (37.18)	668.84 (43.79)	$d = 28.67$ $F = 9.07$ $p = 0.0036$
Post-shock results ***(rounds 31 - 50)***										
H3a	High–aspiration level agents choose the newly changed best option less frequently after a disruptive environmental shock than do agents with moderate aspirations.				7.45 (4.07)	4.99 (3.70)	$d = 2.46$ $F = 20.08$ $p < 0.001$	8.03 (2.74)	4.08 (2.95)	$d = 3.95$ $F = 34.93$ $p < 0.001$
H3b	High–aspiration level agents exhibit worse performance directly after a disruptive environmental shock than do agents of moderate aspiration level.				416.44 (36.22)	406.27 (33.10)	$d = 10.17$ $F = 4.31$ $p = 0.0393$	418.03 (29.96)	399.38 (29.58)	$d = 16.65$ $F = 7.16$ $p = 0.0092$
								rounds 51-80		
H2a	High–aspiration level agents choose the truly best option more frequently in a stable environment than do agents with moderate aspiration levels.							15.17 (3.32)	17.83 (5.90)	$d = 2.67$ $F = 5.64$ $p = 0.0203$
H2b	High–aspiration level agents exhibit better cumulative performance in a stable environment than do agents of moderate aspiration level.							647.39 (42.36)	676.78 (36.92)	$d = 29.39$ $F = 10.01$ $p = 0.0023$

Note: This table reports sample means with standard errors in parentheses. The "Difference between conditions" column reports the ANOVA results.

Table 3. Study 1: Generalized linear mixed model of exploiting the best arm

	Model				
Variable	**[1]**	**[2]**	**[3]**	**[4]**	**[5]**
High aspiration		0.138**		0.211***	–0.001
		(0.054)		(0.055)	(0.119)
Belief difference between best and worst arm			0.460***	0.467***	0.415***
			(0.032)	(0.032)	(0.041)
Belief difference between best and worst arm × High aspiration					0.106**
					(0.053)
Last round performance	0.012***	0.012***	–0.002	–0.002	–0.003
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Average past performance	–0.029**	-0.029**	–0.047***	–0.046***	–0.048***
	(0.011)	(0.011)	(0.012)	(0.012)	(0.012)
Round	0.014***	0.014***	0.004**	0.004**	0.004**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Time spent this round (seconds)	–0.010	-0.010	–0.000	0.000	0.001
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Age	–0.000	-0.000	–0.003	–0.003	–0.003
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Male	–0.003	-0.009	–0.004	–0.014	–0.009
	(0.059)	(0.058)	(0.060)	(0.058)	(0.058)
BART	–0.018***	-0.016**	–0.017**	–0.014**	–0.013**
	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)
Constant	0.975***	0.833**	1.024***	0.800**	0.916**
	(0.369)	(0.368)	(0.376)	(0.370)	(0.372)
Observations	9,457		9,457	9,457	9,457
Number of groups	193		193	193	193

Standard errors are reported in parentheses**$p < 0.05$, ***$p < 0.01$

Table 4. ANOVA results for boundary condition studies

Hypothesis	Findings	Study 4 - positive payoff shock			Study 5a - stable environment - fixed bonus			Study 5b - shock, extended - fixed bonus			Study 6a - stable environment - exchange rate			Study 6b - shock, extended - exchange rate		
	compared to moderate aspiration level condition	Aspiration level		Difference betw. conditions	Aspiration level		Difference betw. conditions	Aspiration level		Difference betw. conditions	Aspiration level		Difference betw. conditions	Aspiration level		Difference betw.
		Moderate	High		Moderate	High		Moderate	High		Moderate	High		Moderate	High	
Across rounds																
H1	High–aspiration level agents choose the option that they believe is best (exploit) more frequently than do agents of moderate aspiration level.	24.68 (5.92)	29.80 (9.53)	$d = 5.12$ $F = 7.94$ $p = 0.0062$	28.13 (10.55)	33.60 (10.15)	$d = 5.47$ $F = 5.22$ $p = 0.025$	41.29 (16.08)	51.31 (13.16)	$d = 10.02$ $F = 7.35$ $p = 0.009$	29.09 (11.02)	32.10 (9.77)	$d = 3.01$ $F = 1.35$ $p = 0.2505$	41.19 (13.12)	46.00 (14.85)	$d = 4.81$ $F = 1.83$ $p = 0.1807$
Pre-shock results																
H2a	High–aspiration level agents choose the truly best option more frequently in a stable environment than do agents with moderate aspiration levels.	13.00 (4.07)	16.39 (5.78)	$d = 3.39$ $F = 8.78$ $p = 0.0041$	26.00 (9.97)	32.27 (10.24)	$d = 6.27$ $F = 7.22$ $p = 0.009$	12.81 (4.90)	19.41 (6.57)	$d = 6.6$ $F = 14.65$ $p = 0.0003$	26.21 (10.78)	31.03 (9.93)	$d = 4.82$ $F = 3.51$ $p = 0.0658$	13.10 (5.78)	15.80 (5.55)	$d = 2.7$ $F = 3.53$ $p = 0.0651$
H2b	High–aspiration level agents exhibit better cumulative performance in a stable environment than do agents of moderate aspiration level.	632.50 (44.71)	655.56 (52.54)	$d = 23.06$ $F = 4.29$ $p = 0.0417$	1069.63 (69.35)	1124.73 (70.32)	$d = 55.1$ $F = 11.67$ $p = 0.001$	645.68 (41.35)	680.60 (51.29)	$d = 34.92$ $F = 8.81$ $p = 0.0043$	1089.47 (80.82)	1113.55 (76.75)	$d = 24.08$ $F = 1.51$ $p = 0.2237$	649.22 (45.41)	662.13 (51.11)	$d = 12.91$ $F = 1.11$ $p = 0.2965$
Post-shock results *(rounds 31 - 50)*																
H3a	High–aspiration level agents choose the newly changed best option less frequently after a disruptive environmental shock than do agents with moderate aspirations.	8.05 (4.42)	3.28 (3.54)	$d = 4.77$ $F = 27.43$ $p < 0.0001$				8.36 (3.80)	5.31 (4.74)	$d = 3.05$ $F = 7.88$ $p = 0.0067$				6.78 (4.32)	6.60 (3.84)	$d = 0.18$ $F = 0.03$ $p = 0.8622$
H3b	High–aspiration level agents exhibit worse performance directly after a disruptive environmental shock than do agents of moderate aspiration level.	512.95 (34.53)	495.67 (33.73)	$d = 17.28$ $F = 4.93$ $p = 0.0293$				416.58 (36.08)	406.00 (36.50)	$d = 10.58$ $F = 1.34$ $p = 0.2519$				401.69 (39.65)	412.97 (28.94)	$d = 11.28$ $F = 1.62$ $p = 0.2082$
								rounds 51-80						***rounds 51-80***		
H2a	High–aspiration level agents choose the truly best option more frequently in a stable environment than do agents with moderate aspiration levels.							15.71 (6.56)	13.46 (9.15)	$d = 2.24$ $F = 1.24$ $p = 0.2695$				15.66 (7.95)	13.33 (6.73)	$d = 2.32$ $F = 1.53$ $p = 0.22$
H2b	High–aspiration level agents exhibit better cumulative performance in a stable environment than do agents of moderate aspiration level.							642.74 (64.39)	634.88 (68.48)	$d = 7.87$ $F = 0.22$ $p = 0.6404$				630.37 (59.42)	642.76 (53.13)	$d = 12.39$ $F = 0.75$ $p = 0.3913$

Note: This table reports sample means with standard errors in parentheses. The "Difference between conditions" column reports the ANOVA results.

References for Essay A

Baum, J. A., Rowley, T. J., Shipilov, A. V., Chuang, Y.-T. (2005). Dancing with strangers: Aspiration performance and the search for underwriting syndicate partners. *Adm. Sci. Q., 50*(4), 536-575.

Benner, M. J., Tushman, M. L. (2003). Exploitation, exploration, and process management: The productivity dilemma revisited. *Acad. Manag. Rev., 28*(2), 238-256.

Benner, M. J., Tushman, M. L. (2015). Reflections on the 2013 Decade Award—"Exploitation, exploration, and process management: The productivity dilemma revisited" ten years later. *Acad. Manag. Rev., 40*(4), 497-514.

Bromiley, P. (2010). Looking at prospect theory. *Strategic Manag. J., 31*(12), 1357-1370.

Bromiley, P., Miller, K. D., Rau, D. (2001). Risk in strategic management research. *The Blackwell Handbook of Strategic Management*, 259-288.

Bromiley, P., Rau, D. (2019). Some problems in using prospect theory to explain strategic management issues (in press). *Acad. Manag. Perspect.*

Cervone, D., Jiwani, N., Wood, R. (1991). Goal setting and the differential influence of self-regulatory processes on complex decision-making performance. *J. Pers. Soc. Psychol., 61*(2), 257.

Cohen, J. D., McClure, S. M., Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1481), 933-942.

Cyert, R. M., March, J. G. (1963). *A behavioral theory of the firm* (Vol. 2). Englewood Cliffs, NJ.

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.

Denrell, J. (2008). Organizational risk taking: Adaptation versus variable risk preferences. *Ind. Corp. Change, 17*(3), 427-466.

Denrell, J., Le Mens, G. (2011). Seeking positive experiences can produce illusory correlations. *Cognition, 119*(3), 313-324.

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

Earley, P. C., Connolly, T., Ekegren, G. (1989). Goals, strategy development, and task performance: Some limits on the efficacy of goal setting. *J. Appl. Psychol., 74*(1), 24.

Edmonds, B. (2001). Towards a descriptive model of agent strategy search. *Comput. Econ., 18*(1), 111-133.

Eisenhardt, K. M., Martin, J. A. (2000). Dynamic capabilities: what are they? *Strategic Manag. J., 21*(10-11), 1105-1121.

Fiegenbaum, A. (1990). Prospect theory and the risk-return association: An empirical examination in 85 industries. *J. Econ. Behav. Organ., 14*(2), 187-203.

Fiegenbaum, A., Hart, S., Schendel, D. (1996). Strategic reference point theory. *Strategic Manag. J., 17*(3), 219-235.

Fiegenbaum, A., Thomas, H. (1995). Strategic groups as reference groups: Theory, modeling and empirical examination of industry and competitive strategy. *Strategic Manag. J., 16*(6), 461-476.

Frank, M. J., Doll, B. B., Oas-Terpstra, J., Moreno, F. (2009). Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nature Neuroscience, 12*(8), 1062.

Gaba, V., Joseph, J. (2013). Corporate structure and performance feedback: Aspirations and adaptation in M-form firms. *Organ. Sci., 24*(4), 1102-1119.

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

Gary, M. S., Yang, M. M., Yetton, P. W., Sterman, J. D. (2017). Stretch goals and the distribution of organizational performance. *Organ. Sci., 28*(3), 395-410.

Gigerenzer, G. (2008). Gut Feelings: Shortcuts to better decision making. London, Uk: Penguin Books.

Gigerenzer, G., Gaissmaier, W. (2011). Heuristic decision making. *Annu. Rev. Psychol., 62*, 451-482.

Gittins, J. C. (1979). Bandit processes and dynamic allocation indices. *J. Royal Stat. Soc.: Series B, 41*(2), 148-164.

Greve, H. R. (2002). Sticky aspirations: Organizational time perspective and competitiveness. *Organ. Sci., 13*(1), 1-17.

Greve, H. R. (2003). A behavioral theory of R&D expenditures and innovations: Evidence from shipbuilding. *Acad. Manag. J., 46*(6), 685-702.

Greve, H. R. (2007). Exploration and exploitation in product innovation. *Ind. Corp. Change, 16*(5), 945-975.

Greve, H. R. (2010). Designing performance feedback systems to guide learning and manage risk. *Organ. Dyn., 39*(2), 104-114.

Greve, H. R. (2017). Organizational learning and adaptation *Oxford Research Encyclopedia of Business and Management*.

Greve, H. R. (2018). Where to Search? *Behavioral Strategy in Perspective* (pp. 91-100): Emerald Publishing Limited.

Greve, H. R., Gaba, V. (2017). Performance feedback in organizations and groups: Common themes: Oxford University Press: Oxford, UK.

Heath, C., Larrick, R. P., Wu, G. (1999). Goals as reference points. *Cognitive Psychology, 38*(1), 79-109.

Junni, P., Sarala, R. M., Taras, V., Tarba, S. Y. (2013). Organizational ambidexterity and performance: A meta-analysis. *Acad. Manag. Perspect., 27*(4), 299-312.

Kanfer, R., Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *J. Appl. Psychol., 74*(4), 657.

Knox, W. B., Otto, A. R., Stone, P., Love, B. (2012). The nature of belief-directed exploratory choice in human decision-making. *Front. Psychol., 2*, 398.

Lant, T., Shapira, Z. (2008). Managerial reasoning about aspirations and expectations. *J. Econ. Behav. Organ., 66*(1), 60-73.

Larrick, R. P., Heath, C., Wu, G. (2009). Goal-induced risk taking in negotiation and decision making. *Sco. Cogn., 27*(3), 342-364.

Le Mens, G., Denrell, J. (2011). Rational learning and information sampling: On the "naivety" assumption in sampling explanations of judgment biases. *Psychol. Rev., 118*(2), 379.

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., ... & Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: the Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied, 8(2),* 75.

Lee, E., Puranam, P. (2016). The implementation imperative: Why one should implement even imperfect strategies perfectly. *Strategic Manag. J., 37*(8), 1529-1546.

Levinthal, D. A. (1997). Adaptation on rugged landscapes. *Manage. Sci., 43*(7), 934-950.

Levitt, B., March, J. G. (1988). Organizational learning. *Annu. Rev. Sociol, 14*(1), 319-338.

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

Locke, E. A., Latham, G. P. (2006). New directions in goal-setting theory. *Current Directions in Psychological Science, 15*(5), 265-268.

March, J. G. (1988). Variable risk preferences and adaptive aspirations. *J. Econ. Behav. Organ., 9*(1), 5-24.

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

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

Ordónez, L. D., Schweitzer, M. E., Galinsky, A. D., Bazerman, M. H. (2009b). On good scholarship, goal setting, and scholars gone wild. *Acad. Manag. Perspect., 23*(3), 82-87.

Ordóñez, L. D., Schweitzer, M. E., Galinsky, A. D., Bazerman, M. H. (2009a). Goals gone wild: The systematic side effects of overprescribing goal setting. *Acad. Manag. Perspect., 23*(1), 6-16.

Posen, H. E., Keil, T., Kim, S., Meissner, F. D. (2018). Renewing research on problemistic search—A review and research agenda. *Acad. Manag. Annals, 12*(1), 208-251.

Posen, H. E., Levinthal, D. A. (2012). Chasing a moving target: Exploitation and exploration in dynamic environments. *Manage. Sci., 58*(3), 587-601.

Puranam, P., Stieglitz, N., Osman, M., Pillutla, M. M. (2015). Modelling bounded rationality in organizations: Progress and prospects. *Acad. Manag. Annals, 9*(1), 337-392.

Schunk, D. (2009). Behavioral heterogeneity in dynamic search situations: Theory and experimental evidence. *J. Econ. Dyn. Control, 33*(9), 1719-1738.

Shinkle, G. A. (2012). Organizational aspirations, reference points, and goals: Building on the past and aiming for the future. *J. Manag., 38*(1), 415-455.

Simon, H. A. (1955). A behavioral model of rational choice. *Q. J. Econ., 69*(1), 99-118.

Simon, H. A. (1997). *Administrative Behavior: A study of decision-making processes in administrative organizations* (4th ed.). New York: Free Press.

Sitkin, S. B., See, K. E., Miller, C. C., Lawless, M. W., Carton, A. M. (2011). The paradox of stretch goals: Organizations in pursuit of the seemingly impossible. *Acad. Manag. Rev., 36*(3), 544-566.

Sterman, J. D. (1989). Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. *Manage. Sci., 35*(3), 321-339.

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

Tushman, M. L., O'Reilly, C. A. (1996). Ambidextrous organizations: Managing evolutionary and revolutionary change. *Calif. Manag. Rev., 38*(4), 8-29.

Wood, R., Bandura, A., Bailey, T. (1990). Mechanisms governing organizational performance in complex decision-making environments. *Organ. Behav. Hum. Decis. Process., 46*(2), 181-201.

Zheng, L., Srikanth, K. (2019). Divide and conquer - When and why exploration and exploitation should be separated. *Academy of Management Proceedings., 2019(1)*.

ESSAY B: Responding to environment turbulence – a bandit model

Introduction

Committing to an appropriate exploration strategy is critical for organizational performance; either over- or under-exploring leads to underperformance and organizational failures (March, 1991). Yet, how much exploration is appropriate, is subject to the environmental dynamism. Organizational response to increased environmental dynamism is an important question (Siggelkow & Rivkin, 2005; Sørensen & Stuart, 2000; Stieglitz, Knudsen, & Becker, 2016; Tushman & Romanelli, 1985): when the environment dynamism increases, should organizations focus more (or less) on exploration? This question is relevant for managers. For example, should traditional auto makers, in response to the volatile oil price, invest more in developing electronic or hybrid vehicles that is beyond their existing knowledge about combustion engines? Should MNEs explore new markets when facing increased turbulence in global economic growth?

Existing literature on organizational response to environmental dynamism has provided mixed suggestions. On one hand, some scholars (Eisenhardt & Tabrizi, 1995; Jansen, Van Den Bosch, & Volberda, 2006; March, 1991) suggest that environmental dynamism makes existing knowledge obsolete, and organizations need to renew their knowledge by increasing exploration. Exploration facilitates the identification of emerged opportunities in a dynamic environment, thus helps organizations' adaptation to the new environment (Uotila, Maula, Keil, & Zahra, 2009; Zahra, 1996). On the other hand, some scholars also argue that increased dynamism undermines the value of exploration (Hannan & Freeman, 1984; Marino, Aversa, Mesquita, & Anand, 2015; Posen & Levinthal, 2012). Posen and Levinthal (2012) illustrate the problem of "chasing a moving target" and challenged the

conventional wisdom about the positive relationship between environmental dynamism and exploration: the knowledge generated by exploration will be quickly eroded in a dynamic environment, thus organization needs to reduce exploration as a response. The opposing viewpoints pose not only conflicting theoretical predictions, but also conflicting managerial guidelines for practitioners (Stieglitz et al., 2016).

Noticing that the existing literature uses single construct for exploration strategy, we revisit the "chasing a moving target" problem and reconcile the opposing viewpoints by disentangling two distinctive aspects of exploration - intensity and width. Specifically, exploration intensity refers to the "*how much*" questions: the amount of attention and resources allocated to exploration, such as R&D intensity (Cohen & Levinthal, 1990) or number of new product introduced (Greve, 2007). Exploration width refers to the "*how*" question: the amount of attention and resources allocated to distant, unfamiliar and risky exploration (versus less distant, unfamiliar and risky exploration), such as boundary-spanning exploration (Rosenkopf & Nerkar, 2001) or new product that are meaningfully distinct from existing ones (Danneels & Sethi, 2011). Using a computational model, we examine how organization should respond to increased environmental dynamism. We find that, if an organization is able to disentangle exploration intensity and width, it can effectively address the problem of "chasing a moving target": not only the optimal exploration (both intensity and width) will monotonically increase as the environment becomes dynamic, the organizations will also achieve superior performance in a dynamic environment than others who are unable to disentangle exploration intensity and width.

Through our model, we highlight two observations. First, in a dynamic environment where past experience has little indication for the future performance, exploration width needs to be high – i.e. all options should be explored regardless the old knowledge built upon past experience – to ensure exploration efficiency and prevent inefficient search misguided by past experience; yet the exploration intensity should be moderate to balance exploration and exploitation. Entangling exploration intensity and width leads to the dilemma of, either inefficient exploration among narrow options or allocation of obsessive resources to exploration, causing the problem of "chasing a moving target".

Second, when an organization focuses more on exploration, it increases two types of costs to generate new knowledge - (i) opportunity cost of exploration and (ii) cost of switching between options. As Posen and Levinthal (2012) pointed out, increased dynamism quickly erodes new knowledge generated through exploration, making exploration less valuable. However, we highlight that increased dynamism also reduces the opportunity cost of exploration – the payoff difference between exploitation and exploration. As old knowledge is eroded by the increased dynamism, exploiting the old knowledge will be less beneficial, and each exploring action will incur less opportunity cost. The conversion of opportunity cost to knowledge is more effective in a dynamic environment than in a stable environment, if exploration width is appropriate. On the other hand, dynamism increases the likelihood of negative experience from exploitation, triggering more passive search – switching action even the organization exploits – and raising switching cost.

Our findings provide theoretical reconciliation and managerial guidance to how organizations should respond in a dynamic environment. If the switching cost is minimal and an organization is able to use separate strategies to disentangle

exploration intensity and width, the organization should do so and focus more on exploration in a dynamic environment. Otherwise, exploration will be ineffective, and the organization should focus more on exploitation in a dynamic environment.

Literature Review

Environment dynamic refers to the unpredictability of environmental changes (Dess & Beard, 1984). It should be differentiated from other aspects of environmental change, such as speed and munificence (Miles, Snow, & Pfeffer, 1974; Siggelkow & Rivkin, 2005): an environmental change can be slow, beneficial, yet unpredictable, imposing administrative challenges for the organization and its decision maker (March & Simon, 1958). Specifically, organizations face two challenges in a dynamic environment. First, the organization's existing knowledge becomes obsolete due to the unpredictable change in environment (Sørensen & Stuart, 2000). Relying on the obsolete knowledge will lead to inferior performance and organizational failures. Secondly, when the environment is dynamic, opportunities emerge in the forms of new customer demands, availability of new technology or the emergence of a new market (Eisenhardt & Tabrizi, 1995; Gavetti, Levinthal, & Rivkin, 2005). Identification of the emerged opportunities leads a better adaptation and organizational success.

Theoretical arguments and formal models suggest that focusing on exploration is an effective organizational response to the obsolete knowledge and emerging opportunities in a dynamic environment. Using a computational model, Kim and Rhee (2009) illustrate that it is important for an organization in a dynamic environment to allow internal variety and encourage its organizational members to experiment with radically new knowledge and update the old knowledge. Davis, Eisenhardt and Bingham (2009) also demonstrate that in an dynamic environment,

organizations should impose less rules and structures driving efficiency, but instead, provide more flexibility and open up the organization to a wide range of opportunities.

These theoretical arguments are further supported by empirical evidences that organizations should focus more on exploration a dynamic environment. Jansen, Van Den Bosch and Volberda (2006) found that pursuing exploratory innovation is more effective in a dynamic environment, leading to better financial performance. Based on the evidence of 279 manufacturing firms in Standard & Poor 500, Uotila and his colleagues (2009) found that exploration effort is more important and profitable when the technological dynamism is high. Garg, Watlers and Priem (2003) found that the CEO's attention to innovation is associated with better performance in a dynamic environment, while the CEO's attention to efficiency is associated with better performance is a stable environment. Larraneta, Zahra & Gonzalez (2014) found that new ventures grow faster from pursuing strategic variety by exploring multiple strategic actions, especially in a dynamic environment. Wang and Li (2008) also found that the negative effect of under-exploration is stronger in a dynamic environment, yet the negative effect of over-exploration is weaker in a dynamic environment. These evidences reaffirm the managerial conventional wisdom: organizations should focus more on exploration in a dynamic environment.

However, such conventional wisdom is not without theoretical challenges from management (Posen & Levinthal, 2012), sociology (Hannan & Freeman, 1984) and economics (Keller & Rady, 1999) literature, questioning the value of exploration in dynamic environment. Particularly, Posen and Levinthal (P&L) use a multi-arm bandit model to illustrate the problem of "chasing a moving target": when the environment is dynamic, organizations are better off avoiding chasing the

moving target by exploration. In a dynamic environment, although exploration facilitates the creation of new knowledge and identification of emerged opportunities, new knowledge and opportunities are quickly eroded, eliminating the expected return of exploration (Posen & Levinthal, 2012). Thus, organization may need to focus less on exploration and more on exploitation in a dynamic environment.

Stieglitz, Knudsen and Becker (2016) highlight the importance to reconcile the converging theoretical arguments and conflicting predictions. Using a multi-arm bandit model similar to P&L, they disentangle three dimensions of environmental dynamism – direction, magnitude and frequency – to provide boundary conditions for the predictions. However, their theoretical model has two shortcomings. First, one of their main predictions contradicts to empirical evidences. Stieglitz and his colleagues propose that increased magnitude of environmental change rewards exploration, and organization should focus more on exploration if the magnitude of change is high. In contrast, Marino and his colleagues (2015) found that when the magnitude of technological regulation change is high, Formula One teams should focus less on exploration as the value of exploration cannot be realized. Such contradiction could be caused by Stieglitz and his colleagues' model specification, that an increase in the magnitude of change will increase the payoff variance across the options, indirectly affect the exploration behavior[22]. Second, the selection process in Stieglitz and his colleagues' model penalizes early explorations. Organizations that explores more face short-term performance consequences in the

[22]In their model, the payoff variance across options increases when magnitude of change increases. Yet, the organizations engage in a payoff-sensitive choice/exploration strategy (softmax). For the same exploration strategy, increased payoff variance across options reduces the likelihood of exploration. To compensate such reduction, organizations need to increase exploration.

early stages and may potentially exit from the population. This selection process creates a bias in the population that favors risk-averse behaviors (Denrell & March, 2001), and potentially contributes to their conclusion that best-performing organizations generally focus more on exploitation and less on exploration in dynamic environments (except high magnitude of change).

In this paper we revisit the "chasing a moving target" problem from a different perspective, by disentangling aspects of exploration. Exploration refers to the pursuit of new knowledge and unknown alternatives (Levinthal & March, 1993) and involves "search, variation, risk-taking, experimentation, play, flexibility, discovery, and innovation" (March, 1991, p. 71). In contrast, exploitation refers to the implementation of old knowledge and known alternatives (Levinthal & March, 1993), and involves "refinement, choice, production, efficiency, selection, implementation and execution" (March, 1991, p. 71). A balance between exploration and exploitation is important organizational success and survival, as exclusively concentrating on exploration or exploitation undermines organizational performance over the long term (Junni, Sarala, Taras, & Tarba, 2013; O'Reilly & Tushman, 2008, 2013; Raisch & Birkinshaw, 2008). One-sided focus on exploitation enhances short-term performance at the cost of long-term performance, as it can cause the firm to fail to discover potential opportunities. On the other hand, one-sided focus on exploration enhances the organization's knowledge at the cost of current performance, and it can trap the organization in endless cycles of instability.

There are two aspects of exploration that have been conceptualized and operationalized. First, as exploration competes organizational resources and managerial attention with exploitation (March, 1991), organizations need to decide

the *exploration intensity*: amount of resources allocated between exploration and exploitation, on decisions such as budget allocation (Coen & Maritan, 2011), managerial focus (Garg et al., 2003), business activities (Uotila et al., 2009), and alliance formation (Lavie & Rosenkopf, 2006; Lin, Yang, & Demirkan, 2007). Second, organizations need to decide on the *exploration width*: how those resources are used, on decisions such as search distance (Levinthal, 1997), search scope (Katila & Ahuja, 2002) and boundary span (Rosenkopf & Nerkar, 2001).

It is unclear whether individuals can behaviorally disentangle the two aspects of explorations. On one hand, an individuals' exploration tendencies, both exploration intensity and width, are constrained by individual preferences (Schoemaker, 1990), past experiences (March, 1996), task incentives (Manso, 2011) and social norms (Hofstede, 1983). On the other hand, Laureiro-Martinez and her colleagues (2015) demonstrate that the attentional control region of human brains (i.e. dorsolateral prefrontal cortex and frontopolar cortex) are responsible for switching between exploration and exploitation, while brain regions invovled with the reward-related uncertainty and attentional control are responsible for exploration activity.

Regardless the behavioral plausibility to disentangle exploration intensity and width at individual level, organizations are capable of effectively disentangling them through deliberate designs. For example, using an organization structure that separates exploration and exploitation activities (Lavie, Stettner, & Tushman, 2010), an organization can strategize its exploration intensity by allocating appropriate budget and manpower to its exploration unit (e.g. R&D unit), while strategize its exploration width by directing appropriate technological search within the exploration unit. Similarly, temporally separating exploration and exploitation

could also disentangle exploration intensity and width: organizations can decide the amount of time they should focus on exploration (and exploitation), and when they explore, they will strategize the search space.

In the following session, we revisit the "chasing a moving target" problem with a multi-arm bandit simulation with specifications similar to P&L.

Simulation Model Specification

The multi-arm bandit model, taking the analogy from multi-arm slot machines, has been extensively used by organization scholars to represent the trade-off between exploration and exploitation (e.g. Aggarwal et al., 2017; Lee & Puranam, 2016; Posen & Levinthal, 2012; Puranam & Swamy, 2016). In this model, the arms of the bandit represent the discrete choices an organization faces such as those between different products, technologies, or investment options, with payoff distributions unknown to the organization ex-ante. In each time period, the agent (organization) makes a choice regarding which arm to invest in, and receives a payoff feedback. This payoff feedback provides information regarding the attractiveness of the choice, which in turn informs future choices. Our model setup builds closely on the models by Posen and Levinthal (2012), Stieglitz, Knudsen and Becker (2016), and Lee and Puranam (2016). Below, we describe the three components of our model: (1) task environment, (2) learning process, and (3) choice strategy.

Task Environment: The *task environment* of a multi-arm bandit model consists of *N* discrete options (i.e. arms). Each period, the organization chooses one arm and receives a payoff $\pi_{n,t} + \epsilon$ in which $\pi_{n,t}$ is the mean payoff of the arm, and ϵ is a normally distributed noise with zero mean and σ^2 variance. The initial mean payoff of the n-th arm, $\pi_{n,1}$, is drawn from a standard normal distribution with zero

mean and unit variance. In such setup, an organization that blindly makes choices will achieve an expected performance of zero.

To model environmental dynamism, we use a stochastic process on $\pi_{n,t}$ with two parameters: frequency of change $\rho \in [0,1]$ and magnitude of change $\mu \in [0,1]$. Specifically, the mean payoff of an arm follows the following process:

$$\begin{cases} \pi_{n,t+1} = (1-\mu)\pi_{n,t} + \vartheta\sqrt{1-(1-\mu)^2}; & \rho \\ \pi_{n,t+1} = \pi_{n,t}; & (1-\rho) \end{cases}$$

in which ρ is the probability with which the mean payoff of an arm is changed from previous period, and ϑ are a normally distributed i.i.d. with zero mean and unit variance, identical to the normally distribution form which the initial mean payoff $\pi_{n,1}$ is drawn. When the magnitude is at its maximum $\mu = 1$, $\pi_{n,t+1} = \vartheta$ if a change happened, that is, the mean payoff of the arm is redrawn from the initial normal distribution. When the magnitude is at its minimum $\mu = 0$, $\pi_{n,t+1} = \pi_{n,t}$ regardless of change. Our manipulation of environment dynamism is slightly different from Stieglitz and his colleagues' model but has two desirable characteristics (see Appendix 1 for proof). First, the mean payoff of the N arms will be normally distributed with zero mean and unit variance, consistently at any point of time. This characteristic ensures any changes in the organizations' choices are caused by the choice strategy rather than the changed payoff distribution across the arms over time. Secondly, frequency and magnitude of change jointly determine dynamism, (i.e. the unpredictability) of the mean payoff of arms over time. Formally, the predictability of the mean payoff of the arm over time can be reflected as the correlation between $\pi_{n,t}$ and $\pi_{n,t+\Delta t}$ is $corr(\pi_{n,t}, \pi_{n,t+\Delta t}) = (1-\rho\mu)^{\Delta t}$.

Learning Process: When an arm is chosen, the organization receives a performance feedback and learns about the attractiveness of the arm. To model the

learning process, we follow the prior literature (Aggarwal, Posen, & Workiewicz, 2017; Lee & Puranam, 2016; March, 1991; Posen & Levinthal, 2012) and use the average learning rule, that is, the organization estimates the attractiveness of an arm by averaging the payoff signals received:

$$q_n = \frac{\sum_{i=1}^{k} p_{i,n}}{k}$$

Here $p_{i,n}$ represents the realized payoff of the i^{th} trial by the organization, and q_n is the estimated payoff (belief) for this arm after it has been tried *k* times.

Choice Strategy: With the beliefs of the arms formed, the organizations now choose which arm to pull. We examine two different choice strategies here: e-greedy and softmax.

The simplest choice strategy is always to be greedy and choose the arm with the highest belief (Auer, Cesa-Bianchi, & Fischer, 2002), that is, a pure exploitation strategy that always exploits the current knowledge for maximized immediate payoff. Epsilon-greedy, or e-greedy, is an alternative that randomly explores amongst all arms with equal chance with a probability ε, and chooses the arm with the highest belief with probability $(1 - \varepsilon)$ (Sutton & Barto, 1998; Tokic & Palm, 2011). ε is the exploration parameter in the e-greedy rule.

A more sophisticated choice strategy is the softmax rule (Posen & Levinthal, 2012) (Knudsen & Srikanth, 2014; Puranam, Stieglitz, Osman, & Pillutla, 2015). The probability of choosing an arm is defined as:

$$Prob(A = n) = \frac{\exp\left(\frac{q_n}{\tau}\right)}{\sum_{i=1}^{N} \exp\left(\frac{q_i}{\tau}\right)}$$

in which $\tau > 0$, also known as temperature, is the exploration parameter and represents the sensitivity of choice probability to the difference in beliefs. When the

τ is very small, the organization is sensitive to value difference – even a small difference in beliefs results in a large difference in the probability of choosing the arms – thus, the agent is highly exploitative. When the τ is large, the organization is insensitive to value difference and thus is highly explorative.

A major difference between e-greedy and softmax is how exploration, particularly exploration intensity and width, is governed by the exploration parameter ε and τ. In the e-greedy rule, exploration parameter ε only governs the exploration intensity: the rate of exploration – choices not exploiting the arm with the best belief – depends on ε. Exploration width on the other hand is always high and independent from ε , and all arms are equally explored whenever the organization explores. In the softmax rule, exploration parameter τ governs both exploration intensity and width. If τ is small, the rate of exploration is low and whenever the organization explores, it is more likely to explore the next-best believed arm than the worst believed arm. If τ is large, the rate of exploration is high and whenever the organization explores, it explores widely: all arms are explored with an (almost) equal likelihood. Posen and Levinthal (2012) argues that this difference is an "important undesirable feature" (p.590) of e-greedy rule, dismissing the e-greedy rule as it is inferior than softmax. However, we will illustrate otherwise in our analysis.

Another difference between e-greedy and softmax is behavioral plausibility. Softmax rule has been found to effectively describe how individuals make choices in trial-and-error learning situations (Camerer & Hua Ho, 1999; Daw, O'doherty, Dayan, Seymour, & Dolan, 2006). Daw and his colleagues (2006) found that individuals make choices in a value sensitive manner, and softmax describes such behaviors better than e-greedy. However, the superior behavioral plausibility of

softmax rule at micro-level may not hold at the macro level for organizations. Organizations could also effectively use organizational culture (Sørensen, 2002), design (Fang, Lee, & Schilling, 2010), employment (Groysberg & Lee, 2009) and incentive schemes (Manso, 2011) to intervene its members' sensitivity to values and exploration behaviors. An organization (e.g. Alphabet) that structurally separates exploration and exploitation activities could embrace cultures and incentives for its exploration unit that drive wide explorations of moonshot projects, while balancing exploration and exploitation by allocating appropriate resources between the units (Lavie et al., 2010). Alternatively, an organization that temporally separates exploration and exploitation activities could embrace cultures and incentives that facilitate wide exploration in the exploration phases, and then transit to exploitation (Lavie et al., 2010). In our formal model, e-greedy rule is a proxy to temporal separation: in the exploration periods, the organization embraces a culture, design, employment or incentive that is "completely insensitive" to values; in the exploitation periods, the organization switches to a culture, design, employment or incentive that is "completely sensitive" to values (Posen & Levinthal, 2012, p. 590). The organization with the e-greedy rule only needs to decide the temporal ratio (i.e. parameter ε) between exploration and exploitation.

Analysis and Results

To make our analysis comparative to the prior literature (Posen & Levinthal, 2012; Stieglitz et al., 2016), we operationalize the model by setting N=10 armed bandit model. The initial belief about each arm is netural and uniformly set to zero. Performance is measured as the average payoff per round over 500 periods. For the baseline analysis, we set the magnitude of change $\mu = 1$ and systematically vary the frequency of ρ. In such setup, when the mean payoff of an arm is changed, it is

a completely redrawn from initial normal distribution, similar to P&L. We then systematically vary exploration parameters τ and ε to identify the optimal exploration strategy that maximizes the performance[23]. For each exploration parameter under each environment dynamism, we report the average performance over 50,000 simulations.

EXPERIMENT 0: Posen and Levinthal 2012 revisited

Figure 1A shows the optimized performance when an appropriate exploration strategy is employed. As expected, for both softmax and e-greedy rules, the optimized performance decreases when the dynamism increases. Surprisingly, the "undesirable" (Posen & Levinthal, 2012, p. 590) e-greedy rule, although performs worse in the stable environments where the frequency of change is below 2%, outperforms the softmax rule when the dynamic environments where the frequency of change is above 2%.

INSERT FIGURE 1A, 1B ABOUT HERE

Figure 1B shows the optimal exploration strategy for both decision rules. As suggested by P&L, optimal exploration parameter under softmax rule decreases when the dynamism increases in dynamic environments (ρ>=0.5%), and organizations should focus less on exploration. However, when choices are guided by e-greedy rule, the optimal exploration parameter increases when the dynamism increases, contradict to P&L's prediction. Figure 1A and Figure 1B suggest: (1) organizations in a dynamic environment should adapt an e-greedy decision rule, and (2) P&L's observation does not hold for the e-greedy rule, that organizations with

[23] τ is set to [0.01, 1.0] with intervals of 0.01, and ε is set to [0.01, 1.0] with intervals of 0.01. Such range is sufficiently large such that the optimal exploration strategy falls within the range, and fine-grain enough that the result is consistent. Robustness checks on the range and fine-graininess does not change the qualitative results.

an e-greedy rule should focus on exploration, rather than exploitation in a dynamic environment. There are several explanations on why P&L's proposed mechanism could not explain the optimal exploration strategy for e-greedy rule when dynamism increases. First, P&L's only focuses on the mechanism of (diminishing) value of exploration, without looking into the cost of exploration. A wholistic view requires investigations on both benefit and cost of exploration. Second, there is a fundamental difference on how exploration is conducted between softmax and e-greedy rule, that the "chasing moving target" is a problem only for the softmax rule, but not for the e-greedy rule (see Appendix 2 for detailed comments about P&L's mechanism).

Experiment 1: Organization Structures

In the baseline model, we compared softmax rule to e-greedy rule. Prior literature has suggested individual decision makers behave in softmax rule, rather than a value-insensitive e-greedy rule. To achieve the approximate to e-greedy, we use a multi-agent organization design.

Organization structure: When an organization has multiple agents, there are different designs to achieve balance between exploration and exploitation. In this paper, we contrast separation versus non-separation of exploration and exploitation, which we model as follows:

Non-separated design: In the non-separated design, the balance between exploration and exploitation is achieved at individual level (i.e. contextual ambidexterity), and thus in this organization as one where each individual in the organization balances her choices between exploration and exploitation activities (see figure 1). We model the organization with J agents such that each of these

agents have the same exploration propensity (or temperature τ). Thus, in an organization with *J* agents, the organization can tune the balance at the individual level and let all agents have the appropriate, intermediate temperature τ_j. This can be achieved in several ways in the real world, such as by setting up an incentive scheme that properly promote exploration and exploitation for all employees or by implementing an HR policy of hiring agents with the desired exploration propensity[24].

Structurally separated design: In the separated design, some agents explore, and other agents exploit. The organization achieves a balance between exploration and exploitation by allocating the appropriate number of agents to either explorative or exploitative roles (see figure 1). We model this by assigning a high value of τ to a fraction x of the *J* agents in the organization, whereas the remaining (1- x) fraction of agents have a very low value of τ. Again, we expect that intermediate values of x will outperform very high or very low values.

In the baseline experiment, there are two types of organizations. In structurally separated organizations, $10x$ agents focus on explorative activities, with a $\tau_{explore} = 1$, and $10(1 - x)$ agents focus on exploitative activities, with a $\tau_{exploit}$=0.002.[25] Therefore, there are 11 possible configurations of the structurally separated organizations ($x = 0.0, 0.1, 0.2, \ldots 0.9, 1.0$), and peak performance (and thus balance) is achieved by tuning the portion of exploration agents$x$.

[24] Organizations in Posen and Levinthal's (2012) model are a special case of non-separated design with one single agent.

[25] We set $\tau_{explore}$=1, which is large enough to represent strong exploration. Robustness tests with different values of $\tau_{explore}$ provide qualitatively similar results. As suggested by Posen and Levinthal (2012), we set $\tau_{exploit}$=0.002 to avoid infinite values in equation (2), yet sufficiently represent a pure exploitation strategy. For example, with two arms with beliefs of 0.30 and 0.29, the probability of choosing the first one is 99.3%, and the latter is less than 0.7%. Tests $\tau_{exploit}$=0.01 and $\tau_{exploit}$=0.005 show our results are robust.

In non-separated organizations, all agents engage in both explorative and exploitative activities. All agents have an identical temperature τ_j (a constraint we drop in robustness). Peak performance is achieved by tuning the value of τ_j from 0.002 to 1.0. By setting the smallest tuning interval at 0.002, there are 500 possible configurations ($\tau_j = 0.002, 0.004 \ldots 0.998, 1.0$).

Figure 2 indicates the accumulated performance for different values of exploration parameter (proportion of exploration agents for structurally separated organizations; temperature τ_j for the non-separated organizations). Both types of organizations can achieve balance: structurally separated organizations' peak performance of 448 is achieved when one agent explores, and the other nine agents exploit. Non-separated organizations' achieve a peak performance of 377 when the temperature τ=0.04. Thus, in this environment, structural separation outperforms non-separation at balance.

INSERT FIGURE 2 ABOUT HERE

Before exploring the contingencies under which this result holds, we attempt to understand the mechanism that underlies this result. First, we compare *accumulated exploration cost* against exploration parameter for these two organizations, as shown in Figure 3A. Exploration, by definition, is choosing an arm that is not optimal based on current belief. Therefore, exploration cost for each agent j at each time period t is the opportunity cost of pulling an arm other than the best-believed arm, formally $\bar{p}_{best\_believed} - \bar{p}_{pulled}$. If an agent pulls the best-believed arm, he/she exploits in this round and the exploration cost is zero. Cumulative exploration cost is the exploration cost summed over all agents across all time periods for a given organization configuration. From figure 3A, we see that exploration cost increases monotonically with increasing exploration parameter in

both designs, as expected. In Figure 3B, we plot the accumulated performance over accumulated exploration cost, and we can observe that structurally separated organization outperforms non-separation at any given exploration cost, and especially when the balance is achieved.[26]

This suggests that the organization with a structurally separated design converts exploration cost to performance more effectively than the organization with a non-separated design. The purpose of exploration is to learn about the task environment and identify potential higher payoff choices for exploitation in the future. Therefore, exploration cost is spent to build *knowledge* that facilitates future exploitation. We measure the organization's knowledge as how close the best-believed arm's mean payoff is to the actual best payoff arm, formally $\bar{p}_{best\_believed} - \bar{p}_{best\_actual}$, ranging from 0 (best knowledge) to -1 (worst knowledge). Our measure of knowledge builds on a similar measure suggested by Lee and Puranam (2016); but unlike their dichotomous measure, we use a continuous measure. Our measure not only considers the match between the best-believed and best-actual arms, but also considers how "wrong" the belief is. We can formally explain performance as function of knowledge and exploration cost in a bandit problem as follows:

$$E(Performance) = E\left(\bar{p}_{pulled} - 0.5\right)$$

$$= E(\bar{p}_{best\_actual} - 0.5 + \left(\bar{p}_{best\_believed} - \bar{p}_{best\_actual}\right) - (\bar{p}_{best\_believed} - \bar{p}_{pulled}))$$

[26] In Figure 2, we cannot make any direct comparison between these two organization designs at any given X-axis coordinate. This is because for the non-separated organization, the X axis denotes the temperature of the agents; in the separated organization, the X axis is the percentage of agents who are exploring. These two cannot be collapsed into a single 'exploration' parameter. Exploration cost provides a uniform handle that we can use to compare these two organization designs, plotted in Figure 3B.

$$= E\left(\bar{p}_{best\_actual}\right) - 0.5 + E(Knowledge) -$$

$$E(Exploration\ Cost) \qquad (3)$$

$E\left(\bar{p}_{best\_actual}\right)$ is a constant, about 0.831 in this set-up, and performance can be explained as spending exploration cost to build knowledge. Table 1 shows the accumulated performance, accumulated exploration cost, and average knowledge for the 'balanced' configuration of the separated and the non-separated organization designs respectively. For these two configurations, we also tabulate the number of exploration and exploitation pulls, and which arms were chosen when the agents explored.

INSERT FIGURES 3A, 3B AND TABLE 1 ABOUT HERE

Table 1 illustrates the mechanism that underlies the superiority of the separated organization design when compared to the non-separated organization design. First, we see that, when balanced, the separated organization has a lower accumulated exploration cost and higher average knowledge when compared to the non-separated organization. Since the organization consists of 10 agents, each of them making a choice 500 times, there are a total of 5000 choices that an organization makes. Out of these, the separated organization exploits (chooses the best believed arm) about 88 percent of the time (4374 of the 5000 choices) and explores (does not choose the best believed arm) about 12 percent of the time (626 choices). In contrast, the non-separated organization exploits only about 74 percent of the time (3678 choices) and explores about 26 percent of the time (1322 choices).

In addition, from Table 1, we see that the separated organization explores more broadly. When the non-separated organization explores, 95 percent of its exploration is concentrated on the second (64 percent), the third (23 percent), and the fourth (8 percent) best believed arms out of the nine arms available; it almost

never chooses the worst believed arm when exploring. In contrast, the separated organization chooses the second (35 percent), third (13 percent) and fourth (9 percent) arms (for a total of 57percent) much less frequently when exploring. Tellingly, the worst-believed arm is chosen over a non-trivial 7 percent of the explorations.

This suggests that non-separated organizations explore more narrowly than the separated organization because of the ***value-sensitive exploration*** by every agent. At balance, each agent in the non-separated organization with moderately low temperature is susceptible to the 'hot stove effect', i.e., a bias against sampling alternatives that initially appeared unattractive (Denrell & March, 2001). Environmental turbulence is a source of such sampling bias, since previously inferior alternatives may become superior in the changed environment. The Bernoulli distributed payoff is another source of such sampling bias, since previous negative experiences with an alternative may not be a good indicator that the alternative is inferior, rather than the agent being initially unlucky. In these cases value-sensitive exploration is inefficient in converting exploration cost into knowledge.

Thus Table 1 provides us with a causal mechanism to explain why the separated organization outperforms the non-separated organization in a turbulent and noisy environment. Non-stationary environments with noisy payoffs require two antithetical behaviors to achieve high performance: (a) more broad-based (value-***in****sensitive*) exploration in order to increase the organization's knowledge regarding the different available alternatives (Denrell and March, 2001), and (b) exploiting that knowledge before the environment changes again (Posen and Levinthal, 2012).

Achieving this balance is tricky because the same organizational mechanisms that influence the choice between exploration and exploitation also influence the breadth of exploration (i.e., how value-sensitive exploration is). In the model, the Softmax exploration parameter (τ) determines the choice between exploration and exploitation as well as the breadth of exploration. When τ is low, the agent exploits more often than explores, and when they do explore, their exploration is narrow (or value-*sensitive*); when τ is high, agent exploration is broad (value *insensitive*), but they also tend to explore rather than exploit.

The structural separation design allows the organization to break this trade-off between the need for *value-**in**sensitive* exploration and *value-sensitive* exploitation. In the separated design, some members with a high temperature explore broadly to generate knowledge about alternatives that other members with a very low temperature exploit to improve organizational performance. This combination thus allows the separated organization to convert exploration cost to knowledge more efficiently than the non-separated organization. The model thus explicitly demonstrates the mechanism underlying the hypothesis suggested by O'Reilly and Tushman (2013) that when faced with a disruptive change, a separated organization is likely to perform better than a non-separated organization.

In the following experiments, we demonstrate that separation of exploration and exploitation buffers against the hot-stove effect and understand its boundary conditions. We demonstrate why value-sensitive exploration of individual agents is inefficient when the environment is turbulent (experiment 2), or the feedback is noisy (experiment 3), or both (experiment 4), and a separated organizational design is needed. Based on this mechanism, we further illustrate how a separated design addresses the problem of "chasing moving target" (experiment 5).

Experiment 2. Effect of Turbulence

In this experiment we more clearly demonstrate the mechanism – *value-(in)sensitive search* – we identified in the previous experiment as critical to performance in turbulent environments. Here, we simplify the baseline model to a 20-period model with one agent per organization (in which structural separation is not possible). The task environment contains only 3 arms, each returns a fixed payoff π_n. When an arm is pulled, the agent instantly updates their belief based on the latest payoff, as each payoff is an accurate representation of the attractiveness of the arm. In this way, learning from the feedback is perfect, and we isolate the sampling problem inherent in a turbulent environment from noisy feedback.

At the beginning, the payoffs of the three arms are 0.2, 0.0 and -0.2. A shock at 11^{th} period reshapes these payoffs. We hypothesize two different shocks as shown in Table 2: (A) less disruptive, i.e. arm 2, the moderate arm (with 0.0 payoff), becomes superior after the shock; (B) more disruptive, i.e. Arm 3, the worst arm (with -0.2 payoff), becomes superior after the shock.

INSERT TABLE 2 ABOUT HERE

At the beginning, the agent has perfect knowledge, i.e. his/her belief of each arm matches the actual payoff. However, he/she does not know whether and when the shock happens and how the shock will reshape the payoffs. The agent needs to have a right temperature or exploration propensity, τ, to balance between exploration and exploitation and maximize her 20-period accumulated performance.

We plot the 20-period accumulated performance with 10,000 iterations for each value τ (between 0 and 0.3, with interval of 0.001) in Figure 4A and 4B (solid lines). With the less disruptive shock, the agent is able to balance exploration and exploitation by having a moderate temperature of 0.121. With the disruptive shock,

however, a balance could not be achieved: performance is maximized when the agent purely exploits (temperature $\tau = 0$).

In other words, value-sensitive exploration (represented by the temperature τ) is ineffective when environment change is disruptive, where past experience is not a good indication for the future. When τ is low, the agent exploits (chooses the best believed arm) often, and is more likely to choose Arm 2 over Arm 3 (narrow exploration). In contrast, when τ is high, the agent is equally likely to choose between the three arms (broad exploration), but does not exploit this superior knowledge.

INSERT FIGURE 4A AND 4B ABOUT HERE

This pathology can be overcome if the decision to explore is divorced from the breadth of exploration (which is what we suggest the separated design accomplishes in the baseline experiment). In Figure 4A and 4B dotted lines, we repeat the above experiment with such a decision rule, which we call the modified softmax. In this rule, the agent first follows a Softmax rule with temperature τ to decide whether the agent explores or exploits. If the agent exploits, he/she pulls the best-believed arm. If the agent decides to explore, in this rule, his/her decision is completely value-insensitive, and the agent pulls any of the non-best-believed arms with equal probability. This modified-softmax decision rule underperforms the value-sensitive rule (Softmax) when the shock is less disruptive (Figure 4A, dotted line), but it outperforms the value-sensitive rule to strike a balanced performance with temperature $\tau = 0.12$ when the shock is disruptive (Figure 4B, dotted line).

Daw and colleagues (2006) used neurological evidence to suggest that such a decision rule is implausible behaviorally for individuals. However, when the organization has multiple agents, such a decision rule could be achieved at the

organization level by structurally separating exploration and exploitation: by a combination of selection and treatment, the organization could choose exploitation members that are completely value sensitive, and the exploration members that are value insensitive.

In addition, given the mechanism illustrated in Figure 4A and 4B, we would expect a separated design outperforms non-separated design when the environment is disruptive and the past experience is a bad indicator of current landscape; a non-separated design outperforms separated design when environment is less disruptive and the past experience is still a good indicator of current landscape.

Next, we repeat Experiment 1 by varying environment turbulence level and removing payoff noise (each pull returns a fixed payoff instead of a Bernoulli payoff) to isolate the mechanism, and compare the advantage (or disadvantage) of structural separation over non-separation across turbulence level[27]. Results, shown in Figure 5 suggest that structural separation underperforms when the turbulence level is low, and outperforms when the turbulence level is high. When the turbulence level is very low, minimal exploration is needed, thus the difference between separated and non-separated designs are minimal. When the turbulence level increases, the advantage of non-separation mode increases and then decreases; until above a threshold turbulence level, after which the separation mode outperforms.

INSERT FIGURE 5 ABOUT HERE

Experiment 3. Effect of Noisy Feedback

In the previous experiment, we illustrate why structural separation outperforms non-separation in a turbulent environment. Strong turbulence makes

[27] Advantage of the structurally separated organization over the non-separated organization is measured as the difference between balanced performance of structurally separated organization and non-separated organization, divided by the balanced performance of the non-separated organization.

the past experience irrelevant to the current landscape, thus a value-sensitive exploration based on past experience is inefficient. High levels of payoff noise can also make past experience a less reliable guide to future payoffs, leading to the hot-stove effect (Denrell & March, 2001). To understand the effect of payoff noise on the two organization designs, again, we simplify the baseline model to a 20-period model with one agent per organization. The task environment contains 5 arms, each returns a Bernoulli payoff, with a p_i success chance returning value 0.5 and $1 - p_i$ failure chance returning value -0.5, as shown in Table 3.

INSERT TABLE 3 ABOUT HERE

At the beginning, the agent has unbiased beliefs of zero payoff for each arm. When the agent pulls an arm, he/she will experience either a success (value of 0.5) or a failure (value of -0.5), and he/she updates the belief using an average learning rule. To illustrate the mechanism, we assume the agent pulls Arm 1 – the best payoff arm – at the first period; now we have two possibilities: (A) the agent is lucky with the first pull and experiences a success; (B) the agent is unlucky with the first pull and experiences a failure.

We compare the performance of the value-sensitive exploration (Softmax decision rule) with value-insensitive exploration (modified-Softmax decision rule) as stated in the previous experiment. Twenty-period accumulated performance with 10,000 iterations for each value τ (between 0 and 0.5, with interval of 0.001) are plotted in Figure 6A and 6B (solid lines for Softmax, dotted lines for the modified Softmax). When the agent is lucky the arm is coded superior. Therefore, her best strategy is to always exploit under both decision rules (recall there is no turbulence) as shown in Figure 6A.

However, when unlucky, the agent forms a negative bias against the best arm, reducing the future sampling probability of the best arm, and leading to the hot stove effect (Denrell & March, 2001; March, 1996). To counter this bias, the agent needs to explore in an unbiased way, that is, value-insensitive. As shown in Figure 6B, the modified-Softmax decision rule outperforms the value-sensitive decision rule at peak performance. The performance of the value-sensitive decision rule is peaked at $\tau = 0.004$, at which value the agent pulls the best arm 0.45 times on average from period 2 to period 20. The performance of the modified Softmax decision rule is peaked at $\tau = 0.097$, at which value the agent pulls the best arm 1.59 times on average from period 2 to period 20. Therefore, separating exploration and exploitation – is a more effective way to reduce the hot stove effect.

INSERT FIGURE 6A AND 6B ABOUT HERE

These results suggest that when the payoff noise is low, that is, when past experiences are good indicators of the actual mean payoff of each arm, value-sensitive decision rule of non-separation organization design outperforms the structurally separated design; when the payoff noise is high, that is, past experiences are not good indicators of the actual mean payoff of each arm subject to luck, value-insensitive exploration by the separated organization design outperforms the non-separated design.

To affirm this speculation, we repeat the baseline Experiment 1 by fixing environment turbulence level at zero, but with Gaussian distributed payoff returns (instead of Bernoulli).[28] We vary the payoff noise by varying the standard deviation of the Gaussian distribution, and compare the advantage (or disadvantage) of

[28] To align with the baseline model, the mean payoff of an arm is drawn from Beta(2, 2) distribution, minus 0.5.

structural separation over non-separation across different payoff noises, as shown in Figure 7. When the payoff noise is zero, minimal exploration is needed, thus the difference between separated and non-separated designs are minimal; when the payoff noise increases, the disadvantage of structural separation increases and then decreases; when the standard deviation is above 0.8, the separation design increasingly outperforms the non-separation design.

INSERT FIGURE 7 ABOUT HERE

Experiment 4: Combining Environment Turbulence and Payoff Noise

In this experiment, we combine the mechanisms from the experiment 2 and 3, namely bounded rationality caused by turbulence and noise. The initial mean payoff of each arm $\bar{p}_{n,0}$ is drawn from a unit Gaussian distribution (mean 0 and standard deviation 1). At each period t, if a shock happens (with turbulence probability η), each arm's mean payoff $\bar{p}_{n,t}$ has a 50% chance to be redrawn from the unit Gaussian distribution; otherwise, $\bar{p}_{n,t} = \bar{p}_{n,t-1}$. When an arm is drawn at time t, the realized payoff is $\bar{p}_{n,t} + \epsilon$ in which ϵ is a Gaussian distributed payoff noise with mean zero and standard deviation σ. We systematically vary both the standard deviation of payoff noise σ and the environment turbulence level η. We set 50 agents per organization and run the simulation for 200 periods. By identifying the right number of exploration agents (separated design) or the right temperature for all agents (non-separated design), the advantage of the separated over the non-separated design at balance is shown in Figure 8.

INSERT FIGURE 8 ABOUT HERE

Structural separation outperforms non-separation when the environment is turbulent and payoff is noisy (top-right area in Figure 8), as expected from our earlier discussion. In contrast, the non-separated design outperforms the separated

design when the environment is stable, and payoff is less noisy (bottom-left area in Figure 8). The advantage of non-separated design (or the disadvantage of separated design) is highest when both environmental turbulence and payoff noise are moderate.

When both payoff noise and turbulence are low, the agents are almost perfectly rational and minimal exploration is needed, thus there is little difference between the separated and non-separated designs. When payoff noise and turbulence increase, rationality becomes more bounded and exploration is needed to realign the experience-based beliefs and the reality. However, when noise and turbulence are moderate, past experiences are still good indicators of reality, thus *value-sensitive exploration* of the non-separated design prevails. When noise and turbulence are high, past experiences are no longer good indicators of reality, thus *value-insensitive exploration* of separated design prevails. This finding formalizes the intuition behind O'Reilly and Tushman's (2013) proposition that structurally separating exploration and exploitation is necessary when the organization faces disruptive change, and extends it to include cases of high feedback noise.

Discussion

Though it is well accepted that superior performance entails achieving a balance between exploration and exploitation, questions remain about how exactly organizations should go about achieving such a balance. Exploitation is defined as the choice to perform actions that are currently known to have superior payoffs; formally, exploitation is modelled as the agent choosing the action that is currently believed to return the *highest* payoff. Exploration is defined as the choice to perform an action that is currently believed to be inferior, in order to learn more about its

true properties (March, 1996); formally, exploration is modelled as the agent choosing an alternative that is currently believed to return a lower payoff than the 'exploitation' option. Unlike exploitation (which is value-*sensitive* by definition), exploration can be value *sensitive* or value ***in****sensitive* – i.e., when the agent explores, how likely is it to choose an option that is currently believed to be almost as good the current exploitation choice (narrow or value-*sensitive* exploration) versus a choice that is currently believed to be truly awful (broad or value-*insensitive* exploration).

When the environment is turbulent and/or payoff is noisy, choices that were previously poorly performing may now perform better; thus knowledge based on past experience may not be a good indicator of the reality. Superior organizational performance under these conditions requires both (1) broad (value-***in****sensitive*) exploration to gain new knowledge, and (2) swift (value-*sensitive*) exploitation to use knowledge before the environment changes again. These two activities are difficult to reconcile since they build on different assumptions about knowledge: the former assumes that knowledge built from past experience does not reflect reality and has little value, whereas the latter assumes that knowledge built from past experience accurately reflects reality and therefore valuable.

Individuals are constrained in meeting these competing requirements: their behaviour does not shift seamlessly between being completely value-insensitive (as required by the first assumption) to completely value-sensitive (as required by the second assumption) and back (Daw et al, 2006). It is for this reason that current formal work does not distinguish between the decision to explore and the breadth of exploration as two distinct decision variables. For example, in the Softmax algorithm (Luce, 1959; Camerer and Hua Ho, 1999), that is frequently used to

model agent choice behaviour, a single parameter, the temperature τ, regulates both the propensity to explore (vs. exploit) and the breadth of exploration. When τ is low, the agent explores less frequently, and its exploration is value-sensitive. When τ is high, the agent explores often, and its exploration is value-insensitive.

Behavioural experiments suggests that the softmax rule better predicts actual human behaviour when compared to the value-insensitive exploration rule (Daw et al, 2006; also see Cohen, McClure and Yu, 2007; Gans, Knox and Croson, 2007; Knox et al, 2012; Weber, Shafir and Blais, 2004; please also see discussion in Posen and Levinthal, 2012, p590 and Puranam et al, 2015). This is likely true of individuals in organizations as well, as social structures such as incentives, authority, culture, and socialization, encourage either (broad) exploration or exploitation, but not both. This propensity for individuals in organizations to either explore or exploit is well recognized and forms the basic behavioural assumption underlying the literature on ambidexterity. Thus, the behaviour assumed by the value-insensitive exploration rule (e.g. ε-greedy rule, Sutton and Barto, 1998), when agents are completely value sensitive, except when they explore with a small probability ε when they are completely value-insensitive, is unlikely to describe the behaviour of human agents in organizations (please see discussion in Adler et al, 2009; Benner and Tushman, 2003; 2015; Shalley and Gilson, 2017 on the broader topic of employees simultaneously exploring and exploiting).

What is often not as well recognized, especially in formal work, is that organizations are collectives of human agents, and therefore, the organization has an "*extra degree of freedom*" when compared to individuals. This can be utilized to achieve a design such that the organization as a whole, rather than its individual parts, balances exploration and exploitation activities. In this case, one unit within

the organization is highly value-sensitive and exploits current knowledge, whereas another unit is value-insensitive, explores broadly and accrues knowledge. Whereas the separated organization effectively utilizes this additional degree of freedom, the non-separated organization does not. Therefore, the separated organization performs better in situations that require exploitation to be coupled with broad (value-*insensitive*) exploration, and the non-separated organization performs better in situations that require exploitation to be coupled with narrow (value-*sensitive*) exploration. In this sense, adopting a richer multi-agent modelling framework allows us to ask questions and arrive at different predictions from typical unitary-agent models, thus contributing to forward formal work on the exploration-exploitation trade-off.

Our results thus formally replicate the intuition presented by O'Reilly and Tushman (2013), who argue that separation mode for achieving ambidexterity is likely superior to the contextual mode when environments are changing radically (also see Benner and Tushman, 2015). We add to prior work by formalizing the mechanism – we show that the ambidexterity problem comes in two distinct flavours: (a) the tension between pursuing value-***in****sensitive* exploration along with exploitation and (b) the tension of pursuing value-*sensitive* exploration along with exploitation (which is always value-sensitive). We argue that the first problem is more difficult to solve than the second problem, and show that different organization designs are optimal under these conditions. The non-separation design is ineffective in the first instance because it is more prone to the hot-stove effect (Denrell & March, 2001) in noisy environments and to the 'chasing a moving target' problem in turbulent environments (Posen & Levinthal, 2012). In contrast, the separation design is effective, but inefficient in the second instance.

In addition, our study proposes the following additional insights. First, we show that the same tension inherent in turbulent environments is also inherent in noisy environments, where prior outcomes are not indicative of the true potential of the arms. For example, consider the pharmaceutical industry, where the success or failure of drug programs (such as statins or beta-blockers) cannot be reliably identified from the success of individual drugs (such as Lipitor or Lopressor). We also show that the effects of noise and turbulence are super-additive, and thus broaden the context for adopting structural ambidexterity from what was known in prior work.

Second, we identify scenarios where contextual ambidexterity is inherently superior, another point of contention in prior work. We show that when turbulence or noise are moderate, the separated organization under-performs the non-separated organization. We note that in highly stable environments, there is almost no difference between separation-mode versus non-separation mode, which is different from intuition. More interestingly, we show that the non-separation mode increasingly outperforms the separation mode when turbulence and noise increase, until some threshold after which the effect reverses. Unlike prior work that does not speculate about whether one these modes of ambidexterity is superior, or whether they are simply equi-final under less turbulent conditions, we show that the non-separated organization is actually superior. We also show that this superiority comes not from coordination problems inherent in structural separation (they are absent by design in our model), but because value-*sensitive* exploration is a superior strategy for adaptation in these environments. Indeed, the superiority of contextual mode is likely to be broader than suggested here, to the extent that coordination problems in the separation mode are pervasive and difficult to solve.

Third, we are more precise about the effects of 'change'. Change can be radical – i.e., arms that were previously believed to be truly inferior (rather than slightly inferior) can become superior, or change can be turbulent – i.e., the nature of the arms fluctuate, without significant reordering of the belief structure, or both. Our results suggest that separation mode is only superior in steady state when there is radical and turbulent change; in fact, non-separation mode is superior when change is turbulent, but not radical. This is because value-***in****sensitive* exploration is only useful when there is radical change. Note that in many of our experiments, similar to Posen and Levinthal (2012), the higher the turbulence, the more likely it is coupled with radical change. In experiment 2 and 4, we specifically decouple radical change from turbulence, and show that separation is only truly superior when there is radical change (in a computational model, the effects of radical change with no turbulence are washed away at steady state, and can only be witnessed in a time-limited model as in experiments 2 and 3).

Finally, our predictions for the 'chasing a moving target' problem differ from those by Posen and Levinthal (2012). They predict that when environments are highly turbulent, the better strategy is to exploit, since the knowledge gained from exploration is fleeting. They argue that as turbulence increases, optimal exploration strategy follows an 'inverted-U' shaped pattern, which we reproduce in the non-separation design. However, their unitary actor model suffers from the lack of degrees of freedom as argued before. Correcting for this, we show that it is better for the organization to devote more (not less) resources to exploring as the environment changes rapidly, where value insensitive exploration is paired with exploitation in the separation design. Thus, the pattern of optimal resource

allocation to exploration activities with increasing turbulence is different across these designs.

Limitations and Conclusions

The study is subject to a few important limitations. First, our modelling approach builds on existing behavioural research, especially on experimental findings regarding how individuals explore versus exploit, as opposed to calibrating to observed real-world data and offering a potential mechanism to explain them. Each modeling approach has its strengths and weaknesses, and we believe that the core research question explored here matches the modelling approach that we have chosen, but other modeling approaches to this question may produce different insights. Our approach is aligned with much prior work in organization theory on balancing exploration and exploitation (Bauman et al, 2018; Denrell and March, 2001; Lee and Puranam, 2016; Posen and Levinthal, 2012). Second, we abstract away from how individual beliefs are formed in order to focus on the research question of interest here, which is the influence of the different organization designs on ambidexterity, though we relax some of our assumptions in robustness. Future work should consider a more rigorous treatment of belief formation with the different designs and using various integration mechanisms.

Despite these limitations, our study contributes to the literature in several ways in teasing out the contingencies under which different modes of achieving ambidexterity are superior. One other strength of this study is that we examine differences in performance when these different modes of ambidexterity are at "balance", allowing us to side-step the thorny issue of tautology in this literature, where balance is defined as the state at which maximum performance is reached (Raisch et al, 2009). By using a formal approach we also throw light on the

underlying mechanisms, which may not always be feasible in empirical studies, especially with a construct such as ambidexterity which has been applied at multiple levels of analysis and with widely different empirical operationalizations (cf: Junni et al, 2013).

Our formal results are aligned with broad empirical patterns. For example, firms with a more decentralized R&D structure, where the same units are more likely to be tasked with future products as well as current improvements, generate fewer patents and these patents have lower technical impact (cf: Argyres & Silverman, 2004; Arora, Belenzon, & Rios, 2011, 2014). In contrast, firms with a more centralized R&D structure, generate innovation that draws from wider range of technologies, and has larger subsequent impact (Argyres & Silverman, 2004). Integration barriers, based on product scope (Arora et al., 2014), geography (Singh, 2008), lack of social capital (Monteiro & Birkinshaw, 2017), or clash of incentives (Bercovitz & Feldman, 2007), however, remain an important liability of pursuing ambidexterity using the separated mode.

The findings derived from our theoretical model have interesting managerial implications. Consider the example of Alphabet. The industry environment it operates in is highly dynamic today. In addition, Alphabet now has projects (e.g. Waymo, Project Loon) that are risky and uncertain. Yet it also operates in businesses (e.g. search engine and advertisement) that generate stable revenues. Conducting explorative projects in dedicated subsidiaries and separating them from revenue generating activities buffer the bias against those risky and uncertain projects that may fail initially. Thus, the company adopted the separation mode, moving away from the non-separated approach by dropping the 80-20 policy (Schrage, 2013; Zenger, 2015). In contrast, for a firm such as 3M that operates in

the more stable technology regime employing chemistry, a non-separated design, where employees pursue their favourite projects some percent of their time appears to be optimal.

This also suggests that perhaps some of the solutions applied to the problem of central R&D are misguided. The point of separating exploration into a central R&D unit is to allow it to explore truly novel alternatives, especially those considered of no value currently. Forcing incentive schemes on central R&D to be responsive to divisions and their current commercial needs, in effect, neuters the very purpose of central R&D. No doubt coordination problems exist between central R&D and the divisions. As Tushman and colleagues have argued (Smith & Tushman, 2005; Tushman & O'Reilly, 1996), it is the job of senior management to achieve an overall collaborative structure across these specialized units. Side-stepping this responsibility and empowering divisions to direct how resources are allocated in the central R&D is likely to not be the appropriate solution, since it pushes the specialized exploration unit to become more value-*sensitive* in its exploration, which in turn guarantees its failure when radical change happens.

List of Figures for Essay B

Figure 1. Achieving optimal performance with Softmax and E-greedy.

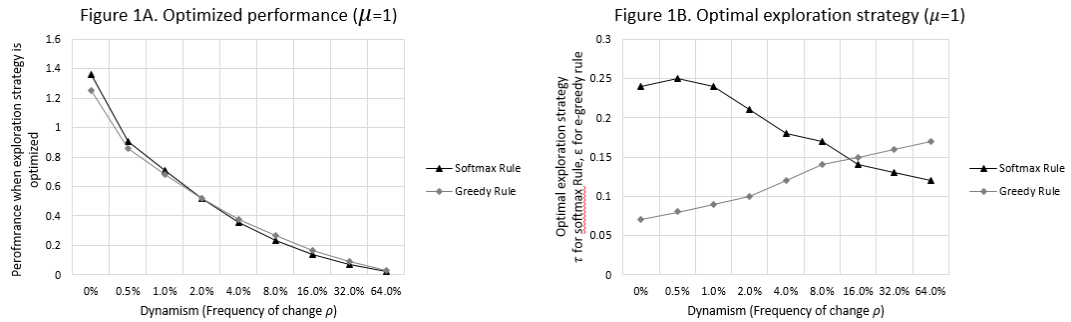


Figure 2. Accumulated performance over exploration parameter (experiment 1)

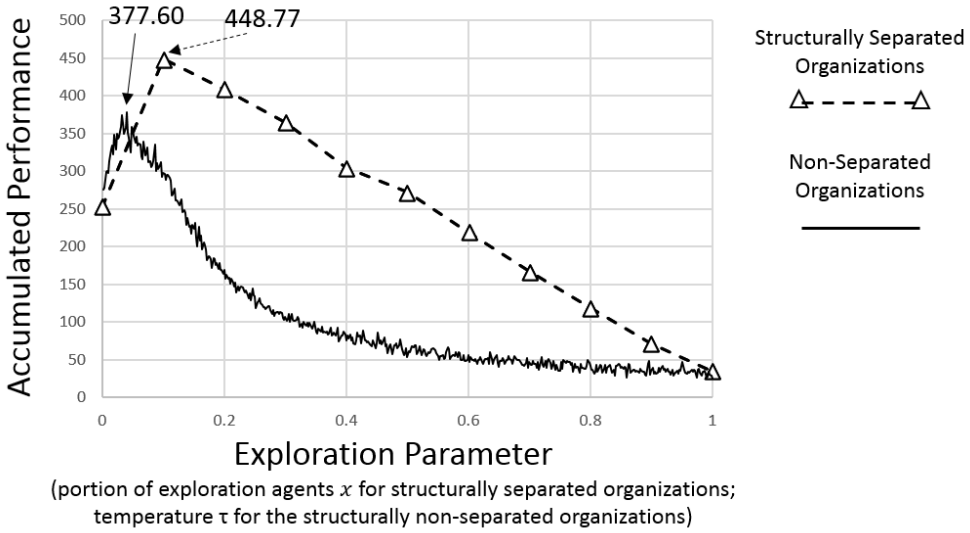


Figure 3A. Accumulated exploration cost over exploration parameter (experiment 1)

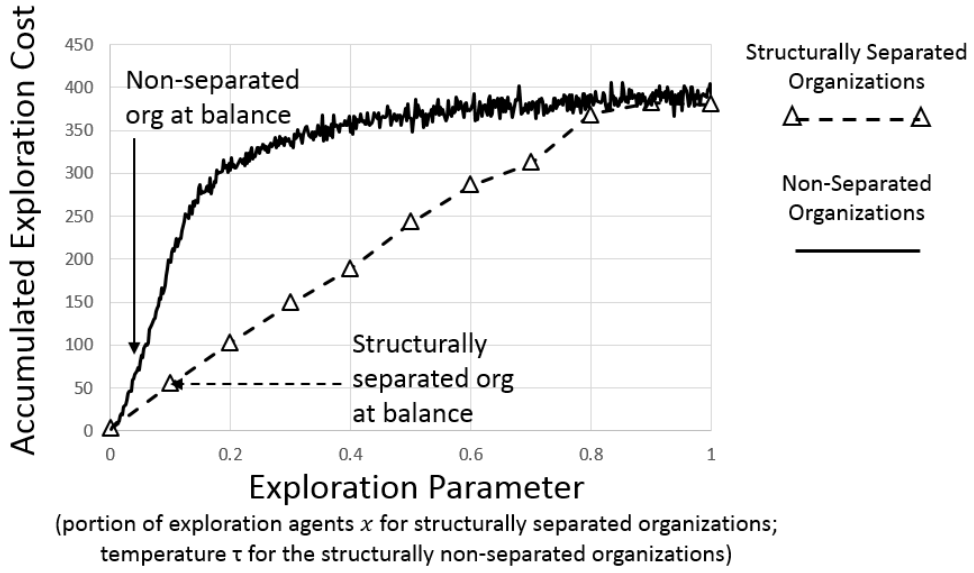


Figure 3B. Accumulated performance over accumulated exploration cost (experiment 1)

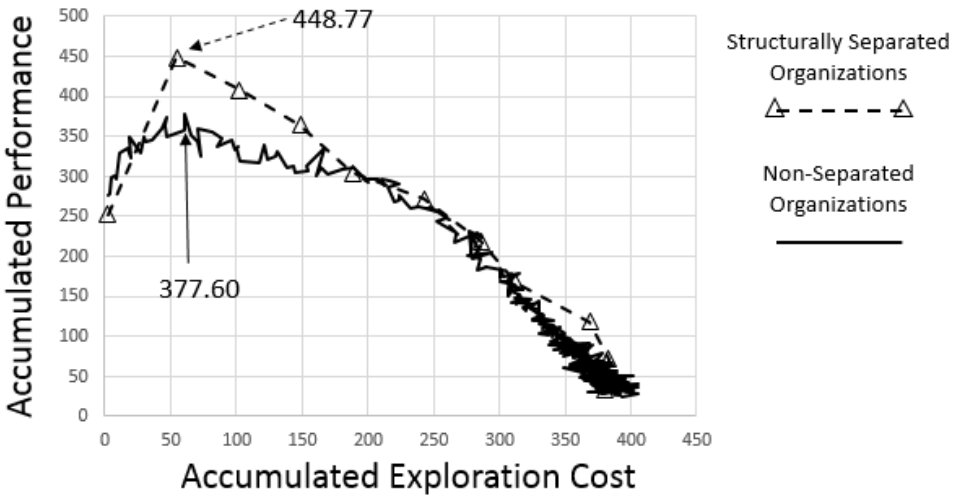


Figure 4. Accumulated performance over temperature, disruptive and less disruptive shocks (experiment 2)

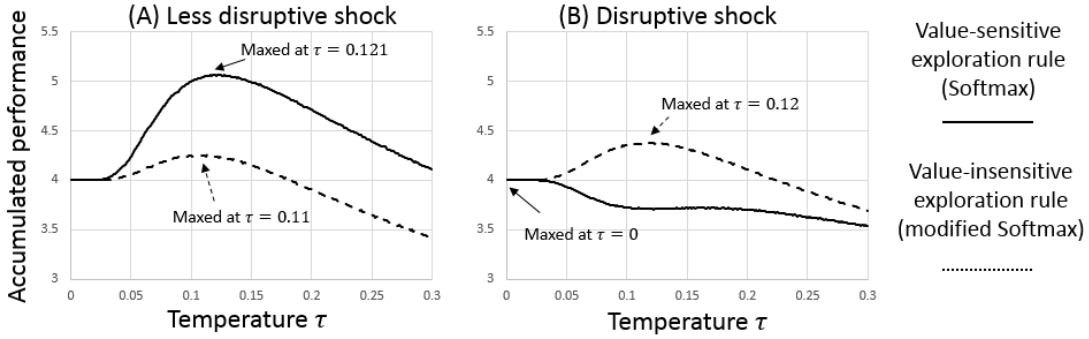


Figure 5. Advantage / disadvantage of structural separation over turbulence level (experiment 2)

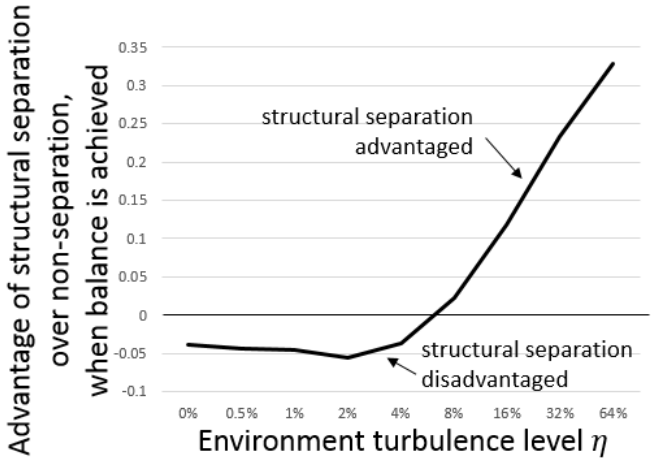


Figure 6. Accumulated performance over temperature, lucky or unlucky with the first pull on Arm 1(experiment 3)

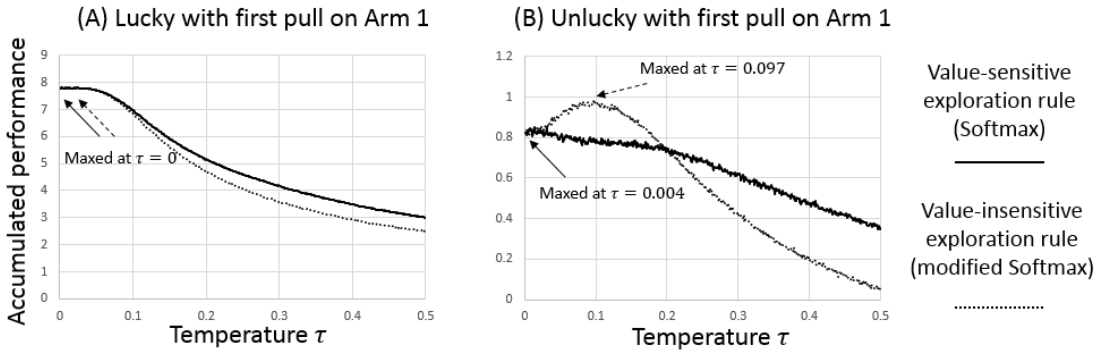


Figure 7. Advantage / disadvantage of structural separation over payoff noise (experiment 3)

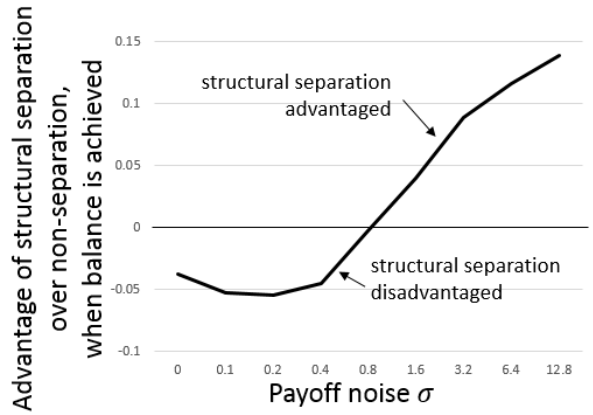
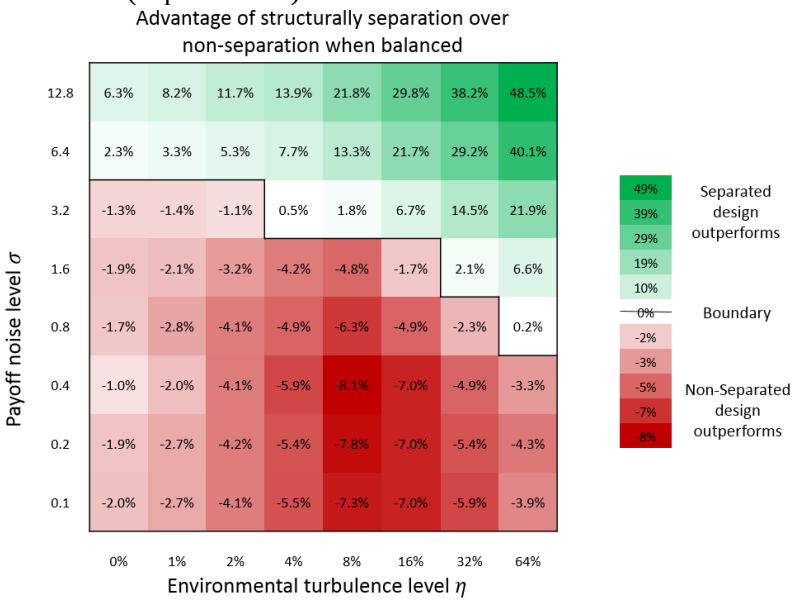


Figure 8. Advantage of structural separation across environmental turbulence and payoff noise level (experiment 4)



List of Tables for Essay B

Table 1: Structurally separated and non-separated organizations at balance (experiment 1)

	Accumulated Performance	Accumulated Exploration Cost	Average Knowledge	No. of exploitation pulls (i.e. on best-believed arm)	No. of exploration pulls (i.e. on non-best believed arms)				
					Total	On 2nd best-believed arm	On 3rd best-believed arm	On 4th best-believed arm	On worst-believed arm
Structurally separated organization (at balance, i.e. 1 exploration agent)	448.77	54.67	-0.2306	4373.97	626.03	220.09 (**35**% of exploration pulls)	79.89 (13% of exploration pulls)	58.94 (9% of exploration pulls)	45.04 (7% of exploration pulls)
Non-separated organization (at balance, i.e. τ=0.040)	377.60	60.64	-0.2422	3677.86	1322.14	848.15 (**64**% of exploration pulls)	315.26 (23% of exploration pulls)	113.56 (8% of exploration pulls)	0.04 (0% of exploration pulls)

Table 2: Payoffs of arms in disruptive and less disruptive shocks (experiment 2)

Less disruptive shock	Period 1 to 10	Period 11 to 20
Arm1 payoff π_1	0.2	0.2
Arm2 payoff π_2	0.0	0.5
Arm3 payoff π_3	-0.2	-0.2

Disruptive shock	Period 1 to 10	Period 11 to 20
Arm1 payoff π_1	0.2	0.2
Arm2 payoff π_2	0.0	0.0
Arm3 payoff π_3	-0.2	0.5

Table 3: success chance of the arms (experiment 3)

Arm1 success chance p_1	0.9
Arm2 success chance p_2	0.7
Arm3 success chance p_3	0.5
Arm4 success chance p_4	0.3
Arm5 success chance p_5	0.1

References for Essay B

Aggarwal, V. A., Posen, H. E., & Workiewicz, M. (2017). Adaptive capacity to technological change: A microfoundational approach. *Strategic Management Journal, 38*(6), 1212-1231.

Argyres, N. S., & Silverman, B. S. (2004). R&D, organization structure, and the development of corporate technological knowledge. *Strategic Management Journal, 25*(8-9), 929-958.

Arora, A., Belenzon, S., & Rios, L. A. (2011). The organization of R&D in American corporations: the determinants and consequences of decentralization: National Bureau of Economic Research.

Arora, A., Belenzon, S., & Rios, L. A. (2014). Make, buy, organize: The interplay between research, external knowledge, and firm structure. *Strategic Management Journal, 35*(3), 317-337.

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

Bercovitz, J. E., & Feldman, M. P. (2007). Fishing upstream: Firm innovation strategy and university research alliances. *Research Policy, 36*(7), 930-948.

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

Coen, C. A., & Maritan, C. A. (2011). Investing in capabilities: The dynamics of resource allocation. *Organization Science, 22*(1), 99-117.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative science quarterly*, 128-152.

Danneels, E., & Sethi, R. (2011). New product exploration under environmental turbulence. *Organization science, 22*(4), 1026-1039.

Davis, J. P., Eisenhardt, K. M., & Bingham, C. B. (2009). Optimal structure, market dynamism, and the strategy of simple rules. *Administrative science quarterly, 54*(3), 413-452.

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.

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

Dess, G. G., & Beard, D. W. (1984). Dimensions of organizational task environments. *Administrative science quarterly*, 52-73.

Eisenhardt, K. M., & Tabrizi, B. N. (1995). Accelerating adaptive processes: Product innovation in the global computer industry. *Administrative science quarterly*, 84-110.

Fang, C., Lee, J., & Schilling, M. A. (2010). Balancing exploration and exploitation through structural design: The isolation of subgroups and organizational learning. *Organization Science, 21*(3), 625-642.

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.

Gavetti, G., Levinthal, D. A., & Rivkin, J. W. (2005). Strategy making in novel and complex worlds: The power of analogy. *Strategic Management Journal, 26*(8), 691-712.

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

Groysberg, B., & Lee, L.-E. (2009). Hiring stars and their colleagues: Exploration and exploitation in professional service firms. *Organization Science, 20*(4), 740-758.

Hannan, M. T., & Freeman, J. (1984). Structural inertia and organizational change. *American sociological review*, 149-164.

Hofstede, G. (1983). The cultural relativity of organizational practices and theories. *Journal of international business studies, 14*(2), 75-89.

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

Junni, P., Sarala, R. M., Taras, V., & Tarba, S. Y. (2013). Organizational ambidexterity and performance: A meta-analysis. *The Academy of Management Perspectives, 27*(4), 299-312.

Katila, R., & Ahuja, G. (2002). Something old, something new: A longitudinal study of search behavior and new product introduction. *Academy of Management Journal, 45*(6), 1183-1194.

Keller, G., & Rady, S. (1999). Optimal experimentation in a changing environment. *The 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.

Knudsen, T., & Srikanth, K. (2014). Coordinated exploration: Organizing joint search by multiple specialists to overcome mutual confusion and joint myopia. *Administrative science quarterly, 59*(3), 409-441.

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

Laureiro-Martínez, D., Brusoni, S., Canessa, N., & Zollo, M. (2015). Understanding the exploration–exploitation dilemma: An fMRI study of attention control and decision-making performance. *Strategic Management Journal, 36*(3), 319-338.

Lavie, D., & Rosenkopf, L. (2006). Balancing exploration and exploitation in alliance formation. *Academy of Management Journal, 49*(4), 797-818.

Lavie, D., Stettner, U., & Tushman, M. L. (2010). Exploration and exploitation within and across organizations. *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.

Levinthal, D. A. (1997). Adaptation on rugged landscapes. *Management science, 43*(7), 934-950.

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

Lin, Z., Yang, H., & Demirkan, I. (2007). The performance consequences of ambidexterity in strategic alliance formations: Empirical investigation and computational theorizing. *Management science, 53*(10), 1645-1658.

Manso, G. (2011). Motivating innovation. *The Journal of Finance, 66*(5), 1823-1860.

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

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

March, J. G., & Simon, H. A. (1958). Organizations.

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.

Miles, R. E., Snow, C. C., & Pfeffer, J. (1974). Organization-environment: concepts and issues. *Industrial Relations: A Journal of Economy and Society, 13*(3), 244-264.

Monteiro, F., & Birkinshaw, J. (2017). The external knowledge sourcing process in multinational corporations. *Strategic Management Journal, 38*(2), 342-362.

O'Reilly, C. A., & Tushman, M. L. (2008). Ambidexterity as a dynamic capability: Resolving the innovator's dilemma. *Research in organizational behavior, 28*, 185-206.

O'Reilly, C. A., & Tushman, M. L. (2013). Organizational ambidexterity: Past, present, and future. *The 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.

Puranam, P., Stieglitz, N., Osman, M., & Pillutla, M. M. (2015). Modelling bounded rationality in organizations: Progress and prospects. *Academy of Management Annals, 9*(1), 337-392.

Raisch, S., & Birkinshaw, J. (2008). Organizational ambidexterity: Antecedents, outcomes, and moderators. *Journal of Management, 34*(3), 375-409.

Rosenkopf, L., & Nerkar, A. (2001). Beyond local search: boundary-spanning, exploration, and impact in the optical disk industry. *Strategic Management Journal, 22*(4), 287-306.

Schoemaker, P. J. (1990). Are risk-attitudes related across domains and response modes? *Management science, 36*(12), 1451-1463.

Schrage, M. (2013). Just how valuable is Google's '20% time'. *Harvard Business Review, 20*.

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

Smith, W. K., & Tushman, M. L. (2005). Managing strategic contradictions: A top management model for managing innovation streams. *Organization Science, 16*(5), 522-536.

Sørensen, J. B. (2002). The strength of corporate culture and the reliability of firm performance. *Administrative science quarterly, 47*(1), 70-91.

Sørensen, J. B., & Stuart, T. E. (2000). Aging, obsolescence, and organizational innovation. *Administrative science quarterly, 45*(1), 81-112.

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). *Introduction to reinforcement learning* (Vol. 135): MIT press Cambridge.

Tokic, M., & Palm, G. (2011). *Value-difference based exploration: adaptive control between epsilon-greedy and softmax.* Paper presented at the Annual Conference on Artificial Intelligence.

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

Tushman, M. L., & Romanelli, E. (1985). Organizational evolution: A metamorphosis model of convergence and reorientation. *Research in organizational behavior*.

Uotila, J., Maula, M., Keil, T., & Zahra, S. A. (2009). Exploration, exploitation, and financial performance: analysis of S&P 500 corporations. *Strategic Management Journal, 30*(2), 221-231.

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.

Zahra, S. A. (1996). Goverance, ownership, and corporate entrepreneurship: The moderating impact of industry technological opportunities. *Academy of management journal, 39*(6), 1713-1735.

Zenger, T. (2015). Why Google became Alphabet. *Harvard Business Review, Harvard Business Review, August, 15*, 1-4.

ESSAY C: Responding to two levels of feedback: evidence from professional basketball teams

Introduction

Decades of research has shown how organizations respond to performance feedback. Drawing insights from the behavioral theory of the firm (BTOF), empirical evidences have been documented about a variety of organizational responses to performance feedback: organizational change (Greve, 2003; Park, 2007; Chen and Miller 2007), risk taking (Bromiley, 1991; Miller and Leiblein, 1996; Kim and Rhee, 2017) and new product introduction (Gaba and Joseph, 2013). However, there is little known about how organizations respond to overall organizational performance feedback and performance feedback of different strategic options.

The importance of this research question comes in three folds. *First*, organizational response to the overall performance can be fine-grained into responses to feedback about different strategic options. Posen and his colleagues (2018) call for a disentanglement of how alternatives are evaluated to understand the black box of organizational responses to performance feedback. Spontaneously analyzing organizational response to the overall performance and feedback from strategic options could uncover the underlying mechanisms. *Second*, organizations often pursue multiple strategic options at the same time, e.g. diversification, strategic alliances, or internationalization, in the form of online search. Therefore, organizations need to respond to a lower level of performance feedback from a strategic option and a higher level of performance feedback of the organization. *Third*, the line of literature on performance response implicitly assumes the unambiguous nature of performance feedback, despite the tradition of emphasis on

bounded rationality arising from ambiguity and adaptation in response to such ambiguity (March and Simon, 1958; March, 2010; March and Olsen, 1976). Recent research has highlighted the importance of relaxing the assumption of unambiguous performance feedback. One stream of literature suggests unambiguous performance can be interpreted ambiguously relative to different aspirations, e.g. historical vs. social aspiration (Joseph and Gaba, 2015; Baum, et al. 2005). The other stream of literature suggests feedback ambiguity arises from the task environment, e.g. noise, uncertainty (Posen & Levinthal, 2012), delay (Rahmandad, 2008; Fang and Levinthal, 2009). The divergence between the performance feedback from a strategic option and the overall organizational performance can be another source of feedback ambiguity, yet unstudied.

Based on attribution theory (Weiner, 1974), this paper proposes that organizations and organizational members tend to attribute organizational success to internal factors for self-enhancement motivation, and attribute organizational failure to external factors for self-protection motivation (Kelley and Michela, 1980). Thus, a negative feedback of overall failure will trigger the attribution towards external factors that are beyond the organization's control, weakening the organization's response to the internal feedback of strategic options.

Using the professional basketball data, I examine how organizations respond to multi-level performance feedback by resource allocation. Specifically, I examine how basketball teams respond to players' performance and the teams' performance in terms of allocation of playing time, a scarce resource to the teams. Allocation of playing time to players not only reflects a team's strategy to the game, but also serves as an online search for better performing options (players). I find that the playing time allocated to a player is positively associated with the player's previous

game performance. The positive relationship between the playing time allocated to a player and the player's performance feedback is weakened (strengthened) if the team experiences a loss (win), suggesting basketball teams attribute team success to the performance of its players, while attribute failure to it less. Align with this argument, I also find that, when the team experiences a blow-out loss, the effect of failure attribution is the strongest and the response to internal feedback is the weakest. The findings of this paper provide a new perspective for BTOF especially when organization receives multiple performance feedback at different levels. It also provides a micro-mechanism on how organizations respond to organizational success and failure.

Theory and Hypotheses

The allocation of resources responds to prior performance feedback reflects a learning process (Levitt & March 1988). Organizations and their members try to identify the superior strategic option by online search and allocate resources accordingly. If resources allocated to a strategic option receive a positive performance feedback, the feedback indicates the strategic option desirable, and more resources should be allocated to this option subsequently. With the assumption of resource scarcity, resources will also be divested from options that receive negative performance feedback, such that these resources could be reallocated to the better performing options. This learning and resource allocation/reallocation process of an organization could find its behavioral root from an individual's heuristic learning strategies such as "win-stay-lose-switch" (Nowak & Sigmund, 1993; Worthy, Hawthorne & Otto, 2013) and "hot-hand" (Gans, Knox & Croson, 2007).

In the context of professional basketball teams, the coaches and managers need to allocate the scarce and valuable resource of playing time among the players. Allocating playing time to the better performing players is critical to the overall success of the team. Coaches and managers learn from, and react to, the latest performance of players on a game-to-game basis.

Proposition 1: The amount of resource allocated to a strategic option is positively related to the option's previous performance.

Hypothesis 1: The minutes allocated to a player in a game is positively related to the player's previous game performance.

Organizations also learn from the organizational performance feedback, namely, organizational successes and failures. Understanding how organizational performance interacts with lower-level performance feedback about strategic options could uncover underlying mechanism of how organizations learn from performance feedback. Prior literature has shown mixed empirical results on whether organizations learn more from organizational successes or from organizational failures. In some cases, organizations (and individuals) learn more from success (Baum and Dahlin, 2007; KC, et al., 2013); in other cases, organizations learn more from failure (Li and Rajagopalan, 1997; Madsen and Desai, 2010). However, these research does not directly examine the mechanisms leading to their predictions. The interaction between organizational performance and performance of strategic options provides a pathway.

Attribution theory (Kelley and Michela, 1980) predicts that individuals tend to attribute failures to uncontrollable external factors while attribute success to internal factors of themselves. Such attribution is driven by individuals' motivation

to preserve self-image: success has the potential for the enhancement of self-esteem, thus attributing success to internal and controllable factors of oneself serves the motivation for self-enhancement; failure has the potential for negative implication for self-regard unless causal responsibility is attributed externally, thus attributing failure to external and uncontrollable factors serves the motivation for self-protection. Consequently, how success and failure are attributed will also influence the subsequent response: when attribution is directed internally towards controllable factors such as effort and learning, motivation for subsequent response will be stronger than when attribution is directed externally to factors such as luck.

Organizations, as a collective of organizational members, are not free from such attribution bias. Resource allocation between strategic options can be viewed as an internal factor within the organizational control. When the organization experiences an organizational failure, the organization and its members such as executives and managers will attribute the failure to external factors rather than the effectiveness of resource allocation between options. The organizational members will self-justify that, the resource allocation has little to do with the failure, and the performance feedback of the strategic options has little meaning for the organizational performance. Following this justification, the organization would respond little to the performance feedback of the strategic options. In contrast, when organization experiences an organization success, the organization and its members will attribute the success to the strategic resource allocation, especially those strategic options with better performance feedback. The organizational members will therefore respond more to the performance feedback of strategic options and allocate more resources to the better-performing options.

In the context of professional basketball teams, attribution theory suggests that when a team experiences a success / failure, the coach and manager will attribute the success / failure towards internal / external factors. That means, if a team experiences a win / loss, the team will respond more / less to its players' performance feedback in terms of minutes allocated to the players as predicted in Hypothesis 1.

Proposition 2. The positive relationship (in P1) between the amount of resource allocated to a strategic option and the option's prior performance, is stronger / weaker if the organization experienced prior success / failure.

Hypothesis 2. The positive relationship (in H1) between the minutes allocated to a player and the player's previous game performance, is positively / negatively moderated by the team's previous game win / loss.

Following the logic of attribution bias, when an organization experiences a strong organizational failure (e.g. large distance from aspiration), the organizational members will face more negative implication of self-regard, thus they may attribute the failure even less to the resource allocation, thus respond less to the performance feedback from strategic options. In contrast, when an organization experience a strong organization success (e.g. large distance from aspiration), the organizational members will have stronger motivation to attribute the success to the prior strategic action, including resource allocation, thus respond more to the performance feedback from strategic options. That means, the moderation effect of organizational performance feedback on the relationship

between resource allocation response and prior performance will be enhanced if the success or failure is stronger.

In the context of basketball teams, when a team experiences a blow-out success over the opponent, the team and its coach / manager will attribute the success more to internal factors, including the performance of its players, thus the team will respond more to the players' performance. In contrast, when a team experiences a blow-out defeat by the opponent, the team will attribute the failure less to the internal factor, thus respond less to the players' performance.

Proposition 3. The moderation effect (in P2) of organizational prior success / failure on the relationship between a strategic option's allocated resource and prior performance, is strengthened if the organization experience strong prior success / failure.

Hypothesis 3. The moderation effect (in H2) of the team's previous game win / loss on the relationship between a player's allocated minutes and his previous game performance, is stronger if the team's previous win or loss is big. That is,

H3(a) The positive relationship between a player's allocated minutes and his previous game performance is stronger if the team won the previous game and the score difference is large.

H3(b) The positive relationship between a player's allocated minutes and his previous game performance is weaker if the team lost the previous game and the score difference is large.

Method and data

To test the above hypotheses, I use longitudinal data of National Basketball Association (NBA) from the 2000/2001 to the 2015/2016 regular season, captured from basketball-reference.com. Basketball teams receive performance feedback following each game: the team receives overarching organizational feedback as wins or losses, as the primary goal of any competitive sports is to win; meanwhile the team also receives feedback regarding the player's on-court performance, in the form of player statistics. The teams will need to adjust the allocation of playing minutes, an important strategic resource to the team (Staw & Hoang, 1995; Ertug and Maoret, 2019; Zhang, 2017), among the players for a better subsequent team performance. Analyzing behavioral response to performance feedback of an organization (a team) and a strategic option (a player) in the setting of professional basketball teams have several empirical advantages. Firstly, the same teams' and the players' performance are available to both the teams and to the researcher. Second, playing minutes of a team, as an important strategic resource, is consistent across teams: a team could only allocate the playing minutes to 5 players for 48 minutes in each game. This setting avoids additional resources such as slack accumulated from prior organizational success. Third, organizational success and failure are clearly defined as win and loss in each basketball game. Fourth, the chance of organizational success and failure, in the form of wins and losses, are equal to 0.5 across the overall population of teams. This setting avoids situations in which successes and failures are extremely rare, yet way more impactful than the common experiences – e.g. in high reliability organizations such as air traffic controls and nuclear power plants, a failure

could be disastrous and catastrophic, containing more information for learning (Baum & Dahlin, 2007).

I gathered a sample of 386,431 player-game observations to test H1A and H1B. I exclude each NBA team's first game of each season as there is no performance feedback from prior game of that season.

Dependent variables

For H1, H2 and H3, I use a player's playing time (minutes) in a game as the dependent variable. It is a valuable organizational resource that the teams possess. A team divides the playing time among the players, and having the best players on court helps winning. For games with overtime, the players' minutes are scaled such that the total game length is constantly 48 minutes.

Independent variables and moderators

I use a players' previous game *relative efficiency per 48 minutes* as his previous game performance. As suggested by NBA official website, efficiency is used by NBA coaches to "evaluate a player's game performance", and is calculated as combination of a players' basic game statistics such as points, rebounds, assist, field goals made and missed, turnovers, etc. A player's efficiency is divided by his playing minutes in the game and times 48 minutes to form efficiency per 48 minutes, and then the efficiency per 48 minutes of the players of the same team in a game is standardized using z-score to form *relative efficiency per 48 minutes*. This measure of performance is highly correlated with other similar measure such as *relative PER* (0.84) and *Game Score per 48 minutes* (0.92).

The team's previous game win (loss) is used as the moderator for H2. The absolute final score difference in the previous game is used as the moderator for H3.

Controls

Following prior literature (Staw & Hoang, 1995; Zhang, 2017), variables are control at both team and player level. At team level, I control whether the game is a *home game*, the *team's past win-loss records*, the *number of games the team has played,* whether the focal team is from the *Western Conference,* and whether the focal team and the opponent are from the *same conference*. At the player level, I control the *personal fouls per 48 minutes* this game, *years in the league (squared) , years with the team (logged), years of collaboration with the coach (logged),* and whether the *player is a defensive player*. I also used year-player fixed effect to control unobserved variables at year and player level.

Results

Table 1 provides summary statistics and correlations for the variables. Table. Table 2 present the results for models predicting a player's allocated minutes. Model 2 suggests a player's previous game performance has a positive and significant effect ($p<0.001$) on the playing time allocated to him this game, supporting H1. Model 3 adds the moderating effect from previous game win. The team's previous game win positively and significantly moderates ($p<0.001$) the effect of player's previous game performance on his playing time this game, supporting H2. Figure 1 illustrate the moderating effect using the coefficient from Model 2. The figure shows the teams responds to individual player's previous performance more strongly when the previous game was a won, than

when the previous game was a loss. Figure 1 also shows that when teams experience failure, their unresponsiveness to performance feedback is not reflected in allocating less minutes to the better performing players. Rather, they do not divest resources from the poor performing players. This observation is an indirect evidence that the teams tend to not attribute poor team performance to poor performing team members.

Model 4 and 5 illustrate the moderation effect based on subsample. Model 4 illustrate the subsample of previous win and model 5 illustrate the subsample of previous loss. The coefficient of players' previous game performance is significant higher when the previous game won.

==Insert Table 1 and 2; Figure 1 and 2 about here==

Model 6 of Table 2 illustrate the three-way interaction between player's performance feedback, team's performance feedback, and absolute score difference. The coefficient of the three-way interaction is significant ($p<0.001$), supporting H3. For better understanding, Model 7 and 8 break down the subsample based on previous game win / loss. In Model 7 with the subsample of previous game won, the interaction between player's previous performance and previous game absolute score difference is not significant, not supporting H3A. In model 8 with the subsample of previous game lost, the interaction between player's previous performance and previous game absolute score difference is significant ($p<0.001$), supporting H3B. The three-way interaction is further illustrated in Figure 2. We can see that the minutes allocation response to the player's previous game performance is significantly weaker (flatter) when the team experience a loss in the previous game, and the score difference is large.

As the data from basketball-reference.com does not contain player's inaction (did not play, or DNP) due to coach's decision prior to 2013-14 season, our results may be flawed due to sampling bias. I also run a sub-sample analysis based on 2013-14 to 2015-16 season, the results are robust.

Discussion

Prior research in BTOF focuses on the organizations' behavioral responses to organizational performance feedback. Attainment discrepancy arises following failures or performance below aspiration, leading to problematic search with the motivation to close the performance gap in response (Gary et al, 2017), creating opportunities for awareness, learning and improvements (Madsen and Desai, 2010). In contrast to these arguments, this research contributes to the literature in two ways. First, organizations may receive performance feedback at different levels, rather than a single overarching performance feedback. Specifically, organization's response to the overall performance feedback and the feedback of strategic options may be intertwined, affecting how these feedbacks are attributed, interpreted and responded. By investigating how the alternative strategic options are evaluated and resources are allocated, conjecture to the organizational performance feedback, this research unveils the black box of organization response to performance feedback. Particularly, this research demonstrates the mechanism why organizations may learn less and respond less from failures, in support of prior empirical findings (KC, et al, 2013; Baum & Dahlin, 2007). That is, organizations tend to not attribute organizational failures to poor performing strategic options, and reluctant to divest resources from these options.

Second, prior literature also suggested that the motivation to close performance gap may lead to adjustment of aspiration. This research shows that the motivation of self-protection leads to adjustment of how failure is attributed. The possibility of attributing performance feedback between two levels of performance feedback give rises to bounded rationality due to attribution error (Foss & Webber, 2016). This argument could also provide alternative explanation to management phenomena. For example, a diversified company received different performance feedback from different business units and an overarching organizational performance feedback. When the company receives a negative organizational performance feedback, the managers may attribute the failure externally rather than internally towards the underperforming units, and are reluctant to divest these units and concentrate resources to the better performing units. In contrast, for an undiversified company with a single (or a limited number of) business unit(s), the organizational performance feedback is identical to the performance of the business unit(s), thus it is difficult for the managers to direct the organization failure away from the business unit's failure, forcing the managers to respond. This could be an explanation to diversification discount.

Although NBA basketball data provides me a detailed context, there are several limitations requiring further studies. Although coaches and managers face pressure from underperformance, NBA teams do not face elimination from the league. In addition, teams with low performance will purposely "tank", or seek loses, such that they will get better rookie picks for the next season. In a more competitive environment with little tolerance for failure, organizations may respond differently to organizational failure. Secondly, as mentioned above,

the minutes allocation in a team reflects a diversified organization with spontaneous pursue of multiple online search, giving opportunity for attribution error. The size of available strategic option space, or the level of diversification, could be a boundary condition. Our findings offer a call for future work to investigate organizational response to multi-level performance feedback.

List of Tables for Essay C

Table 1: Correlation table – player-game level

	Variable	Mean	Std Dev	Min	Max	1	2	3	4	5	6	7	8	9	10	11	12	13
1	Player's allocated minutes	23.47	11.51	0.00	48.00	1.000												
2	Player's previous game performance	0.03	0.92	-3.46	3.44	0.204	1.000											
3	Previous game win	0.50	0.50	0.00	1.00	-0.011	0.006	1.000										
4	Previous game absolute score difference	10.79	7.75	1.00	55.00	0.004	0.007	0.003	1.000									
5	Team's past win-loss record this season	0.50	0.19	0.00	1.00	-0.017	0.001	0.375	0.008	1.000								
6	Team's game number this season	41.40	23.29	2.00	82.00	0.017	-0.003	0.002	0.010	0.006	1.000							
7	Player's fouls per 48 min this game	5.00	5.52	0.00	192.00	-0.274	-0.020	-0.011	-0.001	-0.012	-0.017	1.000						
8	Player's years in the league	5.13	4.24	0.00	31.00	0.112	0.024	0.059	0.005	0.156	-0.023	-0.059	1.000					
9	Player's years with the team (log)	0.72	0.70	0.00	3.00	0.292	0.128	0.059	0.007	0.159	-0.020	-0.074	0.256	1.000				
10	Defensive player	0.49	0.50	0.00	1.00	-0.173	0.085	-0.004	-0.001	-0.010	0.000	0.175	-0.075	-0.016	1.000			
11	Player's year of collaboration with coach (log)	3.71	1.25	0.00	7.08	0.239	0.080	0.078	0.003	0.184	-0.012	-0.056	0.020	0.300	-0.007	1.000		
12	Team is from West Conference	0.50	0.50	0.00	1.00	-0.006	-0.001	0.055	0.010	0.163	0.000	-0.006	0.002	0.034	0.006	0.058	1.000	
13	Both team from the same conference	0.64	0.48	0.00	1.00	0.001	0.000	0.022	0.006	0.002	0.092	0.002	-0.004	-0.003	0.000	-0.006	0.001	1.000

Table 2: Fixed-effect OLS on a player's playing time (minutes)

VARIABLES	(1) Full	(2) Full
Player's previous game performance (H1)		0.961***
		(0.0306)
Previous game win		
Player's previous game performance*Previous game win (H2)		
Previous game absolute score difference		
Player's previous game performance*Previous game absolute score difference (H3A/3B)		
Previous game win*Previous game absolute score difference		
Player's previous game performance*Previous game win*Previous game absolute score difference (H3)		
Team's past win-loss record this season	-5.832***	-5.704***
	(0.391)	(0.384)
Team's game number this season	0.0135***	0.0135***
	(0.00141)	(0.00139)
Player's fouls per 48 min this game	-0.307***	-0.306***
	(0.00774)	(0.00764)
Player's years in the league (squared)	-0.0881***	-0.0854***
	(0.0193)	(0.0189)
Player's years with the team (log)	1.685***	1.630***
	(0.180)	(0.176)
Defensive player	-2.207***	-2.186***
	(0.349)	(0.345)
Player's year of collaboration with coach (log)	0.556***	0.548***
	(0.0342)	(0.0337)
Team is from Western Conference	-0.487**	-0.445*
	(0.241)	(0.237)
Both team from the same conference	0.0755***	0.0726***
	(0.0275)	(0.0272)
Constant	29.45***	29.27***
	(0.983)	(0.962)
Season and player fixed effect	Yes	Yes
Clustered std. err. on players	Yes	Yes
Observations	386,431	386,431
R-squared	0.484	0.489

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 2: Fixed-effect OLS on a player's playing time (minutes) – cont'

VARIABLES	(3) Full	(4) Previous Win	(5) Previous Loss
Player's previous game performance (H1)	0.826***	1.075***	0.832***
	(0.0309)	(0.0394)	(0.0317)
Previous game win	-0.297***		
	(0.0287)		
Player's previous game performance*Previous game win (H2)	0.276***		
	(0.0355)		
Previous game absolute score difference			
Player's previous game performance*Previous game absolute score difference (H3A/3B)			
Previous game win*Previous game absolute score difference			
Player's previous game performance*Previous game win*Previous game absolute score difference (H3)			
Team's past win-loss record this season	-5.408***	-5.940***	-5.501***
	(0.390)	(0.434)	(0.420)
Team's game number this season	0.0135***	0.00252*	0.0240***
	(0.00139)	(0.00139)	(0.00181)
Player's fouls per 48 min this game	-0.306***	-0.288***	-0.323***
	(0.00764)	(0.00824)	(0.00854)
Player's years in the league (squared)	-0.0853***	-0.0874***	-0.0833***
	(0.0188)	(0.0194)	(0.0187)
Player's years with the team (log)	1.627***	1.629***	1.640***
	(0.175)	(0.180)	(0.181)
Defensive player	-2.184***	-2.041***	-2.340***
	(0.344)	(0.367)	(0.347)
Player's year of collaboration with coach (log)	0.549***	0.517***	0.573***
	(0.0337)	(0.0360)	(0.0371)
Team is from Western Conference	-0.445*	-0.396	-0.439*
	(0.237)	(0.254)	(0.235)
Both team from the same conference	0.0788***	0.0908**	0.0625
	(0.0272)	(0.0390)	(0.0381)
Constant	29.26***	30.23***	28.37***
	(0.961)	(1.091)	(0.876)
Season and player fixed effect	Yes	Yes	Yes
Clustered std. err. on players	Yes	Yes	Yes
Observations	386,431	194,259	192,110
R-squared	0.490	0.519	0.467

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 2: Fixed-effect OLS on a player's playing time (minutes) – cont'

VARIABLES	(6) Full	(7) Previous Win	(8) Previous Loss
Player's previous game performance (H1)	1.040***	1.043***	1.048***
	(0.0474)	(0.0472)	(0.0484)
Previous game win	0.129***		
	(0.0487)		
Player's previous game performance*Previous game win (H2)	0.0303		
	(0.0506)		
Previous game absolute score difference	0.0244***	-0.0166***	0.0229***
	(0.00273)	(0.00268)	(0.00269)
Player's previous game performance*Previous game absolute score difference (H3A/3B)	-0.0198***	0.00314	-0.0200***
	(0.00301)	(0.00255)	(0.00300)
Previous game win*Previous game absolute score difference	-0.0409***		
	(0.00399)		
Player's previous game performance*Previous game win*Previous game absolute score difference (H3)	0.0229***		
	(0.00398)		
Team's past win-loss record this season	-5.325***	-5.867***	-5.406***
	(0.389)	(0.432)	(0.418)
Team's game number this season	0.0134***	0.00266*	0.0238***
	(0.00139)	(0.00139)	(0.00180)
Player's fouls per 48 min this game	-0.306***	-0.288***	-0.323***
	(0.00763)	(0.00823)	(0.00854)
Player's years in the league (squared)	-0.0853***	-0.0874***	-0.0832***
	(0.0188)	(0.0194)	(0.0186)
Player's years with the team (log)	1.626***	1.630***	1.637***
	(0.175)	(0.180)	(0.181)
Defensive player	-2.181***	-2.039***	-2.337***
	(0.344)	(0.367)	(0.346)
Player's year of collaboration with coach (log)	0.550***	0.518***	0.575***
	(0.0336)	(0.0360)	(0.0371)
Team is from Western Conference	-0.443*	-0.392	-0.440*
	(0.237)	(0.254)	(0.235)
Both team from the same conference	0.0827***	0.0952**	0.0661*
	(0.0273)	(0.0391)	(0.0380)
Constant	28.95***	30.36***	28.07***
	(0.958)	(1.093)	(0.873)
Season and player fixed effect	Yes	Yes	Yes
Clustered std. err. on players	Yes	Yes	Yes
Observations	386,431	194,259	192,110
R-squared	0.490	0.520	0.468

List of Figures for Essay C

Figure 1: Performance feedback's effect on subsequent minutes allocation, moderated by previous success (H2). Coefficients based on Table 2 model 3 with centered controls

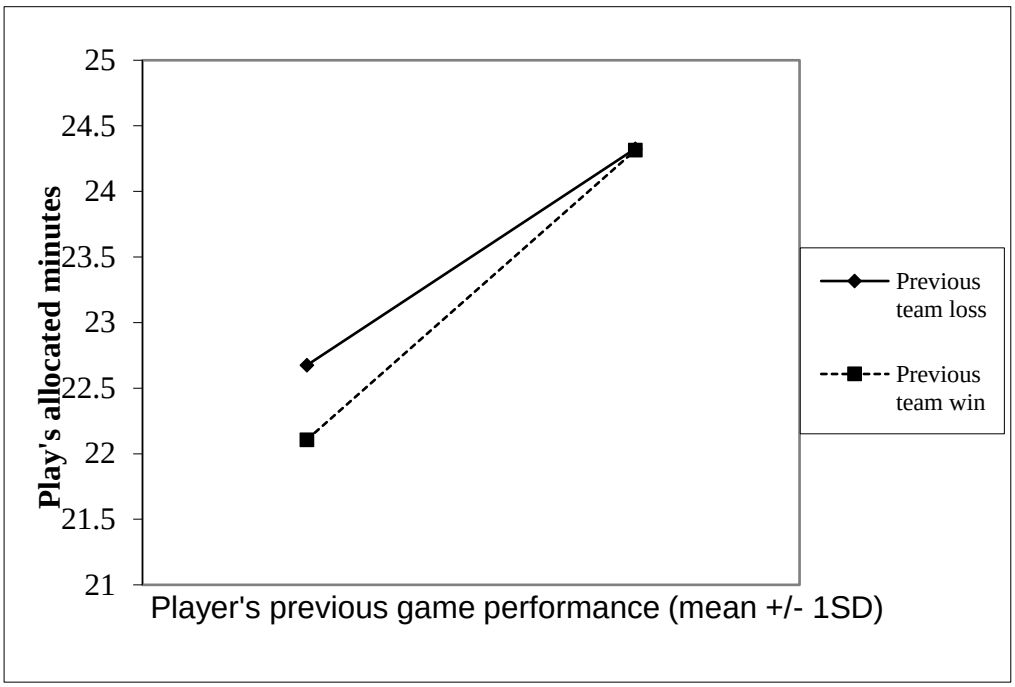
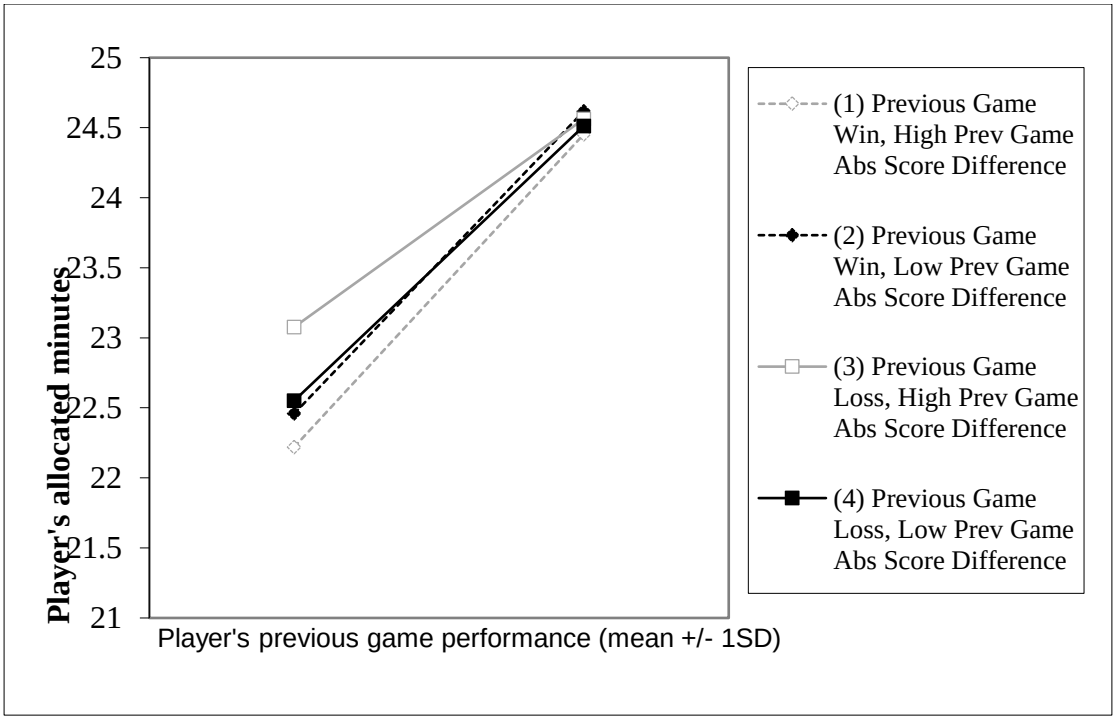


Figure 2: Player's allocated minutes, three-way interaction between player's performance feedback, team's performance feedback, and absolute score difference (H3). Coefficients based on Table 2 model 6 with centered controls.



References for Essay C

Baum, J. A., & Dahlin, K. B. (2007). Aspiration performance and railroads' patterns of learning from train wrecks and crashes. *Organization Science*, *18*(3), 368-385.

Baum, J. A., Rowley, T. J., Shipilov, A. V., & Chuang, Y. T. (2005). Dancing with strangers: Aspiration performance and the search for underwriting syndicate partners. *Administrative Science Quarterly*, *50*(4), 536-575.

Bromiley, P. (1991). Testing a causal model of corporate risk taking and performance. *Academy of Management journal*, *34*(1), 37-59.

Chen, W. R., & Miller, K. D. (2007). Situational and institutional determinants of firms' R&D search intensity. *Strategic Management Journal*, *28*(4), 369-381.

Ertug, G., & Maoret, M. (2019). Do coaches in the National Basketball Association actually display racial bias? A replication and. Academy of Management Discoveries. In-Press

Fang, C., & Levinthal, D. (2009). Near-term liability of exploitation: Exploration and exploitation in multistage problems. *Organization Science*, *20*(3), 538-551.

Foss, N. J., & Weber, L. (2016). Moving opportunism to the back seat: Bounded rationality, costly conflict, and hierarchical forms. *Academy of Management Review*, *41*(1), 61-79.

Gaba, V., & Joseph, J. (2013). Corporate structure and performance feedback: Aspirations and adaptation in M-form firms. *Organization Science*, *24*(4), 1102-1119.

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

Gary, M. S., Yang, M. M., Yetton, P. W., & Sterman, J. D. (2017). Stretch goals and the distribution of organizational performance. *Organization Science*, *28*(3), 395-410.

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

Kc, D., Staats, B. R., & Gino, F. (2013). Learning from my success and from others' failure: Evidence from minimally invasive cardiac surgery. *Management Science*, *59*(11), 2435-2449.

Kelley, H. H., & Michela, J. L. (1980). Attribution theory and research. *Annual review of psychology*, *31*(1), 457-501.

Kim, T., & Rhee, M. (2017). Structural and behavioral antecedents of change: Status, distinctiveness, and relative performance. *Journal of Management*, *43*(3), 716-741.

Levitt, B., & March, J. G. (1988). Organizational learning. *Annual review of sociology*, *14*(1), 319-338.

Li, G., & Rajagopalan, S. (1997). The impact of quality on learning. *Journal of Operations Management*, *15*(3), 181-191.

Madsen, P. M., & Desai, V. (2010). Failing to learn? The effects of failure and success on organizational learning in the global orbital launch vehicle industry. *Academy of management journal*, *53*(3), 451-476.

March, J. S., & Simon, H. A. (1958). HA (1958) Organizations. *New York*.

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

March, J. G., & Olsen, J. P. (1976). The technology of foolishness.

Miller, K. D., & Leiblein, M. J. (1996). Corporate risk-return relations: Returns variability versus downside risk. *Academy of Management Journal*, *39*(1), 91-122.

Nowak, M., & Sigmund, K. (1993). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature*, *364*(6432), 56-58.

Park, K. M. (2007). Antecedents of convergence and divergence in strategic positioning: The effects of performance and aspiration on the direction of strategic change. *Organization Science*, *18*(3), 386-402.

Posen, H. E., Keil, T., Kim, S., & Meissner, F. D. (2018). Renewing research on problemistic search—A review and research agenda. *Academy of Management Annals*, *12*(1), 208-251.

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

Rahmandad, H. (2008). Effect of delays on complexity of organizational learning. *Management Science*, *54*(7), 1297-1312.

Staw, B. M., & Hoang, H. (1995). Sunk costs in the NBA: Why draft order affects playing time and survival in professional basketball. *Administrative Science Quarterly*, 474-494.

Weiner, B. (Ed.). (1974). *Achievement motivation and attribution theory*. General Learning Press.

Worthy, D. A., Hawthorne, M. J., & Otto, A. R. (2013). Heterogeneity of strategy use in the Iowa gambling task: A comparison of win-stay/lose-shift and reinforcement learning models. *Psychonomic bulletin & review*, *20*(2), 364-371.

Zhang, L. (2017). A fair game? Racial bias and repeated interaction between NBA coaches and players. *Administrative Science Quarterly*, *62*(4), 603-625.