ASPIRATION LEVELS AND EXPLORATION-EXPLOITATION: AN ADAPTIVE LEARNING APPROACH

ABSTRACT

How do high and moderate aspiration levels compare in terms of affecting the trade-off between exploration and exploitation? Recent reviews of empirical work have challenged the widely held belief that high aspiration levels lead to more exploration. Although exploration–exploitation decisions are often viewed as a reinforcement learning process, much prior work explores this question as a choice process, i.e., deciding between relatively more or less risky options. After developing a simple agent-based model to understand how aspiration levels influence adaptive learning from feedback, we report on behavioral laboratory experiments used to test our model's predictions. In the canonical multi-armed bandit problem, we show that subjects assigned a high aspiration level exploit more – and that, when they do explore, they do so more narrowly – as compared with subjects assigned a moderate aspiration level. A high (moderate) aspiration level reduces (increases) feedback ambiguity about the relative attractiveness of different options, which affects agents' subsequent sampling and hence their learning. The low levels of exploration engendered by high aspiration levels are advantageous in stable environments, but they impair performance in unstable environments.

INTRODUCTION

Making decisions in uncertain environments is a fundamental inquiry in behavioral theories of organizations. Since the seminal contribution of March (1991), the trade-off in organizational and individual decision making between exploration and exploitation has benefited from a long tradition of research scrutiny (for a meta-analysis, see Junni, Sarala, Taras, and Tarba, 2013; for a recent review, see Posen, Keil, Kim, and Meissner, 2018). This trade-off between *gathering* information (i.e. engaging in activities associated with search, risk taking, experimentation, and innovation) and *using* the acquired information (i.e. activities involving production, efficiency, implementation, and execution) is especially pertinent in changing environments, where learning from prior experience may then be less useful for making choices about future actions (Posen and Levinthal, 2012).

The trade-off between exploration and exploitation is often studied in terms of reinforcement learning (Cohen, McClure, and Yu, 2007; Daw, O'doherty, Dayan, Seymour, and Dolan, 2006; Denrell and March, 2001; Li, Mayhew, and Kourtzi, 2009; March, 1996; Sutton and Barto, 1998). Scholars in this vein of research distinguish between two processes: the *choice* process of deciding between exploiting the currently best-performing option for immediate reward or exploring other, currently inferior options for additional information; and the *adaptive learning* process of translating feedback from accumulated experience into representations about the task environment. These two processes are closely intertwined because (a) choices determine what feedback is received and (b) representations shaped by learning determine subsequent choices (Li et al., 2009).

However, decision makers rarely learn and adapt without some objective in mind. The research and development (R&D) manager searches for new innovations with a performance target; the investor tries to beat the market; and the project manager tries to surpass last year’s performance. Aspiration levels are fundamental to “satisficing” (Simon, 1955, 1997), and they have a strong effect on search and decision making (Cyert and March, 1963). Yet there is hardly any research devoted to exploring exactly how aspiration levels or reference points affect feedback-driven learning. We are therefore motivated to study how aspiration levels influence the exploration–exploitation trade-off in uncertain environments.

It should be noted that behavioral theories of organizations do explore the effect of the aspiration–performance gap on the exploration–exploitation trade-off. However, they do so by focusing mainly on choice making while largely ignoring the adaptive learning process. These literatures study the propensity of a firm to explore versus exploit following performance feedback that deviates from its aspiration level (See Posen et al, 2018; Schilke, 2012 for recent reviews), but they are less concerned with how aspiration levels may influence adaptive learning.

Simon (1955) argued that aspirations perform an encoding function for the decision maker – in effect, categorizing the feedback space into successes and failures. This encoding of continuous outcomes into categorical feedback has the effect of strengthening beliefs about the relative attractiveness of available alternatives. Such increased opinion strength will likely have profound effects on subsequent sampling, which could skew learning in ways not readily explained when one considers only the objective payoffs. Thus, we build on Posen and Levinthal's (2012) insight that the confidence of decision makers in the relative attractiveness of different choices (i.e., their *opinion strength*) may endogenously influence their exploration behavior even when their exploration strategy remains constant. Our paper takes this intuition one step further and suggests that opinion strength is in itself a function of the decision maker's aspiration level.

In comparison with moderate aspiration levels, for example, high aspiration levels will result in relatively more instances of feedback being categorized as failures. Thus a decision maker with a high aspiration level is likely to categorize much of the feedback from any low-payoff option as failures, thereby increasing the high-payoff option's relative attractiveness. This asymmetry in feedback ambiguity, which arises among decision makers with different aspiration levels, affects the perceived relative attractiveness of options and hence the decisions made about exploration and exploitation. Therefore, in our model, an increased aspiration level (and subsequent failure to meet it) need not result in the increased exploration activity posited by prior studies. Rather, a decision maker with the same propensity for exploration but with a higher aspiration level may be more inclined to select higher-payoff alternatives–a strategy that may actually result in more exploitation.

To develop our hypotheses, we use a simple agent-based model that allows us to explore the effects of aspiration levels and feedback ambiguity in a multi-arm bandit task. We use the same underlying model as the backbone of a set of behavioral experiments to test these hypotheses. In the lab experiment, participants are randomly assigned to the condition of a high or moderate aspiration level; then they make choices among uncertain R&D investment options and learn from feedback. The experimental manipulation employs the idea of "mere goals", which simply sets a reference point without attached performance consequences (cf. Heath, Larrick, and Wu, 1999; Larrick, Heath, and Wu, 2009), and we observe how the aspiration level manipulation alters participants' behavior.

We find that high aspiration levels can lead to a reduction in feedback ambiguity, which results in participants more reliably identifying the best option, or *arm*, and so leads not only to greater exploitation but also to narrower exploration. These baseline results are for the case of a stable environment, but we also study the effect of a disruptive change to the environment that alters the relative attractiveness of different payoffs. In that case, we find that the reduction in feedback ambiguity – which is beneficial in stable environments – can become a liability in unstable environments. Thus we document that the propensity of decision makers to exploit, both early and often, what is initially the best arm has a deleterious effect: it delays their adaptation to the exogeneous shock, and as a consequence, their performance recovers more slowly than that of decision makers who have lower opinion strength and therefore engage in more exploration. In addition, we 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. Our empirical findings suggest that high aspiration levels may lead to more exploitation when both the choice and learning processes are considered. This outcome is at odds with the prevailing view that high aspiration levels are prone to induce high levels of exploration, a view that has so far received mixed empirical support (Posen et al., 2018; Shinkle, 2012). From a theoretical standpoint, we identify the mechanism underlying this contrasting result by proposing that the influence of high versus moderate aspiration levels on the exploration–exploitation trade-off depends on how those levels affect the learning process via the relative ambiguity of the feedback received. We derive these insights by applying aspiration levels to a theory of adaptive sampling that integrates choice and learning processes into a unified model (cf. Denrell, 2008) . Thus we enrich the extensive literature on the behavioral theory of the firm, especially with regard to an organization's response to the aspiration–performance gap (Gary, Yang, Yetton, and Sterman, 2017; Greve, 2018; Greve and Gaba, 2017; Posen et al., 2018; Shinkle, 2012).

ASPIRATION LEVELS AND THE TRADE-OFF BETWEEN EXPLORATION AND EXPLOITATION IN UNCERTAIN ENVIRONMENTS

Human decision making in uncertain environments requires a trade-off between exploration and exploitation (March, 1991). A classic instance of this trade-off is the gambler choosing between multiple slot machines. Should she exploit and choose the option with the highest average payoff from past experience; or should she explore and choose the option with which she is less familiar but which may prove to be more profitable? This exploitation–exploitation trade-off is critical to a broad spectrum of human decision-making efforts among uncertain choices; examples include individuals investing for retirement, managers choosing between projects, and investors choosing between asset classes.

The trade-off between exploration and exploitation is perhaps most acute in nonstationary environments, where prior learning from feedback may no longer be an accurate guide to future decisions (Posen and Levinthal, 2012). This decay in the value of prior information implies that decision makers must remain open to a broad set of options, increasing the value of continuing to explore different opportunities (e.g. Eisenhardt and Martin, 2000). Thus a broader exploration of multiple diverse options as well as quick adaptation are necessary for organizations to survive and prosper in rapidly changing environments (Zheng and Srikanth, 2019). Since individuals' choices are influenced by learning from feedback on prior trials, scholars have suggested that explore-exploit decisions are prone to such learning pathologies as the "hot stove" effect – whereby the decision maker forgoes a thorough exploration of unfamiliar choices after experiencing negative payoffs from a few trials (Denrell and March, 2001; March, 1996).

According to behavioral theories of organizations, the extent of exploration and exploitation depends not only on the environment but also on the organization's performance relative to its expectations (Cyert and March, 1963).[1] 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, Miller, and Rau, 2001; Bromiley and Rau, 2019; Greve, 2003; Greve and Gaba, 2017; Posen et al., 2018; Shinkle, 2012). Yet it appears that two key ideas from the literatures building on the behavioral theory of the firm – performance relative to aspirations and adaptive learning – may yield different predictions about the effects of increasing aspiration levels on the exploration–exploitation trade-off. We discuss these contrasting predictions in turn.

High aspiration levels may increase exploration. The behavioral theory of the firm argues that aspirations and performance are closely intertwined. One central idea is that aspiration levels determine a reference point or satisficing criterion that managers aim to reach (Cyert and March, 1963; Simon, 1947). Thus this theory suggests that, when performance is lower than aspirations, managers explore more by engaging in problemistic search whereas, when performance exceeds aspirations, managers exploit more by maintaining the same activities to which (they believe) that superior performance is due (Cyert and March, 1963; Lant and Shapira, 2008; Levitt and March, 1988). The behavioral theory of the firm suggests also that aspirations themselves may change as a function of achieved performance (March, 1988).[2]

The strategic reference point theory (Fiegenbaum, 1990; Fiegenbaum, Hart, and Schendel, 1996) similarly posits that firms' aspirations serve as reference points. In contrast to the behavioral theory of the firm, research along these lines argues that aspirations are often externally induced rather than set by managers (Fiegenbaum and Thomas, 1995; see also Greve, 2002). Drawing from prospect theory, these scholars argue that performance below (above) aspirations is encoded as a failure (a success) that leads to

[1] For ease of exposition, we portray the organization as a decision maker. The processes are obviously more complex, and organizational decisions are more accurately interpreted as choices made by the currently dominant coalition. This interpretation is implied for the purpose of brevity in expressing ideas.
[2] Shinkle (2012) notes that there are only a few studies addressing whether, as suggested by March (1988), organizational aspirations actually change according to achieved performance.

increased risk seeking (risk aversion) by managers. These scholars offer empirical evidence in the form of such observed exploratory, risk-seeking activities as engaging in acquisitions, undertaking organizational change, investing in R&D, and committing to long-term capital expenditures (for reviews, see Posen et al., 2018; Shinkle, 2012).

Overall, these theories of organizations predict that firms with higher aspiration levels (reference points) are likely to explore more. The mechanism proposed as driving this dynamic is the greater difficulty of achieving higher aspiration levels – from which it follows that such organizations are, ceteris paribus, more likely to experience failure than are firms with lower aspiration levels. Experiencing more failure is likely to trigger problemistic search and/or risk-seeking behaviors, in short, more exploration (Greve, 2017, 2018).

These organizational arguments parallel findings at the individual level. For example, several scholars have questioned the practical utility of setting challenging goals for individuals (for reviews and commentary, see Greve, 2010; Ordónez, Schweitzer, Galinsky, and Bazerman, 2009a; 2009b), by identifying several undesirable consequences of setting such goals. Of specific relevance to our discussion is the argument that high aspirations, or challenging goals, induce individuals to make risky decisions (Gary et al., 2017; Heath et al., 1999; Larrick et al., 2009). This research also suggests that, because individuals tend to choose options with a high variance in payoffs, increased aspiration levels does not improve overall performance.

High aspiration levels might not increase exploration. Empirical work has reached mixed conclusions regarding whether or not performance below aspirations increases exploration. In a comprehensive review, Posen et al. (2018) find that only little more than half the studies support the claim that performance below aspirations triggers search; the other studies report non-significant or even opposite results (see also Shinkle, 2012).[3] At the individual level, Locke and Latham (2006) perform a comprehensive review of goal setting literature, and suggest that, to the extent that individuals are committed to goals and are able to attain them, "there is a positive linear relationship between goal

[3] There is a second reference point, the *survival* point, at which firms drop so far below their aspiration levels that they cease exploring (see e.g. March & Shapira, 1987; 1992). Some have argued that, if performance is extremely high relative to aspirations, then the firm engages in "slack search" and may then explore more (Shinkle, 2012).

difficulty and task performance" (Locke and Latham, 2006, p. 265), since challenging goals increase effort and persistence. It is worth noting that this relationship is articulated here in terms of performance – and not with reference to whether higher aspirations influence exploration versus exploitation – although they suggest that challenging goals motivate the search for high-performance strategies (i.e., exploration) and the subsequent exploitation of those strategies (cf. Earley, Connolly, and Ekegren, 1989).

Besides these empirical inconsistencies regarding how aspiration levels affect exploration-exploitation, scholars have posited different theoretical mechanisms that may underlie this relationship. For instance, adaptive learning scholars have argued that, in reality, individuals and organizations are seldom presented with plausible choices and their outcome distributions; instead, these need to be learned from experience (Denrell and Le Mens, 2011; Gigerenzer, 2008; Gigerenzer and Gaissmaier, 2011; Le Mens and Denrell, 2011; also see Bromiley and Rau, 2019). The implication is that such adaptive learning is likely to influence choice behavior *in conjunction* with the aspiration–performance gap (Denrell, 2008). According to Denrell (2008), such an adaptive sampling process, under some circumstances, can result in choice patterns that are similar to those predicted by behavioral theories – even though managers are actually risk neutral.

Though adaptive learning offers an important theoretical lens to understand how aspiration levels may affect exploration-exploitation, empirical work in this tradition is so far limited. At the individual level, empirical studies generally involve a one-time decision of choosing between options whose outcomes exhibit lower or higher variability. The few multi-period studies examine how goals affect self-regulatory processes, such as self-efficacy (Cervone, Jiwani, and Wood, 1991; Wood, Bandura, and Bailey, 1990), or when the links between actions and outcomes are complex and difficult to comprehend (Gary et al., 2017; Kanfer and Ackerman, 1989). Thus, these studies shed little light on how adaptive learning influences the relationship between aspiration levels and exploration-exploitation choices.

At the organizational level, though empirical work often employs longitudinal datasets, they have also not examined the adaptive learning mechanism in detail (Bromiley and Rau, 2019 and Posen et al., 2018 offer more detailed critiques of the theory-data gap in this literature). This is perhaps because in these empirical contexts there may not be a need to learn from feedback about the relative attractiveness of the choices available to the firm. For example, in stable industries, managers may already have insight into the probable outcomes of ongoing products or projects allowing them to choose among the available strategic options. However, adaptive learning becomes important when firms lack such prior knowledge, such as when confronted with a slate of new technologies or business models, and/or when prior knowledge is not useful, such as after environmental shocks that re-order the attractiveness of existing options. It is plausible that this contingency of choice under known conditions versus under adaptive learning accounts for the tension in the empirical literature.

In this paper, we study the choice and learning processes *jointly* in the context of how aspiration levels influence the exploration-exploitation trade-off. We fill an important gap in the empirical literature that has so far mainly focused on choice processes alone, even though theoretical work usually models exploration-exploitation as a reinforcement learning process that contains both choice and learning (Daw et al., 2006; Frank, Doll, Oas-Terpstra, and Moreno, 2009; Li et al., 2009; March, 1996; Sutton and Barto, 1998).

Predictions regarding the consequences of increased aspiration levels are less straightforward under adaptive learning theory. Here, adaptive learning is likely sensitive to how the aspiration levels differentially influence the interpretation of feedback and subsequent learning about the relative attractiveness of the available choices. To examine more carefully how aspiration levels influence the choice between exploration and exploitation in uncertain environments, we build a simple adaptive learning model to develop hypotheses. We then test our hypothesized predictions in a series of behavioral laboratory studies.

ADAPTIVE LEARNING MODEL

In order to inform our theory building, we develop a model for how adaptive learning affects the exploration–exploitation choice at different aspiration levels. Our first step is the following thought experiment. Consider the two-arm bandit task illustrated in Figure 1a, where the "better arm" returns a payoff uniformly distributed between 0.5 and 1.5 and the "worse arm" returns a payoff uniformly distributed between 0 and 1. 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 the figure); and one with a moderate aspiration level of 0.75, which is equal to the average payoff of both arms (DM2).

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

If the agents were aware of the true payoff distributions for both these arms, then the choice between them would be trivial because agents will always exploit the better arm regardless of their respective aspiration levels. More often than not, however, the payoff distributions of the arms are unknown to the decision makers. Since there is considerable overlap in the payoffs of the two arms, it follows that identifying the better arm requires learning from multiple trials.[4]

In Simon (1955, p. 105, Figure 1), aspirations perform an encoding function that reduces a complex environment into a smaller number of states. Simon argues that this encoding function is a fundamental purpose of aspirations, which they serve by partitioning the payoff space into successes and failures (see also Heath et al., 1999). Applied to our thought experiment in Figure 1b, DM1 (with a high aspiration level) has a 50% chance of success when pulling the better arm and a 0% chance of success when pulling the worse arm; whereas DM2 (with a moderate aspiration level) has a 75% chance of success when pulling the better arm and 25% chance of success when pulling the worse arm.

According to adaptive learning theory, agents sample the choices available to them in order to learn their payoff distributions. Then choices that lead to higher utility are more likely to be resampled than are those that lead to lower utility. To translate the successes and failures into utilities, we simply assume that their respective payoffs are +1 and −1. Because aspiration levels influence how decision makers interpret feedback (i.e., via categorization into successes and failures), they also influence the adaptive learning process by affecting the decision makers' sampling choices. Table 1 shows how – as the number of samples taken increases – the probability of choosing the worse arm changes for a decision maker with a high versus moderate level of aspirations. More specifically: with an increase in how often

[4] 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 we isolate the exploration problem from risk taking (or preferences for risk).

the two arms are sampled, the likelihood of choosing the worse arm declines much more rapidly for the high- than for the moderate-aspiration agents.[5]

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

Why does a decision maker's aspiration level affect her propensity to sample the worse arm? When aspiration levels are high, 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 the decision maker that this arm should be avoided. Thus, the high aspiration level - and the resulting categorization of feedback into successes and failures - has the effect of reducing ambiguity in received feedback. In contrast, decision makers with only moderate aspiration levels receive successes and failures from both the better arm and the worse arm. When the number of trials is limited, these mixed successes and failures send an ambiguous signal about both arms; hence the decision maker will be more uncertain (than if she had high aspirations) about which arm is worse (and which arm is better). In our thought experiment, then, an agent with high aspirations is more confident about identifying the inferior arm (and therefore the superior arm) than is an agent with moderate aspirations.

The mechanism driving our thought experiment is Posen and Levinthal's (2012) insight that a reduction in "opinion strength" – that is, in the decision maker's confidence about the two arms' relative attractiveness – endogenously increases exploration behavior even if the decision strategy is constant. We add to this account by suggesting that opinion strength may itself be a function of the decision maker's aspiration level. Overall, these arguments suggest the following proposition.

The influence of different aspiration levels on exploration versus exploitation depends on how much feedback ambiguity is generated by those levels. To the extent that higher aspiration levels

[5] The probabilities in Table 1 can be calculated in closed form. Take, for example, the case of high aspirations; after two trials on each arm, the worse arm will have returned two failures. So in order for the decision maker to pull the worse arm, he must also experience two failures from the better arm (for which the probability is 0.5×0.5) – after which he will randomly choose between them (0.5 probability). That sequence gives a probability of $0.5 \times 0.5 \times 0.5 = 12.5\%$. The choice rule reflects the mindset of a naïve decision maker who only choose the arm with the better average payoff. Alternatively, a more sophisticated (Bayesian) decision maker could choose by calculating the conjugate prior based on previous trial experiences. This alternate approach has no effect on our main result.

are more likely to reduce the ambiguity of feedback on 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-arm choice task in a stable environment. However, choices regarding exploration and exploitation often arise in more complex decision tasks that may involve multiple options or changing environments. In a task involving multiple choices, the possible effect of high aspiration levels on feedback ambiguity has implications not only for the level of exploration but also for how widely the decision maker explores (i.e., the breadth of exploration). The reason is that aspiration levels can change feedback ambiguity, and hence opinion strength, differentially across the available choices. Thus a *three*-arm bandit task is the minimal setup in which we can demonstrate our theory with respect to the level of exploitation and the breadth of exploration. Finally, the mechanism outlined here relies on differences in the beliefs formed through adaptive learning. Hence we can demonstrate the mechanism more convincingly in an environment where such opinion strength can be a liability – for example, in an unstable environment (Posen and Levinthal, 2012). Since a pure thought experiment with three options in a changing environment quickly becomes intractable, we develop a simple agent-based simulation to explore the effect of feedback ambiguity on exploration breadth and to develop our hypotheses. We then use the same task setup to test the hypotheses in a lab-based behavioral experiment.

Agent-Based Simulation

In order to understand how adaptive learning influences the exploration–exploitation trade-off under different aspiration levels, we study a three-arm bandit 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.

Thus we consider a three-arm bandit for which the arms' payoffs are normally distributed with means of [1.0, 0.5, 0], each with a standard deviation of 1. The agents can obtain information about the arms only by sampling and adaptive learning. Following previous work (Denrell and March, 2001; Puranam, Stieglitz, Osman, and Pillutla, 2015; Sutton and Barto, 1998), we assume that an agent's beliefs about the attractiveness of an arm at any moment $\pi_{i,t}$ is a linear function of her prior beliefs $\pi_{i,t-1}$ and her current payoff. That is: when the ith arm is chosen, adaptive learning takes the form

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

where $\theta \in (0,1)$ captures the learning parameter, which reflects the speed of learning. We use a linear learning rule and set $\theta = 0.5$ in the simulation.[6] The term U_{feedback}, which represents utility value based on received feedback, is set equal to 1 if the payoff is above the agent's aspiration or to -1 if below that aspiration – in other words, an agent categorizes feedback relative to her aspiration level.[7] To ensure that the simulation results are not biased by our assignment of an arbitrary value for $\pi_{i,0}$, we assume that the agents will pull each arm once before the start of the simulation; then their feedback (utility) can serve as $\pi_{i,0}$. Thereafter, agents choose the arm with their subjectively highest expected payoff .

In line with our thought experiment, we set the high aspiration level equal to the best arm's mean payoff (1.0) and the moderate aspiration level equal to the average payoff (0.5) of all three arms. Then the middle arm's payoff is distributed around the moderate aspiration and is symmetric to the best arm, which itself is distributed around the high aspiration (as discussed in the thought experiment). The worst arm provides feedback that is unambiguously bad for an agent with high aspirations but that is ambiguous for agents of moderate aspiration levels. Hence, this setup replicates the differential ambiguity across arms discussed previously.

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

The simulation lasts 50 periods. For each aspiration level, we simulate 500,000 agents. Figure 2 plots the likelihood of choosing each arm over time. A high aspiration level leads to more exploitation of the best arm. Even when a high-aspiration agent does not choose the best arm, she is less likely to explore the worst arm – a finding that indicates stronger avoidance of broad exploration. We have argued that this

---

[6] We test the simulation model with different values of the learning parameter θ. The results are qualitatively consistent for $\theta > 0.25$. If θ is too small then, since adaptive learning is absent, choices do not evolve and there is no significant difference between DM1 and DM2.

[7] In reality, individuals might not strictly dichotomize performance feedback as success and failures. For example, performance 10 points below aspirations could be viewed as a much worse failure than performance only 1 point below aspirations. Our assumption of strict dichotomizing can be relaxed by following utility function: $U(x)$ is equal either to x^a (if $x \geq 0$) or to $-(-x)^a$ (if $x < 0$). The power term a $(0 < a < 1)$ reflects the certainty of categorizing a choice as success or failure, and x denotes the performance feedback relative to aspiration level. When $a \rightarrow 0$, feedback is strictly categorized as successes and failures. The simulation results are qualitatively robust regardless of the value assigned to a.

result follows from the high-aspiration agent's relatively stronger opinion about the worst arm's unsuitability (in comparison with the middle arm). Thus we hypothesize:

Hypothesis 1a (H1a)***.*** *Agents with high aspiration levels choose the* best *arm* more *often than do agents with moderate aspiration levels.*

Hypothesis 1b (H1b)***.*** *Agents with high aspiration levels choose the* worst *arm* less *often than do agents with moderate aspiration levels.*

Hypothesis 1a implies that, contra the extant literature, high aspiration agents explore less often than agents with moderate aspiration levels. In addition, Hypothesis 1b implies that high-aspiration agents explore more narrowly than agents with moderate aspiration levels. Since high-aspiration agents exploit more and explore less, they are likely to perform better at tasks (such as this one) for which the arms' payoffs (means and distributions) remain constant. We express this notion formally as follows.

Hypothesis 1c (H1c)***.*** *If the task environment is stable, then agents with high aspiration levels exhibit higher cumulative performance than do agents with moderate aspiration levels.*

To demonstrate our thought experiment's intuition that high aspiration levels reduce feedback ambiguity, we examine the standard deviation of the *difference* in the beliefs of (i.e., in feedback as interpreted by) high- versus moderate-aspiration agents. Figure 3 illustrates this difference for every combination of arms; note that we compare within, not across panels. Recall that the lower the standard deviation, the higher the opinion strength. Overall, a high aspiration level leads to a lower standard deviation in the beliefs about arms' payoffs and to less ambiguity about which arm is better. Thus agents with high aspirations pick the best arm across each pair with greater certainty and hence are less likely to explore other arms or to explore widely. In particular, Figures 3(a) and 3(b) show that agents with a high aspiration level are more certain that the best arm is superior to both the middle arm and the worse arm. This dynamic explains why, in Figure 2 and Hypothesis 1a, high-aspiration agents are more likely to exploit the best arm; it also underscores that agents with a high aspiration level sample the worst arm less often, which further increases their certainty about the best arm's payoffs.

Finally, it is clear from Figure3(c) that high-aspiration agents are more certain (than agents with moderate aspirations) about the middle arm being superior to the worse arm. This result accounts for the observation that, when an agent with high aspirations must choose between the middle and worse arms, she tends to avoid the worse arm and explore the middle arm – thereby reducing the breadth of her exploration (as proposed in Hypothesis 1b).

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

In a stable task environment, where the payoff distributions are not changing, exploitation leads to better performance regardless of aspiration level. High aspiration levels facilitate a strategic focus on exploitation by reducing ambiguity in the feedback from alternatives, which leads to high performance sooner and also to higher cumulative performance. Yet when the task environment is changing, a strategy that focuses on exploitation hinders adaptation to new payoff structures. When agents use adaptive learning to form their beliefs about the relative attractiveness of choices, 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 change. Since higher aspirations can lead to stronger belief formation, it follows that agents with higher aspirations should adapt less quickly to the new environment following a disruptive shock.

We undertake another simulation to assess this intuition. Expanding the previous setup, we introduce a disruptive shock – after period 30 – that changes the mean payoff of arms 1, 2, and 3 from [1.0, 0.5, 0] to [0.5, 0, 1.0]. That is, the worst arm becomes the best arm and the other arms shift down in value.[8]

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

We find that, although the strong beliefs developed by a high-aspiration agent serve him well in a stable environment, those beliefs prevent him from adapting to a *changed* payoff structure in a timely

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

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 confirms that, before the shock, agents with high aspirations exploit the best arm and avoid exploring the worst arm (and so engage in a narrower exploration). However, this strong preference for the best arm *before* a shock delays the agent's identification of the best arm *after* a shock. Our simulation motivates the following hypotheses.

Hypothesis 2a (H2a)***.*** *After a disruptive shock, agents with high aspiration levels choose the previous (i.e., pre-shock)* best *arm* more *often than do agents with moderate aspiration levels.*

Hypothesis 2b (H2b)***.*** *After a disruptive shock, agents with high aspiration levels choose the previous (i.e., pre-shock)* worst *arm* less *often than do agents with moderate aspiration levels.*

There are clear performance implications of the high-aspiration agents' slower adaptation to the changed payoff, as formalized in our next hypothesis.

Hypothesis 2c (H2c)***.*** *After a disruptive shock, agents with high aspiration levels exhibit lower performance than do agents with moderate aspiration levels.*

Thus we have used this agent-based simulation of a multi-choice bandit task in a changing environment to refine the insights derived from our previous thought experiment. Perhaps more importantly, the model itself forced us to make our assumptions about how aspiration levels influence learning from feedback explicit. Doing so enables direct translation into the lab setting, to which we turn next.

METHODOLOGY

In order to test the hypotheses, we design a behavioral lab experiment in which we manipulate the aspiration level and task environment and then observe participants' exploration and exploitation choices as well as their performance. Exploration–exploration decisions are inherently behavioral, since prior probabilities are unknown and must be learned in a trial-and-error process (Simon, 1947). Furthermore, since choice and learning processes are intertwined in these problems (March, 1991; Sutton and Barto, 1998), it follows that experimental studies can add considerable value by controlling the information available (Edmonds, 2001; Schunk, 2009; Sterman, 1989).

Experimental Procedure

In the studies that we shall describe, subjects played a multi-armed bandit game and were randomly assigned to the treatment (high aspiration level) or control (moderate aspiration level) conditions. The multi-armed bandit task is extensively used to study exploration–exploitation behavior in formal (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 studies (Cohen et al., 2007; Daw et al., 2006; Gans, Knox, and Croson, 2007; Knox, Otto, Stone, and Love, 2012). A multi-armed bandit task consists of multiple options, or arms. Each option has an uncertain payoff. The decision maker chooses an option in each time period and receives a payoff, where the arms' respective payoff distributions are unknown ex-ante to the individual. The decision maker's task is to maximize her end-of-game payoffs by choosing an option in each round; that choice is based on her learning from prior feedback. In order to test our research question, we ran four studies within the same broad framework.

Description of the Task. To reduce the task's abstract nature, we framed it as an investment choice. Each participant adopts the role of a R&D manager for a digital business firm and, in each round, decides in which product platform to invest. Participants are informed that the payoffs are uncertain and subject to market turbulence. Also, participants are aware that environmental shocks could alter the relative attractiveness of the arms but do not know when such shocks will occur.

In each period, participants select one out of the three available choices (arms) for investment. Upon making an investment, a participant immediately receives a payoff in virtual dollars. Participants know that they will play for 50 periods and that their payoff accumulates over these rounds. We chose three arms because that is minimum number that allows us to assess whether the treatment subjects explore more than the control subjects *and* whether the former explore more broadly. When there are only two arms, exploring narrowly or broadly result in the same choice. Thus a bandit with three arms allows us to investigate our research question in greater depth.

In some of our studies, the arms are subject to a shock (after round 30) that reorders the arms' relative attractiveness. It is the adjustments that participants make to these shocks that help us understand their exploration behavior: the more a subject explores, the more likely he is to shift choices in response to an abrupt shock. We explore two kinds of shocks: disruptive (Studies 2 and 3) and positive (Study 4). Table 2 summarizes the study conditions.

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

Treatment Manipulation. Participants are randomly assigned to treatment (high aspiration level) and control (moderate aspiration level) conditions. In both cases, subjects are informed that they should attempt to achieve a particular performance goal. We implemented the manipulation for high (resp. moderate) aspiration levels by instructing the subjects as follows: "The previous manager achieved total earnings of $1,250 ($1,000) over 50 rounds, or on average $25 ($20) per round. You are going to play the game for 50 rounds. You should aim to earn more than this amount." We followed the goal-setting literature in using "mere goals" – in other words, those that simply establish a reference point (Heath et al., 1999; Larrick et al., 2009) – and then observe whether such a simple manipulation results in behavioral differences.

The high aspiration level is 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. We selected this value (in the upper 10th percentile) in line with the goal-setting theory for establishing goals that are challenging (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. Each participant can see the previous round's payoff on her computer screen, at any time, and also her accumulated payoff.

Subject Recruitment. For Studies 1 and 3, participants were recruited through Amazon's Mechanical Turk (MTurk). For Study 2, we 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. In Studies 2 and

4, subjects were rewarded with course participation credits. We incentivized the participants to perform well by offering a $10 gift certificate for each of the top five performers. For Studies 1 and 3 we replaced course credit with a fixed payment for participation but otherwise maintained the same performance-based incentives. Because adaptive learning and search behavior are fundamental human behaviors, we decided to run our laboratory studies in different countries so as to increase (albeit just slightly) the external validity of our findings. Table 3 reviews our data collection and the participants' demographics.

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

Measures

Our study's objective is to explain how exploration–exploitation behavior changes with different aspiration levels. We manipulate aspiration levels by setting aspirations as described previously, and we measure exploration–exploitation behavior in several ways. First, at the subject level, we measure participants' propensity to choose the best arm (a proxy for exploitation) versus the worst arm (a proxy for exploration). Next we evaluate to what extent the subject explores narrowly (vs. broadly); here our measure is the pre-shock ratio of how often they choose the worst arm (arm 3) to how often they explore either inferior arm (arm 2 or arm 3). Finally, we measure opinion strength as the average difference in payoffs between the arms.

In the round-level analyses, we follow suggestions in prior work (e.g., Billinger et al., 2013; Levine et al., 2019) and control for such feedback variables as round number, prior round performance, prior average achieved so far, and time taken to make a choice. We also control for participant demographics such as age, gender, college major, and risk-taking propensity, where the latter is measured using the balloon analogue risk task (BART; see Lejuez et al., 2002).

RESULTS

This section reports the results from each of our four studies. For each one, we briefly summarize the study condition and then document our findings. We report analysis of variation (ANOVA) results for subject-level comparisons and run logit regressions to analyze subject–round data.

Study 1: Stable Environment

In Study 1, participants play a three-arm bandit; the payoffs are uniformly distributed, with means [25, 20, 15] and a constant interval of +/-10. The high (resp. moderate) aspiration level is set at 25 (resp. 20) per round. 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. We predicted that the high-aspiration decision makers will exploit more and that, when they explore, they will tend to avoid the worst arm and therefore explore narrowly. We test these hypotheses via analysis of variance.

We find that participants in the high-aspiration condition exploit the best arm significantly more often (mean $M = 29.1$, standard deviation SD = 5.4) than those facing the moderate aspiration level ($M = 26.3$, SD = 4.3; difference $d = 2.8$, $F = 15.18$, $p < 0.001$), outcomes that support Hypothesis 1a. Participants assigned to the high-aspiration level explore the worst arm significantly less often ($M = 7.9$, SD = 2.4) than do those at the moderate-aspiration level ($M = 9.9$, SD = 2.7; $d = 2.0$, $F = 27.68$, $p < 0.001$), which provides support for Hypothesis 1b. We find also that, when subjects explore (i.e., choose arm 2 or arm 3), the breadth of exploration – that is, the rate of choosing the worst arm (arm 3) – is significantly lower for participants in the high-aspiration condition ($M = 0.39$, SD = 0.10) than for those with moderate aspiration levels ($M = 0.42$, SD = 0.10; $d = 0.02$, $F = 4.52$, $p < 0.05$). Finally, cumulative performance is significantly higher in the high-aspiration condition ($M = 1108.5$, SD = 54.8) than in the moderate-aspiration one ($M = 1083.1$, SD = 57.8; $d = 25.4$, $F = 9.84$, $p < 0.01$), in support of Hypothesis 1c.

To illustrate the mechanism underlying these findings, we plot the histogram of the success rate of each arm as experienced by the participants in the prior period in figure 5. These histograms show that high aspiration levels lead to an unambiguous failure feedback for the worst arm (see panel 5c), whereas the best and middle arm provide more ambiguous feedback (panels 5a and 5b). To clarify further, we show the histogram of the success rate difference between each pair of arms in Figure 6. Comparing panels 6a and 6c, we see that the subject in the high aspiration condition is much more likely to infer that the worst arm (arm 3) is inferior to the other two arms (see test statstics in the Figure).

[[INSERT **Figures 5 & 6** ABOUT HERE]]

We exploit the longitudinal nature of our experiment to examine in more detail the mechanism underlying these results. For this purpose, we create a data set at the subject–round level so that we can analyze how prior feedback affects decisions in every round. Given the dynamic nature of this analysis, we use a logit model to predict subjects' exploitation of the best arm in each round. In Table 4, Model [1] replicates the ANOVA effects across periods. As expected, we notice that subjects exploit more in later rounds (an indication of learning) and when they have achieved high outcomes in previous rounds. We also see that subjects' probability of exploiting decreases if they show higher risk-taking propensity (as measured by the BART score). Model [2] confirms that high-aspiration participants exploit the best arm more often than do their moderate-aspiration counterparts.

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

Models [3]–[6] in Table 4 shed more light on the mechanism involved. In particular, Model [3] shows that the focal participant's experience with different arms has a positive effect on the likelihood of exploiting the best arm; this result is reflected in the average feedback difference between the best arm and each of the other arms, where the positive effect indicates that reducing feedback ambiguity facilitates exploitation of the best arm.[9] In sum: the greater is the difference in payoffs from prior feedback about the arms' relative attractiveness, the more likely is the subject to exploit the best arm.

Model [4] reveals a positive interaction effect between high aspiration level and the average feedback difference between the best and the worst arm. We observe that incorporating the 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 and thereby increase exploitation of the best arm. According to Model [5] in the table, that interaction effect is non-significant for the difference in feedback between the middle and the best arm. This result accords with our theory (though we cannot test for "no effect"). Model [6] establishes that these findings are robust when the regression includes all variables and interaction effects.

[9] Note that in the regressions, we measure the feedback as the actual payoff received by the participants rather than success or failure interpreted around aspiration, since the success or failure experience is highly correlated with the treatment.

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

Overall, the baseline results in Study 1 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 a greater difference in feedback is experienced between the best and the worst arm, participants can more rapidly hone in on exploiting the best arm; the effect is stronger for participants with high aspirations. We conclude that, by reducing the ambiguity of feedback on payoffs, high (as compared with moderate) aspiration levels may improve cumulative performance in a stable environment.

Study 2: Unstable Environment with a Disruptive Shock

In this study we introduce a disruptive shock that reshapes the payoff landscape between period 30 and 31. Before the shock, the three arms returned uniformly distributed payoffs with means [25, 20, 15] (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 each) so that the overall 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 treatment condition (high aspiration level) exploit the best arm significantly more often (M = 15.8, SD = 5.1) than do those in the control condition (moderate aspiration level) (M = 14.2, SD = 4.5; d = 1.6, F = 5.94, p = 0.02). Furthermore, participants in the treatment condition explore the worst arm significantly less often (M = 5.7, SD = 3.2) than do those in the control condition (M = 6.8, SD = 3.0; d = 1.1, F = 6.00, p = 0.02). When the participants explore, the rate of exploring the worst arm is significantly lower in the treatment condition (M = 0.39, SD = 0.14) than in the control condition (M = 0.43, SD = 0.12; d =0.04, F =5.25, p = 0.02). Finally, cumulative performance is significantly higher in the treatment group (M = 651.8, SD = 49.1) than in the control group (M =633.6, SD = 44.3; d = 18.2, F = 7.6, p = 0.01). These findings are in line with those we reported for Study 1.

One implication of the adaptive learning model is that an environmental shock should have different effects on decision makers with high versus moderate aspirations. We find that after the shock, high-aspiration participants were more likely to continue choosing the previously best arm (*M* = 10.8, SD = 5.3) than were those with moderate aspirations (*M* = 8.1, SD = 4.3; *d* = 2.0, *F* = 15.6, $p < 0.001$); this finding supports Hypothesis 2a. Also as predicted, high-aspiration participants were less likely post-shock to choose the previously worst arm (*M* = 5.0, SD = 3.7) than were those with moderate aspirations (*M* = 7.5, SD = 4.1; *d* = 2.5, *F* = 20.3, $p < 0.001$) – supporting Hypothesis 2b. As a result, cumulative performance after the shock is 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; *d* = 10.1, *F* = 4.31, $p < 0.05$), which supports Hypothesis 2c.[10]

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

In Study 2 we observe the post-shock choices for only 20 periods (i.e., from period 31 to period 50). However, that duration may not be long enough to observe the subjects' adaptation to the new environment. We 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).

Prior to the disruptive shock, results for Study 3 replicate those for Studies 1 and 2 (see Table 5). 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 itself on learning behavior and choices. Expanding the post-shock period allowed us to observe two different behavioral patterns. More specifically, we compared the choices in the periods immediately after the shock (periods 31–50) with choices in periods long after the shock (periods 61–80).

[[INSERT **Table 5** ABOUT HERE]]

From period 31 to period 50, participants in the high-aspiration condition were more likely to continue choosing the previously best arm (*M* = 11.54, SD = 3.20) than were those with moderate

[10] We also ran this study with greater variance [+/-15] from the mean, which means that the worst arm for high aspiration condition returned a greater percentage of successes (15% instead of 5%). The results are qualitatively robust. Results omitted due to space constraints, available from the authors upon request.

aspirations (M = 7.19, SD = 2.42; d = 4.35, F = 42.57, p < 0.001) – even though that arm now yielded lower payoffs. This result lends support to Hypothesis 2a. During later periods (periods 60–80), however, participants with high aspirations had adjusted their behavior and so were now *more* likely to choose the previously worse but now best arm (M = 13.76, SD = 4.15) than were those with moderate aspirations (M = 10.11, SD = 2.69; d = 3.65, F = 19.74, p < 0.001); H1a is supported also by this finding. Thus decision makers adapted to the shifted payoff landscape, and the reduced ambiguity about the feedback on new payoffs again led high-aspiration subjects to choose the best arm with greater certainty. This result highlights the learning inherent in the effects that our studies uncover.

In line with our post-shock prediction (H2b) and with the Study 2 results, high-aspiration participants were less likely (in periods 31–50) to choose the previously worst arm (M = 4.08, SD = 2.96) than were those with moderate aspirations (M = 8.03, SD = 2.74; d = 3.95, F = 34.93, p < 0.001). This post-shock behavior leads to suboptimal outcomes because the environmental shock should rather encourage broad exploration to identify the new payoff distributions. However, participants in the high-aspiration condition took longer to learn the new payoff structure and also adjusted their performance more slowly, as compared with participants in the moderate-aspiration condition. Yet the former eventually benefited from the reduction in feedback ambiguity, which explains this study's replication of the pre-shock equilibrium effects outlined in H1a and H1b. Figure 8 illustrates these effects by plotting the probability of choosing the best arm before and after the shock as a function of the aspiration level.

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

Studies 1–3 explore the effects of aspiration levels on exploration–exploitation choices made in stable versus unstable environments. To establish the boundary conditions of our major findings, we also explore the effect of different types of environments. Thus Study 4 examines the impact of a positive (rather than a neutral) shock.

Study 4: Unstable environment with a disruptive, *positive* shock

In Study 4, we introduce a positive shock between period 30 and 31 that only changes the payoff of the worst arm. 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]. Similar to study 2, the worst arm before the shock becomes the best arm after, but unlike study 2, the payoff to 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 5.

In comparing Study 2 with Study 4, we expect that the latter's high-aspiration subjects are 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 of 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 high-aspiration subjects do choose arm 1 more frequently after the shock in Study 4 than in Study 2 (13.0 vs. 7.4; standard error S.E. = 0.72, $t = 2.88$, $p < 0.01$). It is noteworthy also that Study 4's high-aspiration subjects 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; S.E. = 0.411, $t = 4.13$, $p < 0.001$). In accordance with our expectations, the pre-shock choice patterns across these two studies do not exhibit any significant differences.[11]

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). In addition, prior research suggests that these activities are often conflicting (Benner and Tushman, 2003, 2015; March, 1991; Tushman and O'Reilly, 1996). Despite an explosion of scholarly research on how the exploration–exploitation trade-off affects choices and learning from experience, it remains unclear how that trade-off is influenced by different aspiration levels.

On the one hand, a substantial body of individual- and organization-level research focuses on the behavioral consequences of performance relative to aspirations and argues that increased aspirations generally lead to greater exploration (Greve, 2017; Shinkle, 2012). On the other hand, empirical work has

[11] We did not assign participants randomly between Study 2 and study 4, so this comparison simply looks at the sample averages across studies.

reported mixed results with respect to this proposition (Posen et al., 2018), and research on adaptive learning suggests that the relationship between increasing aspirations and exploration may be less straightforward than originally proposed (Denrell, 2008; Denrell and Le Mens, 2011). Denrell uses a simulation to show that adaptive learning can produce outcomes similar to predictions from the behavioral theory of the firm *without* any changes in preferences for risk. We combine this insight with the idea that aspiration levels form reference points that induce the subjective categorization of feedback, which has consequences for how subjects learn from feedback. Thus we contribute to the behavioral theory of the firm by combining the two major strands in that literature – adaptive learning and subjective categorization of feedback – to develop a better understanding of the behavioral consequences of setting aspiration levels. It is interesting that empirical work seldom adopts a learning perspective when predicting the exploration–exploitation consequences of having high aspiration levels. At the same time, most of the work that does adopt a learning framework is of a theoretical nature. It is this gap that our paper seeks to bridge.

We draw on elements from both frameworks to develop a model of how differential aspiration levels affect the exploration–exploitation trade-off. We test the hypotheses developed from this model in a laboratory experiment using a multi-arm bandit task, which is the canonical model used for understanding how agents approach exploration and exploitation (Daw et al., 2006; Gittins, 1979). Drawing from behavioral and reference point theories, we argue that aspiration levels are likely to influence how subjects interpret feedback (i.e., as success versus failure). Consistent with adaptive learning theory, we then suggest that this interpretation of feedback affects subsequent sampling behavior and thus exacerbates any preference related to the different choices. Using experimental data, we show that a high aspiration level reduces feedback ambiguity about the relative attractiveness of available choices and leads to (a) more exploitation and (b) narrower exploration. We thus contribute to the literature by showing how aspiration levels and adaptive learning *jointly* influence the choice between exploration and exploitation.

Our theory is especially applicable to situations 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 so must be learned from repeated feedback. Consider, for instance, an R&D manager in a pharmaceutical company 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 observing the outcomes. We 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.[12]

When we consider the impact of aspiration levels in an adaptive learning context, where we 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, then we find that increased aspiration levels actually lead to more exploitation and that the agents who do explore consider a narrower set of choices. The implication is that although 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 organization to try out new ideas, yet the capital budgeting process may hinder that exploration by dictating safer alternatives. So in order for goal setting or aspiration levels to truly unleash exploration, the goals must be sufficiently ambitious that they trigger search for new alternatives – that is, rather than choice between existing ones. Nevertheless, both Sitkin et al. (2011) and Gary et al. (2017) point out that each of these responses comes with its own drawbacks.

Overall, our findings are counterintuitive when one considers the prevailing view in research on performance relative to aspirations and goal setting (Greve and Gaba, 2017; Ordóñez et al., 2009). That our results run counter to previous work reflects our *joint* study of the choice process (which, as usual, we examine in the context of aspiration levels) and the learning process. This broader approach allowed us to explore how aspiration levels may affect not just the explore versus exploit decision but also the

[12] 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.

interpretation of payoff feedback as either success or failure. This subjective interpretation influences 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 the agent-based simulation described here allowed us to make concrete predictions about this interplay between aspiration levels and adaptive learning about the exploration–exploitation trade-off, we believe that this paper has but scratched the surface of a promising avenue of enquiry.

REFERENCES

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)*.

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

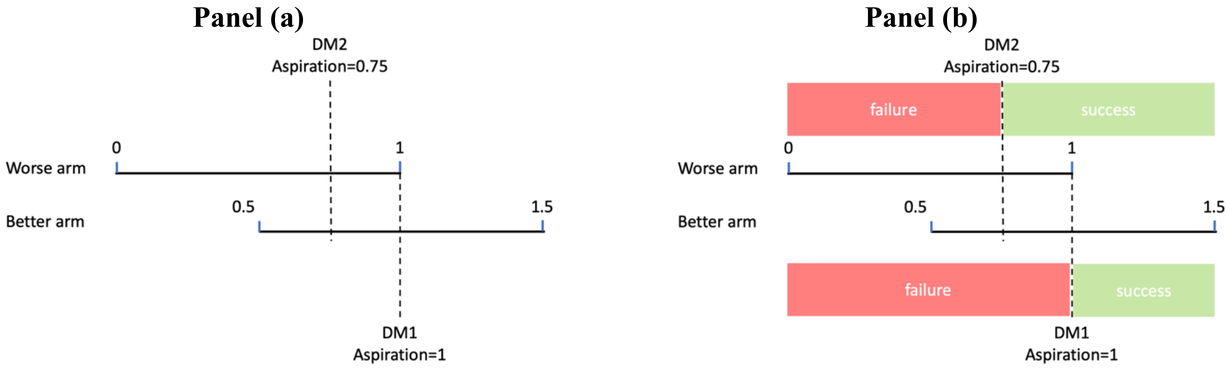
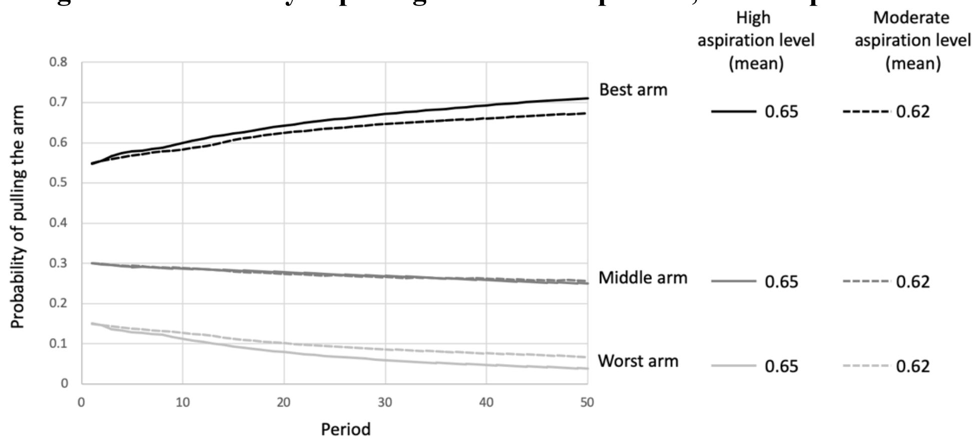
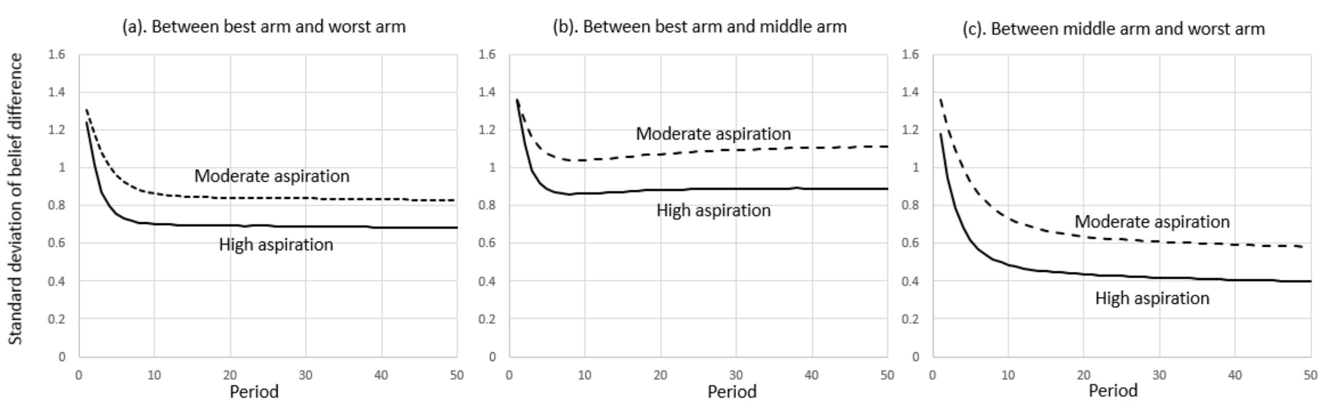


Figure 2. Probability of pulling the arms: 50 periods, no disruptive shock



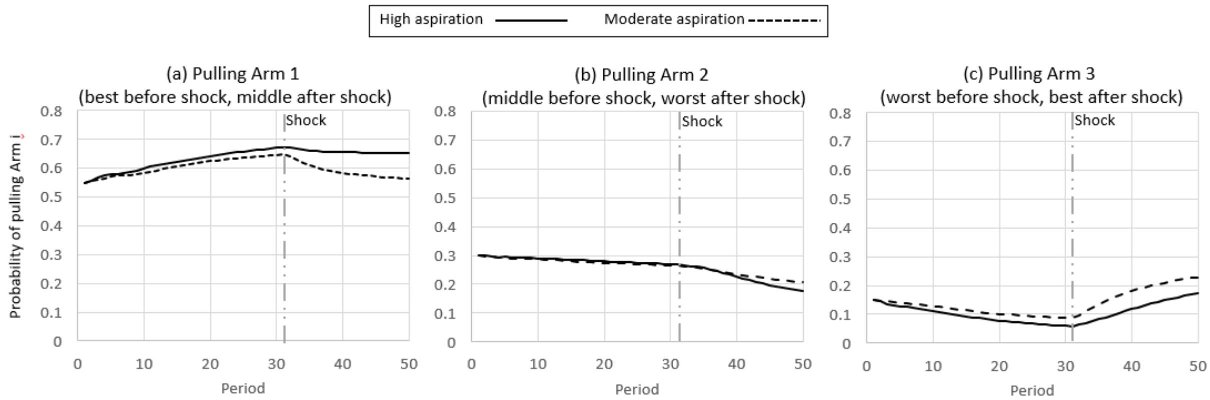
High aspiration levels lead to more exploitation of the best arm. Also, when agents with high aspiration levels explore, they tend to avoid exploring the worst arm.

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



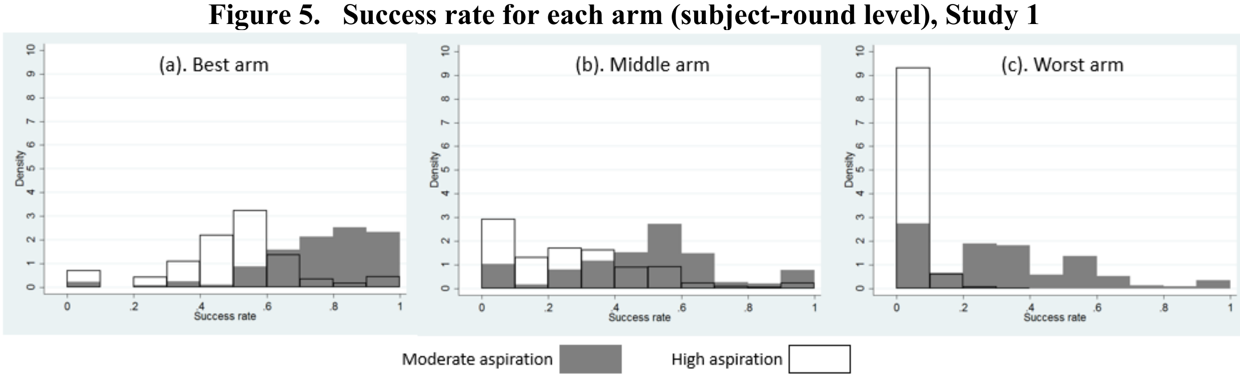
The smaller the standard deviation of belief difference between two arms, the less the agent's feedback ambiguity that one arm is better than the other. High aspiration levels always lead to a lower standard deviation of these differences between arms. Hence high-aspiration agents experience less ambiguity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

Figure 4. Probability of pulling the arms: Disruptive shock after period 30



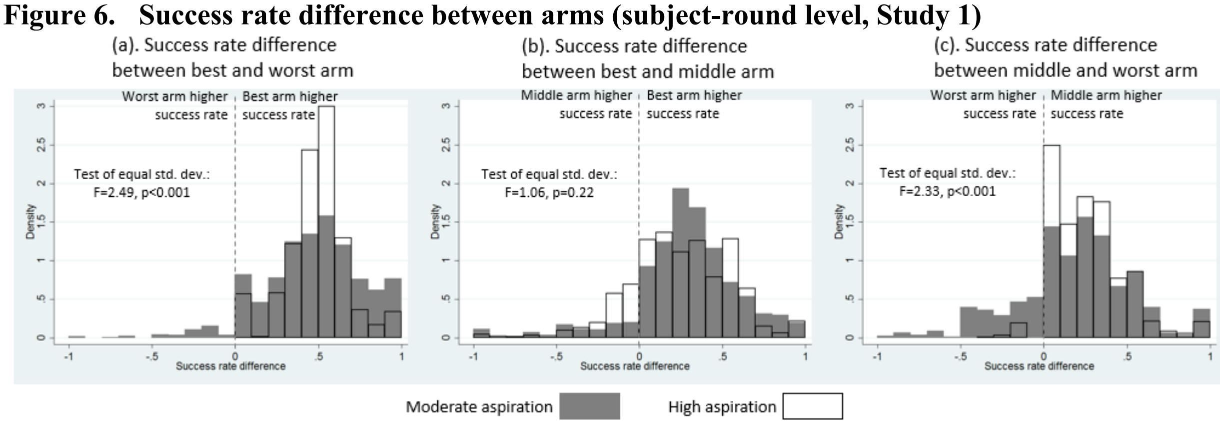
A high aspiration level leads to a strong preference for the best arm with little or no exploration of inferior alternatives. When a disruptive shock alters the payoff structure, this strong preference makes it difficult for high-aspiration agents to identify quickly the new, superior alternative.

Figure 5. Success rate for each arm (subject-round level), Study 1



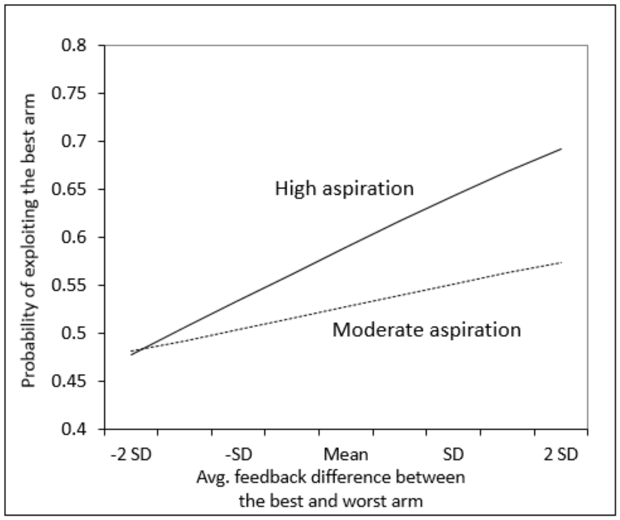
At every round, for each subject, the figure shows the distribution of success rates in the prior round across the three arms. For example, if a participant had 2 experiences above aspiration and 3 experiences below aspiration from pulling an arm prior to the focal period, then she has experienced a 0.4 success rate for this arm. *A high aspiration leads to consistent feedback of failure (panel c).*

Figure 6. Success rate difference between arms (subject-round level, Study 1)



In each panel we subtract the former arm's success rate from the latter arm's success rate. If the difference is above zero, then the former arm is better than the latter arm. As shown in panels (a) and (c), the moderate aspiration level leads to more ambiguous feedback about the worst arm in comparison to the best and middle arm (higher variance). In contrast, the high aspiration level leads to consistent feedback that the best and the middle arm are better than the worst arm (lower variance and consistently positive difference).

Figure 7. Average feedback difference and probability of exploiting the best arm as a function of aspiration level (coefficients from Table 4, Model [6])



For the same increase in actual feedback difference between the best and the worst arm, participants with high aspiration level show a greater increase in their probability of exploiting the best arm.

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

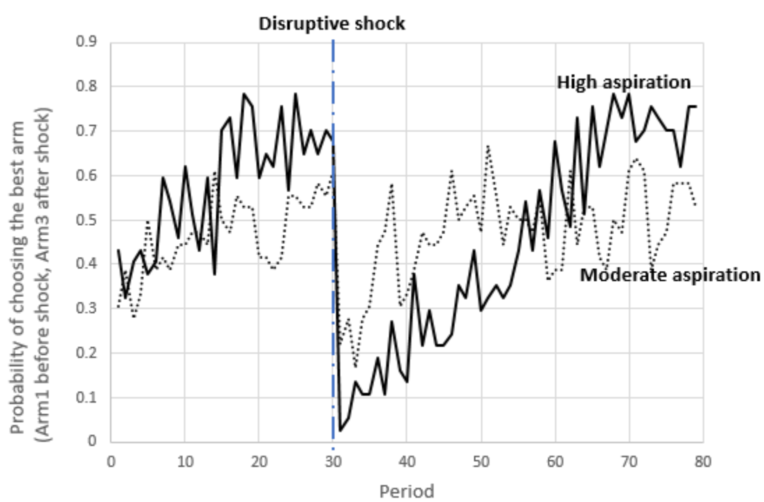


Table 1. Probability of choosing the worse arm by previous number of trials

	After number of trials on each arm					
	1	2	3	4	5	6
High aspiration (DM1)	25.0%	12.5%	6.3%	3.2%	1.6%	0.8%
Moderate aspiration (DM2)	25.0%	15.6%	10.3%	7.1%	4.9%	3.4%

The decision maker samples each arm for a certain number of trials and then chooses the arm that has returned more successful positive payoffs. Here we illustrate the likelihood of re-exploring the worse arm after 1, 2, …, 6 trials on each arm. For example, after two trials on each arm, the decision maker with high (resp. moderate) aspirations has a 12.5% (resp. 15.6%) chance of choosing the worse arm in the fifth pull.

Table 2. Study conditions

Study	Arm payoff before shock			Shock Period	Arm payoff after shock			Interval *Constant across all arms*
	Arm 1	*Arm 2*	*Arm 3*		*Arm 1*	*Arm 2*	*Arm 3*	
1	$25	$20	$15	None	*Same as before*			[±10]
2	$25	$20	$15	30	$20	$15	$25	[±10]
3	$25	$20	$15	30	$20	$15	$25	[±10]
					30 additional periods after shock			
4	$25	$20	$15	30	$25	$20	$30	[±10]

All arms are drawn from a uniform distribution, means and intervals displayed in the table. In studies 2 and 3, a disruptive shock occurred after period 30. There was no shock in study 1. Study 3 had 50 periods after the shock (compared to 20 for studies 2, and 4). Study 4 had a positive shock after period 30, where only arm 3's payoff increased and the other arms remained unchanged.

Table 3. Data collection

Study	Study condition	Total number of participants	Number of participants in treatment condition	Location / Source	Mean age [range]	Female (%)
1	Stable environment: no shock	193 (220)	98	MTurk (international)	35.0 [19, 73]	33.2
2	Unstable environment, disruptive shock: replicated Study 1 *with* shock	200 (220)	103	US university	20.8 [19, 31]	54.5
3	Unstable environment, disruptive shock and extended post-shock period: replicated Study 2 but with an additional 30 rounds post-shock	73 (80)	37	MTurk (international)	33.6 [20, 65]	30.2
4	Unstable environment, disruptive *positive* shock: shifting average payoffs upward (compared to Study 2)	77 (90)	39	Singapore university	20.7 [19, 24]	45.5

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 in all parts of the questionnaire. When possible, we performed analyses on each sample (both the full and incomplete ones); no significant differences were observed.

Table 4. Study 1: Logit regression on exploiting the best arm

		DV: Exploiting the best arm (binary variable)				
VARIABLES Model:	[1]	[2]	[3]	[4]	[5]	[6]
Treatment (high aspiration)		0.200***		-0.035	0.179**	-0.045
		(0.055)		(0.114)	(0.083)	(0.117)
Avg. payoff difference between the best and worst arm			0.035***	0.020***	0.038***	0.022**
			(0.006)	(0.008)	(0.006)	(0.009)
Treatment * Avg. payoff difference between the best and worst arm				0.033***		0.031***
				(0.011)		(0.012)
Avg. payoff difference between the best and middle arm			0.061***	0.066***	0.053***	0.063***
			(0.006)	(0.007)	(0.009)	(0.010)
Treatment * Avg. payoff difference between the best arm and middle arm					0.020	0.006
					(0.012)	(0.013)
Previous round performance	0.014***	0.014***	0.012***	0.011***	0.012***	0.011***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Average cumulative performance	0.023**	0.022**	0.004	0.003	0.002	0.003
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Round	0.019***	0.019***	0.017***	0.017***	0.017***	0.017***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Time spend	-0.013	-0.013	-0.008	-0.007	-0.008	-0.007
	(0.009)	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)
Age	0.001	0.002	0.001	0.002	0.002	0.002
	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
Gender (male)	-0.067	-0.068	0.005	0.015	0.016	0.018
	(0.060)	(0.057)	(0.056)	(0.050)	(0.052)	(0.051)
BART score	-0.031***	-0.029***	-0.026***	-0.021***	-0.023***	-0.021***
	(0.006)	(0.006)	(0.006)	(0.005)	(0.006)	(0.005)
Constant	0.173	-0.002	-0.195	-0.436	-0.424	-0.430
	(0.332)	(0.330)	(0.340)	(0.325)	(0.329)	(0.326)
Log-likelihood	-6361.9	-6350.7	-6250.0	-6224.0	-6227.4	-6223.9
Clustered subjects	193	193	193	193	193	193
Observations	9,457	9,457	9,457	9,457	9,457	9,457

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

Table 5. ANOVA Results for Studies 3 and 4

Hypothesis	Findings (compared to moderate aspiration level condition	Study 3 – extended post-shock			Study 4 – positive payoff shock		
		Aspiration level		Difference between conditions	Aspiration level		Difference between conditions
Pre-shock results		Moderate	High		Moderate	High	
H1a	High aspiration level subjects exploit the best arm significantly more often	14.22 (3.00)	17.51 (4.79)	d=3.29 F=12.32 p<0.001	13.00 (4.07)	16.38 (5.78)	d=3.38 F=8.78 p<0.01
H1b	High aspiration level subjects explore the worst arm significantly less	5.94 (2.32)	4.14 (1.78)	d=1.81 F=12.32 p<0.001	7.95 (3.08)	5.85 (3.01)	d=2.10 F=9.17 p<0.01
	When high aspiration level subjects explore, their percentage of exploration is lower	0.40 (0.13)	0.35 (0.13)	d=0.05 F=1.30 p<0.01	0.46 (0.12)	0.42 (0.12)	d=0.04 F=2.92 p<0.1
H1c	High aspiration level subjects have higher performance before the shock	640.17 (37.18)	668.84 (43.79)	d=28.67 F=9.07 p<0.01	632.50 (44.71)	655.56 (52.54)	d=23.06 F=4.29 p<0.05
Post-shock results		**Moderate**	**High**	**Difference**	**Moderate**	**High**	**Difference**
		rounds 31-50			***rounds 31-50***		
H2a	High aspiration level subjects are more likely to continue choosing the previously best arm	7.19 (2.42)	11.54 (3.20)	d=4.35 F=42.57 p<0.001	7.42 (4.84)	12.95 (5.67)	d=5.53 F=21.10 p<0.001
H2b	High aspiration level subjects are less likely to choose the previously worst arm	8.03 (2.74)	4.08 (2.96)	d=3.95 F=34.93 p<0.001	8.05 (4.42)	3.28 (3.54)	d=4.78 F=27.43 p<0.001
H2c	High aspiration level subjects have lower performance post-shock	418.03 (29.96)	399.38 (29.58)	d=18.65 F=7.16 p<0.01	512.95 (34.53)	495.67 (33.73)	d=17.28 F=4.93 p<0.05
		rounds 61-80					
H1a	High aspiration level subjects exploit the (new) best arm, arm3, significantly more often	10.11 (2.69)	13.76 (4.15)	d=3.65 F=19.74 p<0.001			
H1b	High aspiration level subjects explore the (new) worst arm, arm2, significantly less	3.53 (1.76)	1.22 (1.94)	d=2.31 F=39.96 p<0.001			

The table reports sample means, standard errors in parentheses.