

MOTIVATING INNOVATION: THE EFFECT OF LOSS AVERSION ON THE WILLINGNESS TO PERSIST

Yaroslav Rosokha and Kenneth Younge*

Abstract—We investigate the willingness of individuals to persist at exploration when confronted by prolonged periods of negative feedback. We design a two-dimensional maze game and run a series of randomized experiments with human subjects in the game. Our results suggest individuals explore more when they are reminded of the incremental cost of their actions, a result that extends prior research on loss aversion and prospect theory to environments characterized by model uncertainty. In addition, we run simulations based on a model of reinforcement learning that extend beyond two-period models of decision making to account for repeated behavior in longer-running, dynamic contexts.

I. Introduction

Universal foreknowledge would leave no place for an “entrepreneur.” His role is to improve knowledge, especially foresight, and bear the incidence of its limitations.

—F. H. Knight (1921)

WHAT motivates innovation? A growing literature views this question as a problem of optimally structured incentives to explore new prospects (Che & Gale, 2003; Manso, 2011; Morgan & Sisak, 2016), and recent empirical work has studied the causal effect on innovation of government subsidies (Howell, 2017), ownership (Guadalupe, Kuzmina, & Thomas, 2012; Seru, 2014), bankruptcy laws (Acharya & Subramanian, 2009), career concerns (Aghion, Van Reenen, & Zingales, 2013), and wrongful discharge laws (Acharya, Baghai, & Subramanian, 2014). Much of the empirical research on innovation has been at the organizational level, for broad samples of individuals have been hard to obtain and mechanisms hard to identify (for a notable exception, see Azoulay, Graff Zivin, & Manso, 2011).

While incentives clearly are an important motivator to engage in innovation (Lerner & Wulf, 2007), behavioral research suggests that they do not always shape individual behavior as one might expect (Gneezy, Meier, & Rey-Biel, 2011). In this study, we draw on methods from experimental economics to examine how the structure of incentives may affect a decision by an individual to pursue (or avoid) an unproven path to an uncertain payoff. We construct an experiment to induce an uncertain environment where exploration

is possible and focus on loss aversion as our mechanism of interest. Specifically, we test how incentive structures that generate loss aversion affect an individual’s willingness to persist at an exploratory task. We find that individuals vary considerably in their willingness to explore an uncertain prospect and that treatments that give rise to loss aversion increase attempts at exploration.

Our results contribute to a behavioral perspective of what motivates innovation, with implications for how organizations might structure incentives to motivate persistence when attempting a breakthrough innovation. For example, the literature on tolerance for failure (Azoulay et al., 2011; Manso, 2011; Tian & Wang, 2014) has emphasized the need for organizations to tolerate early failure and reward long-term success. Our behavioral findings, however, suggest that an optimal incentive structure would induce innovators, at the individual level, to experience the potential benefits of innovation in tandem with the urgency of ongoing losses during exploration.

We contribute to an emerging literature that uses experiments with human subjects to study innovation. For example, Ederer and Manso (2013) test how pay for performance affects effort to explore; Buchanan and Wilson (2014) investigate how intellectual property protection encourages innovations in the market; Herz, Schunk, and Zehnder (2014) study overoptimism and overconfidence and show that while overoptimism is positively associated with innovation, overconfidence is negatively associated with innovation; Elfenbein, Knott, and Croson (2016) test how an equity stake affects the optimal timing of exit from a losing proposition; and Kagan, Leider, and Lovejoy (2017) investigate how to divide limited time between design (exploration) and execution stages during new product development. In this study, we focus on one aspect of the innovation problem: the decision to explore an unproven path when there is no direct evidence from prior search about the likelihood of success. In such a context, individuals must rely on their own “foresight” to decide whether success is possible.

We build on Frank Knight’s (1921) proposition that foresight is a key aspect of innovation and entrepreneurship. It is well known that Knight distinguishes between risk (an unknown draw from a known probability distribution) and uncertainty (an unknown draw from an unknown probability distribution). What is less known is that the context for his distinction is an innovation problem. Specifically, Knight argues that innovations are uncertain (by definition) because the probability distribution is unknown by virtue of being new. Furthermore, he argues that when little is known about the true feasibility or returns of a prospect, then markets will require the entrepreneur to “warrant” a decision to innovate. The entrepreneur might act on his own (such as starting a

Received for publication August 7, 2017. Revision accepted for publication April 10, 2019. Editor: Shachar Kariv.

*Rosokha: Purdue University; Younge: École Polytechnique Fédérale de Lausanne.

This paper benefited from discussions with and comments from Saurabh Bansal, Tim Cason, and Stephen Leider, as well as workshop participants at the 2017 North American ESA, 2017 INFORMS Annual Meeting, 2018 Workshop on Experimental Economics and Entrepreneurship, and seminar participants at Imperial College London.

A supplemental appendix is available online at http://www.mitpressjournals.org/doi/suppl/10.1162/rest_a_00846.

new venture) or act within an organization (such as leading a new initiative for a firm), but in either case, the Knightian perspective implies that innovation requires an individual to first develop an intuition about how the environment might be and to then bear the consequences (positive or negative) of being right or wrong. The decision to innovate must be warranted by the individual innovator because the innovator cannot present probabilistic evidence about the likelihood of success and purchase insurance for the innovation outcome. When one can provide probabilistic evidence, the problem is then not one of innovation but, rather, one of risk management.

We therefore focus on an innovation problem that is somewhat different from the innovation problems presented by other studies in the literature on innovation. In particular, bandit models have been used to study learning under uncertainty as a type of innovation problem (Manso, 2011). In a bandit model, an actor learns about the efficacy of a prospect over time. In our framework, individuals may learn about the potential benefit of a prospect, but the benefit is conditional on getting the prospect to work. Getting the prospect to work, however, remains fundamentally uncertain because there is no positive experience from which to learn. Thus, our innovation problem focuses on model uncertainty (Hansen & Sargent, 2001; Lim et al., 2006), where several specifications are possible, and it is up to the individual to proceed based solely on intuition about the true underlying model—what Knight calls *foresight*.

One might think that our innovation problem is an extreme case. After all, innovators have strong incentives to try to learn about the efficacy of an innovation path as they explore it, and so it would be reasonable to expect innovators to accumulate probabilistic evidence as quickly as possible. For example, Nanda and Rhodes-Kropf (2016) present a model in which innovators stage investments specifically to learn more about the probability distribution of an innovation prospect prior to full commitment to the prospect. In their model, innovators (or investors, or both) learn from first-stage experiments in order to either invest more into continuing down that innovation path or to abandon course. We point out, however, that in the Nanda and Rhodes-Kropf (2016) model (and in the real world), there are periods during which action must be motivated by little or no evidence about the probability of the ultimate outcome (i.e., the first stage in their model). During such periods of blindness, when cash keeps flowing out the door and there is little to show for it, one or more key individuals must take on the responsibility to warrant the decision to continue pursuing the innovation. The most critical decision during this time is simply whether to keep on trying or to give up.

Given the nature of the innovation problem defined for this study, we examine how the structure of incentives for this problem may affect a decision by an individual to engage in (or avoid) exploration. In particular we consider one of the best-established biases in the field of behavioral economics: loss aversion. Prior research has shown that individuals tend

to interpret gains or losses in different ways (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). Specifically, they tend to interpret each gain or loss relative to a reference point (Kahneman, Knetsch, & Thaler, 1990; Barberis, Huang, & Santos, 2001), such that a loss affects a value function more than an equivalent gain (Tversky & Kahneman, 1991).

A brutal fact of innovation and entrepreneurship is that most prospects fail; moreover, a relatively small share of successes earn a majority of the rewards (Kerr, Nanda, & Rhodes-Kropf, 2014). In such situations, the willingness of individuals to bear the incidence of losses in the pursuit of rewards may be an important behavioral factor to what motivates innovation. However, to the best of our knowledge, there are no studies explicitly investigating loss aversion in the context of a Knightian innovation. While a related stream of literature on effort provision (Hannan, Hoffman, & Moser, 2005; Hossain & List, 2012; Armantier & Boly, 2015; Rubin, Samek, & Sheremeta, 2018) and team productivity (Dickinson, 2001; Hong, Hossain, & List, 2015) has found that loss aversion plays a significant role in determining an individual's decision to expend resources to achieve a goal, the extent to which loss aversion affects an individual's willingness to persist at innovation remains unexplored.

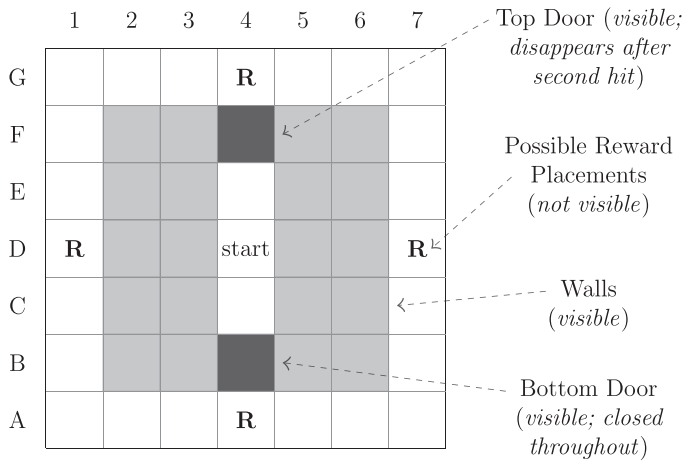
In summary, we investigate the intersection of two topics: innovation problems in which there is little (or no) evidence about the probability of success, and the behavioral tendency to avoid losses. As we will show in section III, the effect of loss aversion on the decision to explore is not obvious, especially when one considers the role of foresight in motivating the decision to explore. To examine this question, we designed an environment where we could manipulate incentives for exploration while independently testing conditions that promote or reduce loss aversion.

II. Experimental Design

In this section, we explain the design of our experiment. Given the unusual setup of our innovation problem, it is important to first explain how subjects are allowed to behave within the experiment and the instructions provided. In section III, we describe how we model loss aversion and develop predictions for our experiments, given the nature of the environment described here.

Our research question called for the development of an environment in which the structure, rules, and incentives would (a) lead each subject to develop a multiple-model view of the environment; (b) lead each subject to develop foresight into a superior solution to the game under one of those models; (c) avoid revealing information about whether the foresight would work (i.e., maintain uncertainty); (d) allow researchers to manipulate incentives to induce loss aversion without otherwise changing the expected earnings from the game; and (e) unfold over repeated trials so that researchers could test subjects' persistence at exploration. We were unable to find such an environment in the literature, so we designed a custom environment to meet our research needs—one that we call the

FIGURE 1.—MAZE SETUP



Subjects start the game at D4. They can navigate around empty cells, cannot backtrack, and cannot hit walls. Walls are light gray. Doors at B4 and F4 are dark blue (dark gray in the figure). Subjects can hit a door from either side. A subject who hits a closed door then jumps to the start position at D4. The top door at F4 is opened on the second hit and then disappears for the remainder of the experiment. The bottom door at B4 remains closed and is displayed as dark blue for the entire experiment. Potential reward placements are marked in the figure with an R (at G4, D1, D7, and A4), but neither the potential nor actual reward placements are displayed to subjects. A cell flashes green when a subject reaches a reward, and the subject then moves back to the start position at D4. Grid coordinates are not visible. Online appendix B provides step-by-step screenshots of a sample game.

Maze Game.¹ Although the game lacks verisimilitude as an example of innovation, it does reflect the underlying tension between an option to exploit and an option to explore, and the game allows us to carefully and independently vary each of the factors of interest we identified earlier.

A. Rules of the Environment

The Maze Game is played on a 7 × 7 grid, illustrated in figure 1. At the start of the game, participants are placed in the center cell (marked “start” in the figure) and given 500 moves to play. The game ends and subjects are paid only when they complete all 500 moves. Participants are instructed to move around the grid to discover and earn rewards, and the number of moves is updated and displayed with each move.

Four types of cells appear on the grid: empty cells (in white), walls (shaded in gray), doors (shaded in blue, but dark gray in the figure), and rewards (shaded in green, once discovered). While we describe gray cells as “walls” and blue cells as “doors,” we do not provide any such description or interpretation to subjects. Instead, the instructions for the game state only: “Your moves may be blocked and/or you may be forced to restart from a given position”; they are then left on their own to infer what they will about the environment (see section IIC and online appendix A for the additional details about the instructions used in the experiment). Subjects can

move through empty cells one move at a time. Subjects cannot move into walls or attempt to move into walls because the relevant button in that case is disabled. Subjects can move into (or “hit”) a door from any side, but immediately upon hitting a door, they jump back to the start position at the center of the board. The top door (at F4 in figure 1) will “open” and permanently disappear after the second hit. The bottom door (at B4 in figure 1) remains closed at all times throughout the experiment—that is, no matter how many times a subject hits the door, the door always flashes red and the subject always returns to the start position.² Online appendix B provides step-by-step screenshots for the opening moves of a sample game.

The spatial distribution and frequency of rewards are stochastic, and (unlike doors) the potential positions for rewards are not marked on the grid or shown to subjects. Instead, participants need to learn about both the positions in which rewards can appear and the relative frequency of rewards in those positions. When a participant discovers a reward, it is temporarily shown to her on the board (as a green square and a reward amount); the reward then is added to the game balance (displayed at the top of the game in dollars), and the participant returns to the start position at the center of the board. We refer to all of the steps taken from starting at the center of the board to finding a reward (or hitting a door) and returning to the center of the board as a *cycle*. At the start of each cycle, a new reward is randomly and invisibly positioned into one of the four locations on the grid (G4, D1, D7, or A4 in figure 1).

The game does not allow backtracking, which limits the number and types of strategies available to subjects and makes it possible for us to model the relative expected payoff for each strategy. Specifically, the no-backtracking rule works in conjunction with the structural layout of the maze to ensure that rewards are always discovered with 3, 9, 15, or 21 steps when a subject chooses to go through the top door.³ Because the layout of both the maze (figure 1) and the distribution of rewards (table 1) are left/right symmetric, the subject actually makes only one substantive choice at the start of each cycle: to go up from the start position or to go down. Once the top door is opened (after the second hit, and as shown in figure 2a), then the two options also map to stylized notions of exploitation (go up and find a reward with relative certainty) and exploration (go down in anticipation

²We considered using an alternative configuration in which the bottom door would open after perhaps five, ten, or fifteen attempts, but we determined that such a design would only truncate evidence about the extent to which subjects would be willing to persist at trying to open the bottom door. Given that the objective of our study is to test such persistence, we selected a configuration in which the bottom door always remains closed.

³A participant who chooses to go through the top door could choose to go clockwise or counterclockwise. In either case, however, the rewards are discovered with 3, 9, 15, or 21 steps because of the symmetric layout and the fact that the participant cannot backtrack. The layout also guarantees that regardless of clockwise or counterclockwise choice, the participant learns the same information about the position of rewards.

¹Designing an environment to vary incentives independent of loss aversion is a difficult problem. Although other researchers have identified the problem (see note 10 in Elfenbein et al., 2016—“We chose to shift the RL parameters by +130 to avoid problems associated with loss aversion”), our paper is the first study we know of to test changes in both the incentive to explore and conditions for loss aversion without changing the expected earnings from the game.

TABLE 1.—DESIGN PARAMETERS

a. Placement of Rewards			b. Payoff Structure		
Location	Baseline	Breakthrough	Parameters	Gains	Losses
Top (G4)	0.25	0.25	Starting balance	\$1.50	\$6.50
Left (D1)	0.25	0	Moves per game	500	500
Right (D7)	0.25	0	Cost per move	\$0.00	\$0.01
Bottom (A4)	0.25	0.75	Reward amount	\$0.06	\$0.06

(a) Rewards are placed on the grid according to the two probability distributions. (b) "Starting balance": endowment at the beginning of the game. "Moves per game": total number of moves available to the subject in the game. "Cost per move": cost assessed on each move and deducted from the current balance. "Reward amount": amount added to the participant's total balance upon reaching the reward.

of opening the bottom door—an innovation that might lead to superior outcomes).

B. Treatments

Given the layout and rules of the environment described above, we vary the structure of incentives within the experiment in two ways: we vary the placement of rewards to vary the relative benefit of exploration and vary the direction of earnings to vary exposure to losses. Thus, we implement a 2×2 factorial design. We vary each factor such that a change in one dimension does not affect expected payoffs resulting from differences in the other dimension. In this section, we describe the configuration of two probability distributions for the placement of rewards (baseline versus breakthrough) and the two earning policies for the avoidance or inducement of loss aversion (gains versus losses). In section IIC, we highlight the instructions provided regarding the four treatments.

Factor 1: Placement of rewards. Table 1a presents two configurations for the placement of rewards used in the experiment. First, in the baseline treatment, rewards are equally likely to appear in any of four potential positions: a top position (G4), a left position (D1), a right position (D7), and a bottom position (A4). Second, in the breakthrough treatment, rewards are skewed to appear only at the top position 25% of the time and at the bottom position 75% of the time. The distributions of rewards were chosen carefully so that the expected number of moves to the reward through the top door (F4) is the same in both the baseline and breakthrough treatments, even as the relative incentive to explore through the bottom door (A4) increases between the baseline and breakthrough. In the baseline treatment, the expected number of moves to earn a reward through the bottom door is the same as the expected number of moves to the reward through the top door; as such, once the top door is open there is no incentive to explore the possibility of opening the bottom door. In the breakthrough treatment, however, the expected number of moves to the reward through the bottom door (if it were to open) is fewer than the expected number of moves to the reward through the top door; as such, a participant with *foresight* that the bottom door might open would then have an incentive to try the bottom door. In this way, we create a breakthrough opportunity (i.e., a shortcut), but only for those willing to act on such foresight. We summarize this key in-

sight to the baseline versus breakthrough manipulation with the following remark:

Remark 1. The expected number of moves to the reward through the top door is the same in both the baseline and breakthrough treatments.

Factor 2: Framing of rewards. Table 1b presents two framings of earnings used in the experiment. In the gains treatment, subjects begin the game with an initial balance of \$1.50 and each move is free (i.e., costs \$0.00); the reward amount remains constant at \$0.06. In the losses treatment, subjects begin the game with a higher initial balance of \$6.50, and each move costs \$0.01; the reward amount remains constant at \$0.06. Notice that because every subject has 500 moves and the number of moves is both fixed and known to participants at the start of the game, the expected payoffs at the end of the game are equivalent for both the gains and losses treatments. We summarize this key insight to the gains versus losses manipulation with the following remark:

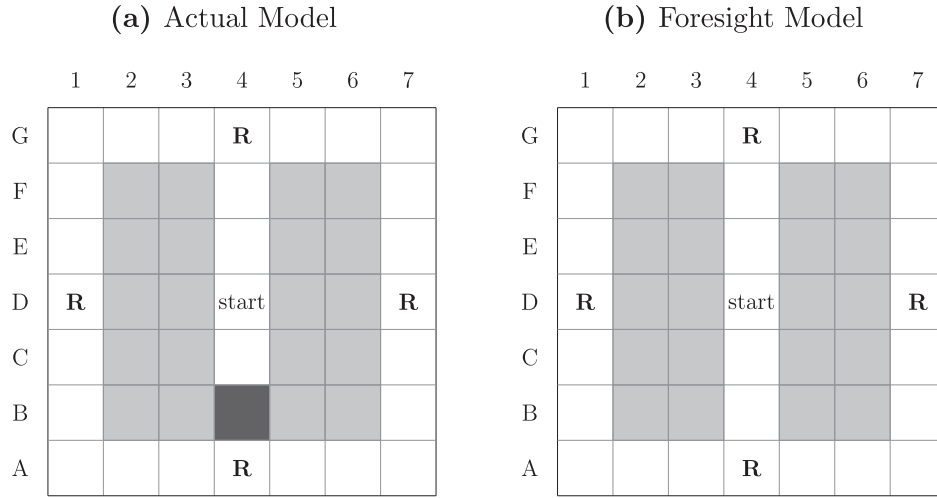
Remark 2. The payoffs for both the gains and losses treatments are equivalent.

C. Instructions

The instructions provided to the subjects were identical in the four treatments. In particular, the main part of the instructions, which pertained to the rules of the maze game, was as follows: "In the experiment, you will navigate through a maze and collect rewards. You can observe your location, your balance, and the remaining number of moves available to you in the game. You will start the game with 500 moves. You must complete all 500 moves in order to be paid a bonus. Rules for your game were determined prior to the start of play. Your moves may be blocked and/or you may be forced to restart from a given position. You must navigate around the board, subject to the rules, in order to collect rewards." (See online appendix A for a full set of instructions.) This is all that subjects knew about the experimental setting prior to start of the game.

The instructions were ambiguous for several reasons. First and foremost, our goal was to focus on behavior in uncertain environments. Thus, we did not provide any information about the distribution of rewards or about the behavior of doors. Instead, subjects learned information about rewards and doors through experience. Second, we did not want to induce wrong beliefs about the environment through the instructions, as doing so could be considered a form of deception. To that end, we chose not to tell the participants that the doors might (or might not) be permanently removed after an unknown number of tries. Instead, we chose to remove the top door after the second hit to demonstrate that doors might be removed, but we chose not to explicitly say anything about the bottom door. As such, participants learned about the possibility that doors may open (in general) only through experience rather than through instructions. Third,

FIGURE 2.—MODELS OF THE ENVIRONMENT



Potential reward positions are marked with a capital R (G4, D1, D7, A4). The reward positions are not shown to subjects but must be discovered. The top door (F4) is opened upon the second hit, and it remains open for the duration of the experiment. The bottom door (B4) remains closed for the duration of the experiment. Subjects can hit a door from any side; if they do, the door temporarily turns red and they return to the start position. (a) Actual model of the environment with the bottom door closed. (b) Foresight model of the environment with the bottom door open.

we wanted to start each of the four treatments with the exact same information regarding the game to ensure a clean comparison between treatments. Specifically, we did not want to emphasize any asymmetry between losses or gains in the instructions, because that might induce differences in behavior due to differences in explanations and not due to differences in the desired treatment conditions. Instead, the framing of loss aversion was imposed by varying the starting balance and cost of moves between the two settings and the intuitive design of the maze game environment.

III. Predictions

In this section we develop predictions for our experiment. A challenge with predicting outcomes for the Maze Game is that the environment is inherently uncertain: there is no objective information about the distribution of rewards, and we leave it up to the individual to evaluate whether the bottom door will, or will not, open (thereby changing the future valuation of rewards). Furthermore, because each period in the game is unique and the environment is uncertain, any strategy that did, or did not, work in the past could lead to a new outcome at a later stage in the game. We therefore do not rely on probability updating by subjects but rather develop predictions based on how loss aversion interacts with Knightian foresight (i.e., a subjective belief about a possible distribution of future rewards in the game) to increase or decrease expected per period rewards of future exploration.⁴

Our game allows for competing views (i.e., multiple models) about an uncertain environment. Figure 2 illustrates these

two views. In the actual model (panel a), the top door is opened, but the bottom door always remains closed; this view is consistent with both the experience of subjects and the actual rules of the environment. In the foresight model (panel b), the top door is opened but subjects consider the possibility that the bottom door also may be opened through some unknown set of actions at some point during the game. This view is consistent with an intuition of how the rules of environment might be, given what happened to the top door.

While we are able to manipulate the placement and framing of rewards (baseline versus breakthrough and gains versus losses), we are not able to manipulate whether individuals do in fact consider a foresight model of the environment when making decisions. We therefore pursue two lines of analytical inquiry to predict how being loss averse (or not) and/or how having foresight (or not) affect the likelihood of exploration. We then compare different cases under the model to derive testable predictions for the experiment. Specifically, we derive the probability that an agent will, or will not, take an exploratory action, which we define as choosing to go down from the starting position during the game.

A. Stochastic Choice Model

To derive our predictions, we apply a classic stochastic model of choice (Luce, 1959) with four simplifying assumptions about subjects' behavior. The first assumption is regarding how subjects bracket the payoffs. Specifically, when making a decision in period t , we assume that subjects will integrate all of the payoffs while following the same strategy, s , in each period of the remaining horizon, h_t . To do this, we begin by calculating the expected per period

⁴Note that there also could be another level of crossed factors affecting our predictions: whether individuals actually are, or are not, loss averse. The assumption of loss aversion has been widely tested in the literature, however, and so for the purposes of this study, we assume that individuals are in fact loss averse. The effect of loss aversion, however, has not been

investigated in the context of Knightian uncertainty. That is the focus of this study.

rewards for each of the four available strategies $s \in \{\text{Up-Right, Up-Left, Down-Right, Down-Left}\}$ for each of the models $m \in \{\text{actual, foresight}\}$: $u_t^{s,m} = \int_{\mathcal{F}_t} \frac{r^{s,m}}{c^{s,m}(x)} dx$, where $\mathcal{F}_t$ is the subject's subjective assessment of the distribution of rewards; $c^s(x)$ is number of steps taken using strategy s until either hitting the door or reaching the reward at position x ; and $r^{s,m}$ is the reward amount obtained by following strategy s when considering model m .⁵ Then, defining $u_t^{U,m} = \max\{u_t^{UR,m}, u_t^{UL,m}\}$ and $u_t^{D,m} = \max\{u_t^{DR,m}, u_t^{DL,m}\}$, we calculate the value of going up or down for each of the two models of the environment under each of the gains and losses treatments:

	Up	Down
Gains	$V_{t,g}^{U,m} = v(h_t \times u_t^{U,m})$	$V_{t,g}^{D,m} = v(h_t \times u_t^{D,m})$
Losses	$V_{t,l}^{U,m} = v(h_t \times u_t^{U,m} - h_t)$	$V_{t,l}^{D,m} = v(h_t \times u_t^{D,m} - h_t)$

The second assumption pertains to the value function, $v(\cdot)$. Specifically, we will use the simplest model of loss aversion: a piece-wise linear function with a kink at 0: $v(x) = x$ if $x \geq 0$ and $v(x) = \delta x$ if $x < 0$ with $\delta > 1$ denoting loss aversion. The two assumptions considerably simplify the predictions. Specifically, to examine the probability that an agent will choose to explore the breakthrough opportunity, we use the logit specification of the stochastic choice model (Luce, 1959): $p_t^{D,m} = \frac{1}{1 + e^{V_{t,g}^{U,m} - V_{t,g}^{D,m}}}$, where $p_t^{D,m}$ denotes the probability of taking an exploratory action (choosing to go down from the starting position) in period t under model $m \in \{\text{actual, foresight}\}$ of the environment. The stochastic choice model has been widely used to study individual choice across economics and operations domains (Luce, 1959; McFadden, 1974; Su, 2008), as well as outcomes in multiagent strategic environments (McKelvey & Palfrey, 1995). Note that the probability depends on the difference in the valuation between going up and going down, which is simple given the piece-wise linear value function.

The third assumption is that subjects' assessment of the distribution of rewards is consistent with the theoretical distribution presented in table 1a. In particular, this assumption implies that (a) the realized placement of rewards and the subjective assessment of that distribution are not substantially different between gains and losses (across the two treatments, subjects expect the same number of moves to the reward through a given door); and (b) the realized placement of rewards and the subjective assessment of that distribution in baseline and breakthrough are such that (a) there is no difference in subjective assessment of the expected moves to the rewards through the top door and (b) the expected number of moves to the reward through the bottom door (if open) is less in breakthrough than in baseline. While the third assumption is helpful to derive theoretical predictions for our experiment in expectation, we will relax this assumption in our em-

pirical analysis (section IVC) and simulated learning model (section V).

The fourth and final assumption pertains to subjects' beliefs about the behavior of the bottom door. In particular, we assume that under the actual model, subjects believe that the bottom door is closed and will remain closed for the remainder of the experiment. Under the foresight model, however, subjects believe that the door will open on the next attempt and stay open for the remainder of the experiment. Given the ambiguous nature of the environment, we adopt a binary representation of beliefs about the bottom door as a simplification. We relax this assumption later (section V) when we introduce a learning model in which agents may hold and update more complicated beliefs about the bottom door.⁶

B. Predicted Probability of Exploration

In view of the stochastic model of choice and assumptions reviewed above, we now derive the predicted probability of exploration (of going down to the bottom door) for the experiment. We consider four cases of our 2×2 design, depending on whether participants do (do not) consider a foresight model and whether participants are (are not) subject to loss aversion.

Case 1: Actual model, no loss aversion. In the first case, agents do not have foresight and are not loss averse. In other words, they believe that the bottom door is closed and their value function is linear. Then the difference between payoffs for the up path and the down path will be the same between the gains (subscript g) and losses (subscript l) treatments, as well as between the baseline (subscript ba) and breakthrough (subscript br) treatments:

$$V_{ba,g}^U - V_{ba,g}^D \approx V_{ba,l}^U - V_{ba,l}^D \approx V_{br,g}^U - V_{br,g}^D \approx V_{br,l}^U - V_{br,l}^D \\ \Rightarrow p_{ba,g} \approx p_{ba,l} \approx p_{br,g} \approx p_{br,l}.$$

Thus, in case 1, there should be no significant difference in the number of exploratory actions among the four treatments of the experiment.

Case 2: Actual model, loss aversion. In the second case, agents do not have foresight, but they are loss averse. That is, they believe that the bottom door is closed, but agents value losses more than gains. Then the valuation of gains and losses will differ, but there should be no difference between the baseline and breakthrough treatments:

$$V_{ba,g}^U - V_{ba,g}^D \approx V_{br,g}^U - V_{br,g}^D < V_{ba,l}^U - V_{ba,l}^D \approx V_{br,l}^U - V_{br,l}^D \\ \Rightarrow p_{ba,g} \approx p_{br,g} > p_{ba,l} \approx p_{br,l}.$$

⁶It also is possible that subjects could believe that the bottom door will open after some number of, or combinations of, future attempts on the door. For the predictions made in this section, it is important only that subjects believe that the bottom door can open under foresight.

⁵Notice that $r^{DR,actual} = r^{DL,actual} = 0$ and $c^{DR,actual} = c^{DL,actual} = 2$.

FIGURE 3.—SUMMARY OF PREDICTIONS

		Model						
		Actual			Foresight			
Loss Aversion	No	①	Baseline	Breakthrough	③	Baseline	Breakthrough	
		Gains	$N_{ba,g}$	$\sim$	$N_{br,g}$	Gains	$N_{ba,g}$	$<$ $N_{br,g}$
			$\wr$			$\wr$		
		Losses	$N_{ba,l}$	$\sim$	$N_{br,l}$	Losses	$N_{ba,l}$	$<$ $N_{br,l}$
	Yes	②	Baseline	Breakthrough	④	Baseline	Breakthrough	
		Gains	$N_{ba,g}$	$\sim$	$N_{br,g}$	Gains	$N_{ba,g}$	$<$ $N_{br,g}$
		$\vee$		$\vee$	$\wr$	$\wedge$		
	Losses	$N_{ba,l}$	$\sim$	$N_{br,l}$	Losses	$N_{ba,l}$	$<$ $N_{br,l}$	

N denotes the number of exploratory actions for configuration { ba for baseline, br for breakthrough} and framing { g for gains, l for losses}. The $>$ symbol indicates a significant difference between conditions; the $\sim$ symbol indicates no significant difference.

Thus, in case 2, there should be fewer exploratory actions in the losses treatment than in the gains treatment, but there should be no difference in the number of exploratory actions when comparing baseline and breakthrough treatments.

Case 3: Foresight model, no loss aversion. In the third case, agents have foresight, but they are not loss averse. In other words, agents believe that the bottom door is open and their value function is linear. Then there will be a difference in the payoffs between the baseline and breakthrough treatments, but that difference will be the same for the gains and losses treatments:

$$V_{ba,g}^U - V_{ba,g}^D \approx V_{ba,l}^U - V_{ba,l}^D > V_{br,g}^U - V_{br,g}^D \approx V_{br,l}^U - V_{br,l}^D \\ \Rightarrow p_{ba,g} \approx p_{ba,l} < p_{br,g} \approx p_{br,l}.$$

Thus, in case 3, there should be fewer exploratory actions in the baseline than in the breakthrough treatments, but there should be an equal number of exploratory actions when comparing the gains and losses treatments.

Case 4: Foresight model, loss aversion. In the fourth case, agents have foresight and are loss averse. In other words, they believe that the bottom door is open, and agents value losses more than gains. Then there will be a difference in valuations between the baseline and breakthrough conditions. Additionally, there will be a difference in valuations between the gains and losses conditions in the breakthrough condition, but there will not be a difference in valuations between the gains and losses conditions in the baseline condition:

$$V_{ba,g}^U - V_{ba,g}^D \approx V_{ba,l}^U - V_{ba,l}^D > V_{br,g}^U - V_{br,g}^D > V_{br,l}^U - V_{br,l}^D \\ \Rightarrow p_{ba,g} \approx p_{ba,l} < p_{br,g} < p_{br,l}.$$

Thus, in case 4, there should be fewer exploratory actions in the baseline condition than in the breakthrough condition. Furthermore, in the breakthrough condition, there should be fewer exploratory actions in the gains condition than in the losses condition.

C. Summary of Predictions

Figure 3 summarizes the predictions from the four cases. Specifically, the number of exploratory actions should be the same in the gains and losses treatments if subjects are not loss averse (cases 1 and 3), because costs associated with moving cancel out when evaluating the differences in the values V_t^U and V_t^D . However, if subjects are loss averse (cases 2 and 4), the direction of the prediction will depend on whether subjects consider the actual model or the foresight model.⁷ A comparison between the baseline and breakthrough conditions is more complicated, because rewards come from different probability distributions. However, we chose the distributions of rewards such that the expected number of moves to a reward through the top door is the same on average. Therefore, if subjects act according to the actual model, we expect no difference in the number of exploratory actions on average (cases 1 and 2); but if subjects act according to the foresight model, then the path through the bottom door will be more attractive under the breakthrough treatment, on average, than under the baseline treatment (cases 3 and 4).

To summarize, the effect of loss aversion depends on whether subjects do, or do not, act with foresight. If they do

⁷If subjects act according to the actual model, the value difference between V_t^U and V_t^D will increase with an increase in loss aversion, leading to fewer exploratory actions. But if subjects act according to the foresight model, the value difference between V_t^U and V_t^D will decrease with an increase in loss aversion, leading to more exploratory actions.

TABLE 2.—RANDOMIZATION CHECK AND DEMOGRAPHICS BY CONDITION

	Female	Male	Age ≥ 34	Stat ≥ 2	Econ ≥ 2	Biz ≥ 2	College
Gains	50.5%	49.5%	49.0%	22.5%	28.5%	32.0%	56.0%
Losses	48.0%	51.0%	51.0%	20.5%	27.5%	32.0%	53.0%

“Stats”: percentage of subjects with more than one course in statistics. “Econ”: percentage of subjects with more than one course in economics. “Biz”: percentage of subjects with more than one course in business. “College”: percentage of subjects with a college degree.

act with foresight, then loss aversion will lead to more persistence at exploration, but if they do not act with foresight, then loss aversion will lead to less persistence at exploration. The basic intuition for this result is that loss aversion magnifies mistakes: if subjects believe that the bottom door can be opened, then it would be a mistake to go up, and going up is more costly under loss aversion. In the next section, we test these predictions with human subject experiments.

IV. Results

To test our predictions, we recruited 300 subjects from the Amazon Mechanical Turk labor market (“M-Turk”) and randomly assigned subjects to a treatment such that each baseline treatment had 50 participants and each breakthrough treatment had 100 participants. We chose the M-Turk population for the experiment as it allowed us to run a large number of human participants through the experiment with strong incentives. Participants were restricted to M-Turk workers located in the United States, with a “master” status and an approval rating of at least 90% on prior work conducted at M-Turk. The M-Turk population is now widely used to recruit subjects for social science experiments (Paolacci, Chandler, & Ipeirotis, 2010; Buhrmester, Kwang, & Gosling, 2011; Horton, Rand, & Zeckhauser, 2011; Rand, 2012; Goodman, Cryder, & Cheema, 2012).

Participants were told in an advertisement that they could earn up to \$4.00 as part of the experiment, and after recruitment, they were told that they would earn a base payment of \$2.00, plus the possibility for a substantial bonus, depending on decisions they made within the game. Final earnings ranged between \$3.24 and \$4.38 (mean = \$3.86). As the experiment lasted approximately 15 minutes, on average, the effective average hourly earnings of \$15.44 was a high compensation rate for M-Turk workers (Horton & Chilton, 2010, for example, find a median reservation wage of \$1.38 per hour). Experiments were conducted online on a private web server. We ran experiments between the gains and losses conditions simultaneously to avoid potential differences in populations related to the hour of the day or the day of the week (i.e., any such population differences should be randomly assigned equivalently across the conditions). An *ex post* randomization check (table 2) between the gains and losses conditions suggests a relatively uniform assignment between conditions for educational and demographic characteristics.

The rest of this section is organized as follows: First, we present the data collected by the experiment. Second, we con-

duct permutation tests for our main inferential test of causality and our main result. Next, we perform regression analyses to control for heterogeneity in the random placement of rewards and to correlate the effect of demographic variables with exploratory behavior.

A. Data

The number of exploratory actions (decisions to go down from the starting position after the top door has been opened) is the main outcome of interest for the study. Figure 4 plots the average cumulative number of exploratory actions by condition. Two observations stand out in the figure. First, subjects tended to explore the unproven path more frequently in the breakthrough treatment than in baseline. Second, they tended to explore the unproven path more in the losses treatment than in gains. Thus, we find qualitative support for the predictions in section III with respect to loss-averse subjects who act with foresight. Note that greater exploration under breakthrough compared to baseline is a basic requirement for our experiment.⁸ If individuals did not try to take a shorter path to rewards in the game after discovering an advantageous distribution of rewards, then the game itself would not be a valid test for exploration.

Table 3 reports descriptive statistics for aggregated behavior in our experiment. In the following two sections, we test the statistical significance of differences in the level of exploration with permutation tests and regression analysis.

B. Permutation Tests

Table 4 reports statistical comparisons between levels of exploration by treatment condition made through two-tailed permutation tests. Permutation tests are nonparametric randomization tests in which the distribution of the test statistic is obtained through random permutation of labels for treatment among observations (Phipson & Smyth, 2010; Good, 2013). The *p*-value for the statistical comparison is obtained by comparing the actual test statistic to the constructed distribution. For example, consider the cells in the breakthrough column: there were 100 observations for gains and 100 observations for losses. Thus, there were 100 observations labeled “G” and 100 labeled “L,” for a total of 200 observations in the breakthrough column. Considering the difference of means as the statistic of interest, let us denote the original difference as *d*-original. Under the null hypothesis, the labels are interchangeable among subjects because treatment does not matter. Therefore, in order to construct the empirical distribution of the test statistic under the null hypothesis, we generate *m* random permutations of the labels (e.g., 10,000), and then, for each permutation, calculate the statistic of interest,

⁸As a robustness test, we ran experiments for an intermediate configuration that weighted rewards toward the bottom more than baseline, but less than breakthrough (see online appendix C). As expected, the level of exploration was greater than baseline and lower than breakthrough.

FIGURE 4.—CUMULATIVE NUMBER OF EXPLORATORY ACTIONS FOR HUMAN SUBJECTS

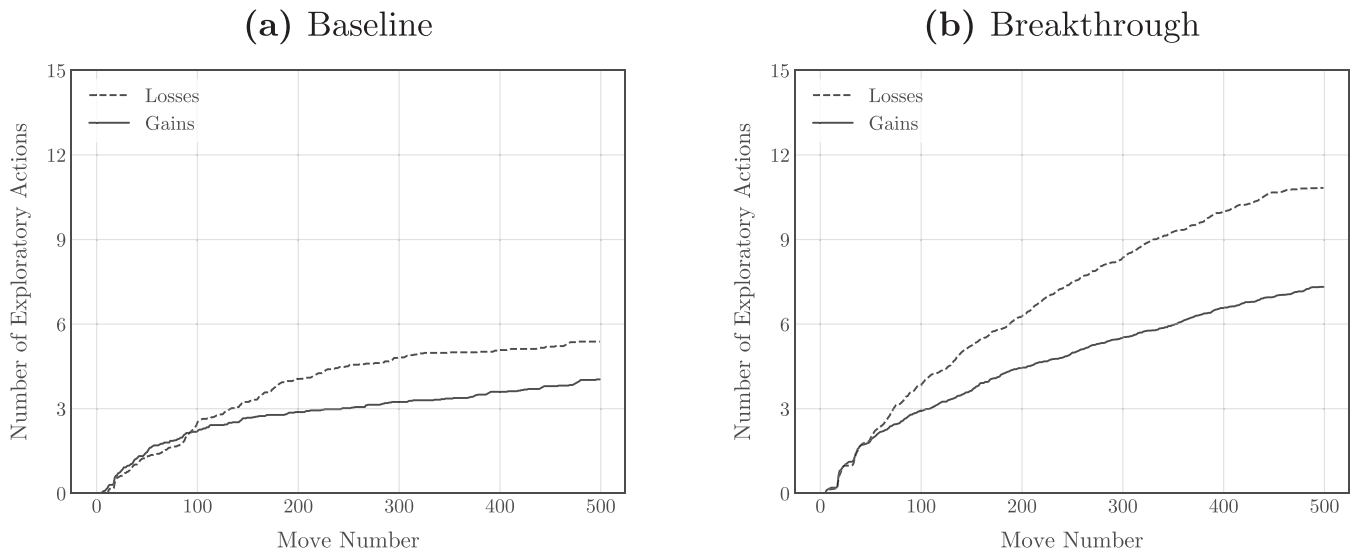


TABLE 3.—SUMMARY STATISTICS

	(1) Baseline		(3) Breakthrough	
	Gains	Losses	Gains	Losses
Time per move	1.38 (0.54)	1.27 (0.45)	1.28 (0.6)	1.32 (0.55)
Open top door	3.12 (0.84)	3.38 (0.82)	3.47 (2.29)	3.35 (1.25)
Exploration	4.04 (2.93)	5.38 (5.38)	7.32 (6.79)	10.83 (13.79)
Exploration—Interim moves	246.74 (153.18)	276.08 (140.23)	217.52 (159.94)	221.59 (157.32)
Exploration—Final attempt	252.12 (153.47)	222.82 (140.39)	281.53 (160.01)	277.21 (157.7)
Exploitation	41.14 (3.53)	39.46 (3.47)	39.37 (3.15)	38.91 (3.09)
Exploitation—Clockwise	15.0 (3.62)	15.86 (3.6)	17.17 (7.46)	15.7 (7.11)
Exploitation—Counterclockwise	16.16 (4.17)	15.3 (3.96)	13.11 (7.36)	14.13 (7.06)
Number of subjects	50	50	100	100
Number of moves	500	500	500	500

Values averaged by condition. Standard deviations in parentheses. “Time per Move” in seconds. “Open Top Door” is number of attempts at a door before the top door opened. All values for exploration and exploitation are calculated after top door opened. “Exploration” is number of exploratory actions (decisions to go down from the starting position after the top door has been opened). “Exploration—Interim moves” is number of moves between consecutive exploration actions. “Final exploration” is number of moves taken in game at time of final exploration action. “Exploitation” is number of times moving up. “Exploitation—Clockwise” is number of times moving up and left. “Exploitation—Counterclockwise” is number of times moving up and right.

d -permut. Finally, by counting the number of permutations, b , for which the absolute value of the statistic of interest exceeds or is equal to the absolute value of d -original, the two-tailed p -value rejecting the null hypothesis can be calculated as $p = \frac{b+1}{m+1}$ (Ernst, 2004; Phipson & Smyth, 2010).

As reported in table 4, there are two key results to consider. First, we find that subjects were significantly more likely to explore the unproven path (bottom door) in the breakthrough treatment, regardless of whether they are under a condition of gains or losses. Although this is unsurprising, it confirms that subjects do act with foresight in the experiment and that we therefore achieved our objective of inducing the innovation

TABLE 4.—AVERAGE NUMBER OF EXPLORATORY ACTIONS

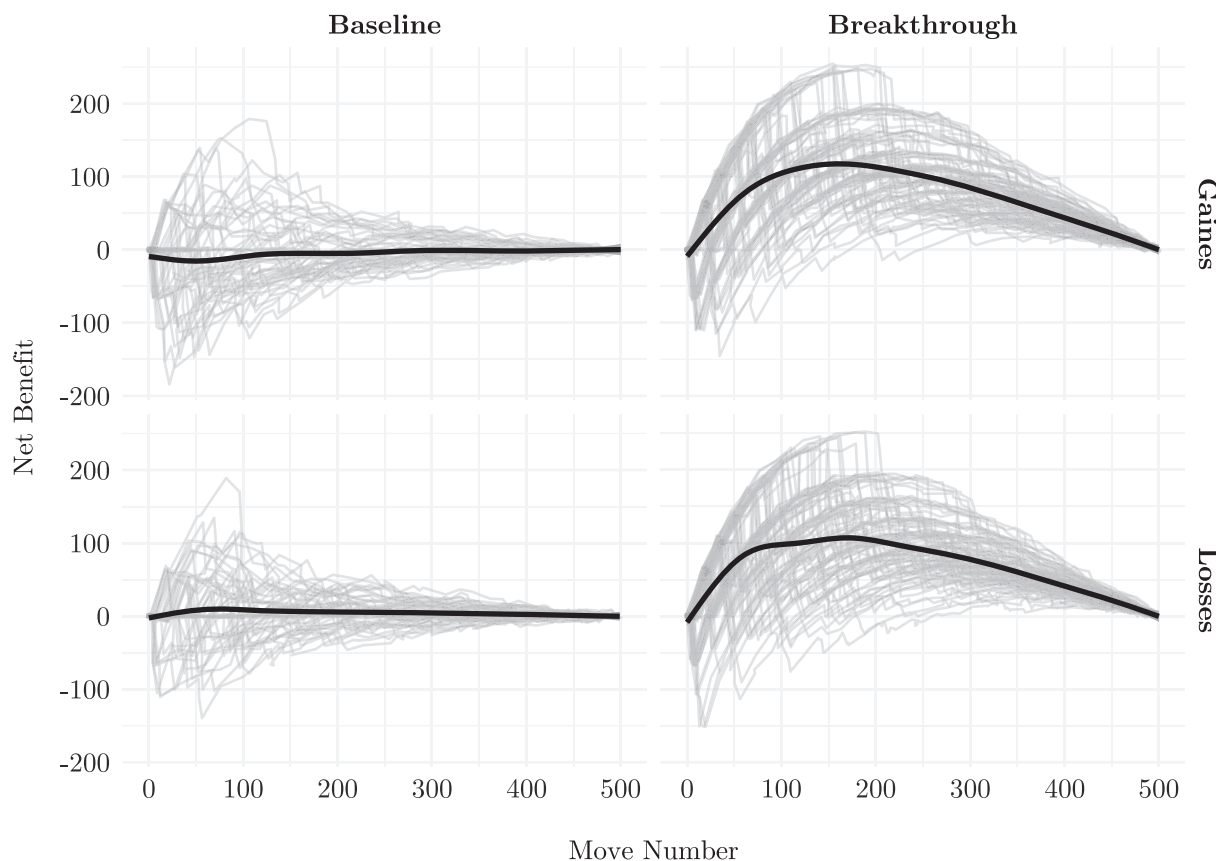
	Baseline		Breakthrough
Gains	4.04 (0.414)	»»	7.32 (0.678)
Losses	5.38 (0.752)	»»	10.83 (1.374)

Bootstrapped standard errors are in parentheses. » and »» denote significance at the 0.05 and 0.01 levels, respectively. ~ denotes no significance. p -values are determined using two-tailed permutation tests. The unit of observation is a unique subject.

problem described at the outset of the paper. Second, in the main result of the paper, we find that human subjects attempt to go down the unproven path (in an attempt to open the bottom door) significantly more when incentives are framed as losses, as opposed to when equivalent incentives are framed as gains. These two results are in line with the predictions in section III for loss-averse agents who act with foresight.

It is important to point out that randomness in the placement of rewards is likely to generate substantial heterogeneity with respect to the realized sequence of signals in each treatment. It is not inconceivable that a sequence of the placement of rewards from the baseline treatment might actually be more representative of a breakthrough treatment than a baseline treatment (and vice versa) due to the random nature of the environment. Although it would be unlikely for that to happen on average (and the descriptive statistics suggest that it did not), our research design is “between-subjects” and so results may be clearer if we take the heterogeneity of signals into account. In the sections that follow, we do so in two ways. First, we condition on the theoretical potential net benefit of succeeding at the exploration task (opening the bottom door), regardless of whether the subject comes from the baseline or breakthrough treatment. Second, we construct matched samples where subjects from different treatment conditions for gains versus losses are paired up and compared to each other

FIGURE 5.—NET BENEFIT, CONDITIONAL ON FORESIGHT MODEL BEING TRUE



Thin lines represent the expected difference between going through the bottom door (B4) and the top door (F4) under the foresight model for the remainder of the experiment. Expected values are obtained from realized signals by applying Bayes' rule, starting with a uniform prior. Each subject has his or her own trajectory with respect to the expected net benefit over time. Thick black lines represent the average potential net benefit for each treatment.

based on the exact sequence of reward placements (i.e., “signals”) that they observe.

C. Regression Analysis

Consider the following scenario. Two participants—one in the baseline treatment and the other in the breakthrough treatment—both find their first reward at location A4. In other words, both participants receive the same first signal about the distribution of rewards, regardless of treatment. Although the two subjects are assigned to different reward distributions, the situation is the same from their perspective. Therefore, to better control for homogeneity between treatments and heterogeneity within treatments, we pursue two regression approaches.

In our first regression approach, we develop a new measure for “net benefit, conditional on foresight.” Specifically, we calculate the difference in the expected rewards one would obtain between opening and going through the bottom door (location B4) for the remainder of the experiment, compared to going through the top door (location F4). We updated expectations using Bayes' rule, starting with a uniform prior. Knowing the potential net benefit of opening the bottom door allows us to compare decisions made by subjects with differ-

ent realized signals at different points in time. Moving to a regression framework also allows us to control for other demographic and educational characteristics. Figure 5 presents the new net benefit measure for each treatment. As expected, there is substantial heterogeneity within each treatment in terms of the placement of rewards actually observed by subjects. Furthermore, there is substantial overlap between the baseline and breakthrough treatments—variation that is statistical noise in our permutation tests. The average net benefit of opening the bottom door (the bold black lines in figure 5) is also consistent with the parameters we set for the distribution of rewards for each treatment.

Results for our first regression are presented in table 5. We determine whether a subject goes down (coded 1) or up (coded 0) from the starting position and run a logit model for the probability of taking an exploratory action (attempting to go down through the bottom door) once the top door is opened. We regress on treatment condition (losses coded 1, gains coded 0), and control for net benefit, moves into the game, and demographic characteristics.⁹ Coefficients are

⁹ Although including net benefits and moves in the regression introduces endogeneity into our model specification, the goal of the analysis in table 7 is to assess whether the estimation of losses changes after controlling

TABLE 5.—LOGIT MODEL OF EXPLORATION, GIVEN POTENTIAL NET BENEFIT

	(1)	(2)	(3)	(4)
Losses	1.397** (0.168)	1.394** (0.167)	1.413** (0.179)	1.434** (0.182)
Net benefit		1.001 (0.001)	1.001 (0.001)	1.001 (0.001)
Moves			0.996*** (0.000)	0.996*** (0.000)
Sex				0.771 (0.107)
Age				0.991 (0.006)
College				1.159 (0.155)
Statistics				1.026 (0.085)
Biz and Econ				0.938* (0.025)
Constant	0.180*** (0.013)	0.171*** (0.012)	0.388*** (0.033)	0.640* (0.141)
Observations	15,180	15,180	15,180	14,633
Log likelihood	−7,065	−7,060	−6,695	−6,399

The dependent variable is *Exploration*, coded 1 when going down. The model controls for the “Net Benefit” of exploration, conditional on *Foresight*. Coefficients are odds ratios, with robust standard errors, clustered by subject, in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

reported as odds ratios, and robust standard errors are clustered by subject. The results in table 5 are consistent with the main result in table 4 from the permutation analysis: we find a strong, causal association between being in the losses condition and the likelihood of taking an exploratory action. Unsurprisingly, there also is a tendency to eventually give up as one progresses through the game, as indicated by the highly significant coefficients on moves. Finally, the only demographic variable significant at the 0.05 level is Biz & Econ, which represents that business and economics students are marginally less likely to explore overall.

For our second regression approach, we match subjects based on the realized distribution of signals observed in the first half of the experiment. We restrict the sample to the first 250 moves in order to limit the effects of sample attrition from failing to find a match and to confirm that results are not driven by idiosyncratic behavior in the second half the game. To match observations, we build a cycle-by-cycle distribution of reward locations observed by each subject, and then match a subject from the losses treatment to a subject from the gains treatment with the same distribution of rewards. When there was more than one potential match, we picked one at random; when there was no match, we dropped the observations. We determine whether a subject goes down (coded 1) or up (coded 0) from the starting position and run a logit model for the probability of taking an exploratory action. We again regress on treatment condition and control for moves into the game. Results from the matched analysis are presented in table 6.

for the level of net benefits and moves and not to specifically estimate the coefficients for net benefits or moves. For robustness, we address potential endogeneity issues in table 8 by removing the variable Net Benefit from the analysis and controlling for differences in observed signals with a matched design.

TABLE 6.—LOGIT MODEL OF EXPLORATION, GIVEN MATCHED SAMPLES

	(1) Matched by Signal	(2) Matched by Signal	(3) Matched by Signal and Move	(4) Matched by Signal and Move	(5) Matched by Signal and Move
Losses	1.407** (0.177)	1.423** (0.186)	1.391*** (0.138)	1.415*** (0.147)	1.385* (0.220)
Moves		0.994*** (0.001)		0.991*** (0.001)	0.991*** (0.001)
Constant	0.265*** (0.012)	0.518*** (0.052)	0.250*** (0.019)	0.578*** (0.061)	0.536*** (0.090)
Observations	4,920	4,920	2,731	2,731	1,134
Log likelihood	−2,709	−2,642	−1,462	−1,388	−529

Dependent variable is *Exploration*, coded 1 when going down. Coefficients are odds ratios, with robust standard errors, clustered by subject, in parentheses. Columns 1 and 2 match observations by the realized distribution of rewards. Columns 3, 4, and 5 match observations by both the realized distribution of rewards and for being within ± 1 move into the game. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

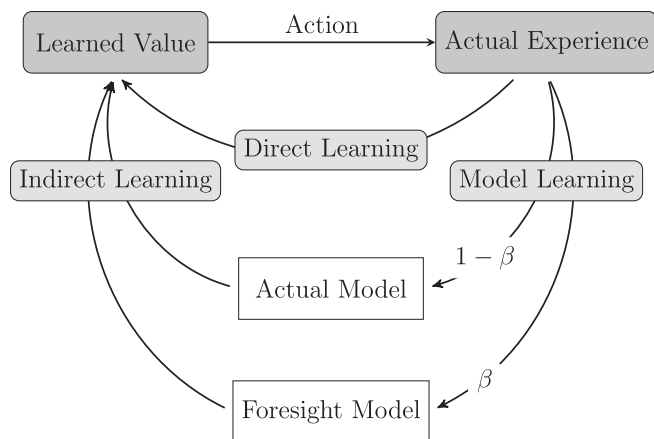
In columns 1 and 2 of table 6, we again find that subjects with incentives structured for loss aversion were significantly more likely to explore. The matching procedure drops 10,260 observations (from 15,180 observations in table 5, to 4,920 observations in column 1 of table 6), but the level of statistical significance and magnitude of effects remains essentially the same as in table 5. Next, in columns 3 and 4, we further restrict the matching procedure to also require that matched pairs of observations be within a window of ± 1 move into the game. The more restrictive matching procedure again drops the sample size (now down to 2,731 observations), and we again find support for our main result with similar sizes of effects and levels of significance. Finally, in an attempt to control for some of the sample selection bias that may result from the matching procedure itself, in column 5 we resample observations from column 4 based on the inverse probability of being selected into the sample by the matching procedure at a given move into the game. Down-sampling in this way rebalances the sample so that matches made from earlier in the game do not dominate the analysis (see online appendix D for more information). Rebalancing the sample again drops the sample size by more than half, now to just 1,134 observations, but the effect size of losses is unaffected and the statistical significance of the coefficient remains better than $p = 0.05$.

Summarizing across all of our empirical results, we find that human subjects are more likely to explore an uncertain strategy, with a potentially higher net benefit, when incentives are structured to induce loss aversion. Permutation tests demonstrate that participants motivated by loss aversion explore more overall, and regression analyses demonstrate that participants motivated by loss aversion are more likely to make an exploratory decision. While these results provide strong evidence that loss aversion is an important determinant of exploratory behavior, they do not capture how participants may learn over time. We use simulation analysis in the following section to examine this question.

V. Simulations

Qualitative evidence from our human subjects experiments suggests that individuals learn about their environment and

FIGURE 6.—REINFORCEMENT LEARNING WITH A CHANNEL FOR FORESIGHT



adjust their behavior over time. As seen in figure 4, subjects tend to experiment during the start of each game, and then behavior differentiates by treatment condition. Specifically, the trajectory of exploration is higher for subjects with incentives structured to induce loss aversion. Also, while persistence at exploration attenuates over time for all treatments, the attenuation is more gradual under the losses condition. In this section, we incorporate foresight and loss aversion into a model of reinforcement learning and calibrate the model to match results from our experiments.

A. Reinforcement Learning with Foresight

We use Q-learning (Sutton, 1990; Watkins & Dayan, 1992) to simulate the action selection and learning process of agents in our environment. Q-learning is a type of reinforcement learning from the machine learning literature. It differs from seminal models of reinforcement learning in economics (Roth & Erev, 1995; Erev & Roth, 1998) in that agents learn (reinforce) values for state-action pairs. In the context of our environment, a state is a position in the maze, and an action is one of the potential directions that one can move, {Up, Down, Left, Right}. We chose a Q-learning approach for several reasons. First, it can be applied to classical multiarmed bandit problems, as well grid-world problems such as our environment (Sutton & Barto, 1998). Second, Q-learning has already been used successfully to model learning behavior in economic environments (Waltman & Kaymak, 2008; Greenwald, Kannan, & Krishnan, 2010). Third, we can build on prior research to incorporate model uncertainty into the reinforcement learning framework; specifically, we add in a step for indirect learning (Sutton, 1990; Sutton & Barto, 1998).

Figure 6 provides intuition as to our use of the Q-learning.¹⁰ The figure illustrates two channels for learning: a direct channel that updates the value function based on actual experience and an indirect channel that bootstraps experience from ei-

ther an actual model or a foresight model based on β , the likelihood of using the foresight model during indirect learning. Even though the bottom door never opens, the foresight model simulates experience as if it were open, making an exploratory action (choosing down in the starting position) more attractive in the breakthrough treatment. Model learning captures an important aspect of our innovation problem: the agent is not sure which model of the environment is true. The environment provides visual cues (such as walls, doors, and the disappearance of the top door) to condition human subjects to consider an alternative model; we simulate that ambiguity by including an indirect channel for learning.

B. Simulation Results

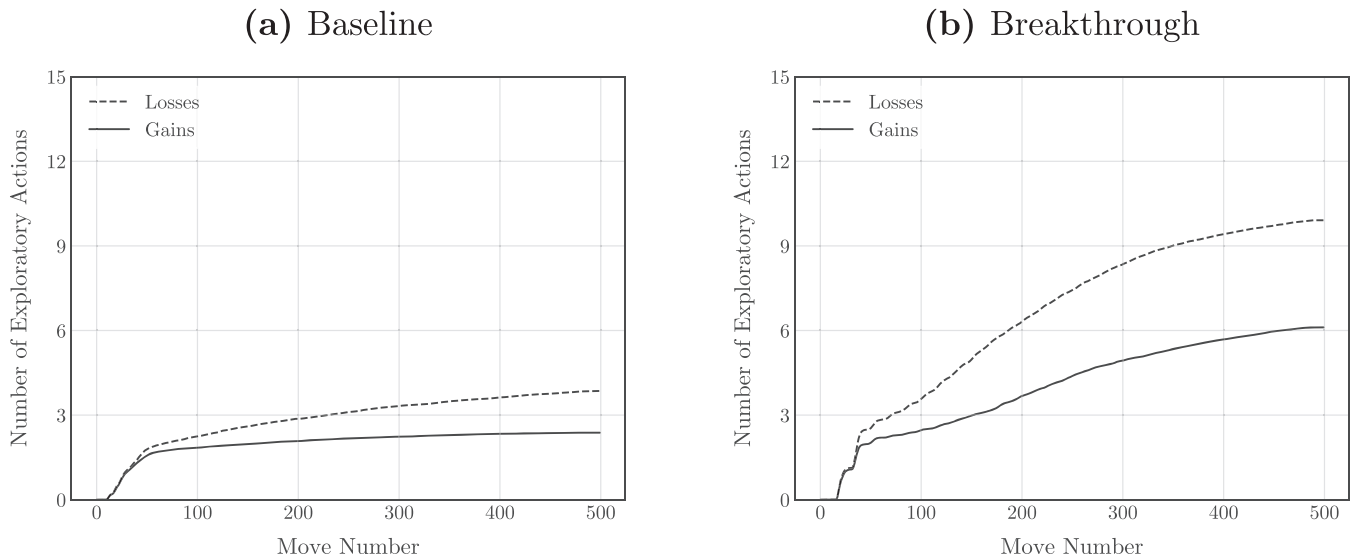
We fit a Q-learning model to our experimental results by minimizing the total squared error between results from the simulation algorithm and results from human experiments. We run Q-learning simulations across 500 moves, for 1,000 games, for each of the four treatments, and then sum the squared difference between the averaged simulated value and the averaged observed value, by treatment. Having thus obtained an average squared error for a given level of parameter values, we then use a Bayesian optimization algorithm to repeat the entire estimation procedure and adjust parameters in a direction determined by the algorithm to be likely to reduce error (see online appendix E2 for a complete description of the optimization procedure).

Figure 7 plots the simulation results for the best-fit parameter values (presented in table E1 of the online appendix). The figure shows that Q-learning simulations closely match data from our experiment. Importantly, while the value of the loss-aversion parameter in the gains condition is set to 1.0, the value of the loss-aversion parameter in the losses condition that is discovered by the optimization procedure is 2.64. This means that levels of loss aversion that are observed in the literature (e.g., 2.25 in Tversky & Kahneman, 1992) could explain the difference between the two treatments. In addition, we consider counterfactuals to show the effect of removing mechanisms for loss aversion and for foresight. As expected, if loss aversion is set to 1.0 then there is no difference between the gains and losses conditions (figure E2 in the online appendix), and, if the likelihood of learning from the foresight model during indirect learning is set to 0.0, then agents do not persist at exploration (figure E3 in the online appendix).

In summary, our simulation results suggest that loss aversion can lead to greater persistence at exploration in environments where actors hold Knightian foresight. Although existing models (either a multiarm bandit or a reinforcement learning from actual experience) might appear to be an attractive way to model our problem, such models will never give any weight to new ideas where there is no probabilistic evidence of success. Thus, our simulations differ from prior research in that our model of learning accounts for other

¹⁰Technical details for Q-learning are presented in online appendix E1.

FIGURE 7.—CUMULATIVE NUMBER OF EXPLORATORY ACTIONS FOR SIMULATIONS



motivations for innovation, such as intuition, analogical reasoning, and entrepreneurial insight.

VI. Conclusion

In this paper, we investigate the willingness of individuals to persist at innovation in the face of failure. Specifically, we incorporate Knight's idea of foresight into a stochastic model of choice and find that it helps to explain human exploration in an uncertain environment with the possibility for innovation. Moreover, our results suggest that individuals explore more when they are reminded of the incremental cost of their actions, a counterintuitive result for research on innovation, but a result that extends findings on loss aversion and predictions from prospect theory.

Our results have implications for how incentives can be framed to increase persistence at a breakthrough innovation. While the literature on tolerance for failure (Azoulay et al., 2011; Manso, 2011) emphasizes the need to tolerate early failure and reward long-term success, our behavioral findings suggest that an optimal incentive structure in an uncertain environment may also benefit from inducing individuals to experience losses along the way (although perhaps while also shielding them from the long-term consequences of being wrong). Studies on escalation of commitment (Staw, 1976, 1981; Staw & Ross, 1978), overconfidence (Camerer & Lovallo, 1999), and fear of failure (Kihlstrom & Laffont, 1979) often portray such factors in a negative light. Our results, however, suggest that incentives tailored to induce loss aversion can be beneficial when the goal is to induce more exploration. In this respect, Hirshleifer, Low, and Teoh (2012) and Galasso and Simcoe (2011) find that overconfident managers and CEOs are associated with greater innovative activity by the firm (as captured by the number of patents filed by the firm). These results align well with our interpretation of Knightian foresight: overconfident managers should have

stronger beliefs about their foresight, and thus they should be more willing to persist at innovation despite a lack of confirmatory evidence to justify their actions.

Finally, we note several limitations to our study that open promising avenues for future research. First, there is substantial heterogeneity in exploratory behavior between subjects. A question for future research is the extent to which differences in the willingness to persist are driven by heterogeneity in loss aversion and/or how much is driven by other individual characteristics. Second, the focus of this study is on the effect that loss aversion has on exploration in uncertain environments. Future work could compare whether the willingness to persist in a risky (as compared to uncertain) environment is different. Third, our experiments do not examine interpersonal or organizational considerations, and we do not test how loss aversion may be moderated by a social context. Future research could examine how loss aversion affects innovation decision making in teams. Finally, in our implementation of the learning model, we assumed a deterministic path for the evolution of beliefs. Future research could build on our model of reinforcement learning to explore how the dynamics of model learning and belief updating relate to a willingness to persist.

REFERENCES

- Acharya, V. V., R. P. Baghai, and K. V. Subramanian, "Wrongful Discharge Laws and Innovation," *Review of Financial Studies* 27 (2014), 301–346.
- Acharya, V. V., and K. V. Subramanian, "Bankruptcy Codes and Innovation," *Review of Financial Studies* 22 (2009), 4949–4988.
- Aghion, P., J. Van Reenen, and L. Zingales, "Innovation and Institutional Ownership," *American Economic Review* 103 (2013), 277–304.
- Armantier, O., and A. Boly, "Framing of Incentives and Effort Provision," *International Economic Review* 56 (2015), 917–938.
- Azoulay, P., J. S. Graff Zivin, and G. Manso, "Incentives and Creativity: Evidence from the Academic Life Sciences," *RAND Journal of Economics* 42 (2011), 527–554.
- Barberis, N., M. Huang, and T. Santos, "Prospect Theory and Asset Prices," *Quarterly Journal of Economics* 116 (2001), 1–53.

- Buchanan, J. A., and B. J. Wilson, "An Experiment on Protecting Intellectual Property," *Experimental Economics* 17 (2014), 691–716.
- Buhrmester, M., T. Kwang, and S. D. Gosling, "Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?" *Perspectives on Psychological Science* 6 (2011), 3–5.
- Camerer, C., and D. Lovo, "Overconfidence and Excess Entry: An Experimental Approach," *American Economic Review* 89 (1999), 306–318.
- Che, Y.-K., and I. Gale, "Optimal Design of Research Contests," *American Economic Review* 93 (2003), 646–671.
- Dickinson, D. L., "The Carrot vs. the Stick in Work Team Motivation," *Experimental Economics* 4 (2001), 107–124.
- Ederer, F., and G. Manso, "Is Pay for Performance Detrimental to Innovation?" *Management Science* 59 (2013), 1496–1513.
- Elfenbein, D. W., A. M. Knott, and R. Croson, "Equity Stakes and Exit: An Experimental Approach to Decomposing Exit Delay," *Strategic Management Journal* 38:2 (2016), 278–299.
- Erev, I., and A. E. Roth, "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria," *American Economic Review* 88 (1998), 848–881.
- Ernst, M. D., "Permutation Methods: A Basis for Exact Inference," *Statistical Science* 19 (2004), 676–685.
- Galasso, A., and T. S. Simcoe, "CEO Overconfidence and Innovation," *Management Science* 57 (2011), 1469–1484.
- Gneezy, U., S. Meier, and P. Rey-Biel, "When and Why Incentives (Don't) Work to Modify Behavior," *Journal of Economic Perspectives* 25 (2011), 191–209.
- Good, P., *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses* (New York: Springer Science and Business Media, 2013).
- Goodman, J. K., C. E. Cryder, and A. Cheema, "Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical-Turk Samples," *Journal of Behavioral Decision Making* 26 (2012), 213–224.
- Greenwald, A., K. Kannan, and R. Krishnan, "On Evaluating Information Revelation Policies in Procurement Auctions: A Markov Decision Process Approach," *Information Systems Research* 21 (2010), 15–36.
- Guadalupe, M., O. Kuzmina, and C. Thomas, "Innovation and Foreign Ownership," *American Economic Review* 102 (2012), 3594–3627.
- Hannan, R. L., V. B. Hoffman, and D. V. Moser, "Bonus versus Penalty: Does Contract Frame Affect Employee Effort?" (pp. 151–169), in *Experimental Business Research* (Berlin: Springer, 2005).
- Hansen, L. P., and T. J. Sargent, "Robust Control and Model Uncertainty," *American Economic Review* 91 (2001), 60–66.
- Herz, H., D. Schunk, and C. Zehnder, "How Do Judgmental Overconfidence and Overoptimism Shape Innovative Activity?" *Games and Economic Behavior* 83 (2014), 1–23.
- Hirshleifer, D., A. Low, and S. H. Teoh, "Are Overconfident CEOs Better Innovators?" *Journal of Finance* 67 (2012), 1457–1498.
- Hong, F., T. Hossain, and J. A. List, "Framing Manipulations in Contests: A Natural Field Experiment," *Journal of Economic Behavior and Organization* 118 (2015), 372–382.
- Horton, J., and L. Chilton, "The Labor Economics of Paid Crowdsourcing" (pp. 209–218), in *Proceedings of the 11th ACM Conference on Electronic Commerce* (New York: ACM, 2010).
- Horton, J. J., D. G. Rand, and R. J. Zeckhauser, "The Online Laboratory: Conducting Experiments in a Real Labor Market," *Experimental Economics* 14 (2011), 399–425.
- Hossain, T., and J. A. List, "The Behavioralist Visits the Factory: Increasing Productivity Using Simple Framing Manipulations," *Management Science* 58 (2012), 2151–2167.
- Howell, S. T., "Financing Innovation: Evidence from R&D Grants," *American Economic Review* 107 (2017), 1136–1164.
- Kagan, E., S. Leider, and W. S. Lovejoy, "Ideation–Execution Transition in Product Development: An Experimental Analysis," *Management Science* 64 (2017), 2238–2262.
- Kahneman, D., J. L. Knetsch, and R. H. Thaler, "Experimental Tests of the Endowment Effect and the Coase Theorem," *Journal of Political Economy* 97 (1989), 1325–1348.
- Kahneman, D., and A. Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica* 47 (1979), 263–291.
- Kerr, W. R., R. Nanda, and M. Rhodes-Kropf, "Entrepreneurship as Experimentation," *Journal of Economic Perspectives* 28 (2014), 25–48.
- Kihlstrom, R. E., and J.-J. Laffont, "A General Equilibrium Entrepreneurial Theory of Firm Formation Based on Risk Aversion," *Journal of Political Economy* 87 (1979), 719–748.
- Knight, F. H., *Risk, Uncertainty and Profit* (New York: Hart, Schaffner and Marx, 1921).
- Lerner, J., and J. Wulf, "Innovation and Incentives: Evidence from Corporate R&D," this REVIEW 89 (2007), 634–644.
- Lim, A. E., J. G. Shanthikumar, and Z. M. Shen, "Model Uncertainty, Robust Optimization, and Learning" (pp. 66–94), in M. P. Johnson, B. Norman, and N. Secomandi, eds., *Models, Methods, and Applications for Innovative Decision Making* (INFORMS, 2006).
- Luce, R. D., *Individual Choice Behavior: A Theoretical Analysis* (Hoboken, NJ: Wiley, 1959).
- Manso, G., "Motivating Innovation," *Journal of Finance* 66 (2011), 1823–1860.
- McFadden, D., "Conditional Logit Analysis of Qualitative Choice Behavior" (pp. 105–142), in P. Zarembka, ed., *Frontiers in Econometrics* (New York: Wiley, 1974).
- McKelvey, R. D., and T. R. Palfrey, "Quantal Response Equilibria for Normal Form Games," *Games and Economic Behavior* 10:1 (1995), 6–38.
- Morgan, J., and D. Sisak, "Aspiring to Succeed: A Model of Entrepreneurship and Fear of Failure," *Journal of Business Venturing* 31:1 (2016), 1–21.
- Nanda, R., and M. Rhodes-Kropf, "Financing Entrepreneurial Experimentation," *Innovation Policy and the Economy* 16:1 (2016), 1–23.
- Paolacci, G., J. Chandler, and P. G. Ipeirotis, "Running Experiments on Amazon Mechanical Turk," *Judgment and Decision Making* 5 (2010), 411–419.
- Phipson, B., and G. K. Smyth, "Permutation P-Values Should Never Be Zero: Calculating Exact P-Values When Permutations Are Randomly Drawn," *Statistical Applications in Genetics and Molecular Biology* 9:1 (2010), art. 39.
- Rand, D. G., "The Promise of Mechanical Turk: How Online Labor Markets Can Help Theorists Run Behavioral Experiments," *Journal of Theoretical Biology* 299 (2012), 172–179.
- Roth, A. E., and I. Erev, "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term," *Games and Economic Behavior* 8:1 (1995), 164–212.
- Rubin, J., A. Samek, and R. M. Sheremeta, "Loss Aversion and the Quantity–Quality Tradeoff," *Experimental Economics* 21 (2018), 292–315.
- Seru, A., "Firm Boundaries Matter: Evidence from Conglomerates and R&D Activity," *Journal of Financial Economics* 111 (2014), 381–405.
- Staw, B. M., "Knee-Deep in the Big Muddy: A Study of Escalating Commitment to a Chosen Course of Action," *Organizational Behavior and Human Performance* 16:1 (1976), 27–44.
- , "The Escalation of Commitment to a Course of Action," *Academy of Management Review* 67 (1981), 577–587.
- Staw, B. M., and J. Ross, "Commitment to a Policy Decision: A Multi-Theoretical Perspective," *Administrative Science Quarterly* 23 (1978), 40–64.
- Su, X., "Bounded Rationality in Newsvendor Models," *Manufacturing and Service Operations Management* 10 (2008), 566–589.
- Sutton, R. S., "Integrated Architectures for Learning, Planning, and Reacting Based on Approximating Dynamic Programming" (pp. 216–224), in *Proceedings of the Seventh International Conference on Machine Learning* (San Mateo, CA: Morgan Kaufmann, 1980).
- Sutton, R. S., and A. G. Barto, *Reinforcement Learning: An Introduction* (Cambridge, MA: MIT Press, 1998).
- Tian, X., and T. Y. Wang, "Tolerance for Failure and Corporate Innovation," *Review of Financial Studies* 27 (2014), 211–255.
- Tversky, A., and D. Kahneman, "Loss Aversion in Riskless Choice: A Reference-Dependent Model," *Quarterly Journal of Economics* 106 (1991), 1039–1061.
- , "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty* 5 (1992), 297–323.
- Waltman, L., and U. Kaymak, "Q-Learning Agents in a Cournot Oligopoly Model," *Journal of Economic Dynamics and Control* 32 (2008), 3275–3293.
- Watkins, C. J., and P. Dayan, "Q-Learning," *Machine Learning* 8 (1992), 279–292.