Purdue University Economics Department Working Paper **No 1301**

Motivating Innovation:

The Effect of Loss Aversion on the Willingness to Persist

Yaroslav Rosokha[†] Kenneth Younge[‡]

August 2017

Abstract:

We investigate the willingness of individuals to persist at exploration in the face of failure. Prior research suggests that the organization's "tolerance for failure" may motivate greater exploration by the individual. Little is known, however, about how individuals persist at exploration in an uncertain environment when confronted by prolonged periods of negative feedback. To examine this question, we design a two-dimensional maze game and run a series of randomized experiments with human subjects in the game. We develop predictions for the game using computational models of reinforcement learning. Our methods extend beyond two-period models of decision-making under uncertainty to account for repeated behavior in longer-running, dynamic contexts. Our results suggest that 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. We discuss implications for future research and for managers.

Keywords: Experiments, Innovation, Persistence, Loss Aversion, Model Uncertainty

---

[†] Krannert School of Management, Purdue University • Email: yrosokha@purdue.edu

[‡] École Polytechnique Fédérale de Lausanne • Email: kenneth.younge@epfl.ch

1 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.*
>
> *Knight, 1921: page lxii*

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

While incentives clearly are an important motivator to engage in innovation (Lerner and Wulf, 2007), behavioral research suggests that incentives do not always shape individual behavior as one might expect (Gneezy, Meier, and 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.

In addition to behavioral experiments, we also run computer simulations derived from machine learning and Bayesian models to show that our treatments have no effect on the outcome under the assumption of no model uncertainty. However, we find that the same treatments have large and persistent effects with humans. 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, Graff Zivin, and Manso, 2011; Manso, 2011; Tian and 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; 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 (forthcoming) investigate how to divide limited time between design (exploration) and execution stages during the 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 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 one cannot ensure the decision to innovate to other actors (or markets) by presenting probabilistic evidence about the likelihood of success. When one *can* provide probabilistic evidence, then the problem is 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" (e.g., Manso, 2011). In a bandit model, an actor learns about the probability distribution of a prospect by updating priors over time. For our innovation problem, however, there is no empirically driven prior probability distribution of any sort to be updated, because there is no positive experience about the prospect gathered over time. While, in our framework, individuals may learn about the *potential* benefits of a prospect, any such evidence is conditional on getting the prospect to work – getting the prospect to work, however, remains fundamentally uncertain. Thus, our innovation problem focuses on *model uncertainty* (Hansen and Sargent, 2001; Lim, Shanthikumar, and Shen, 2006), where several specifications are possible, and it is up to the the individual to proceed based solely on intuition about the true underlying model, or 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 (and/or investors) 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 most well-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 and Tversky, 1979; Tversky and Kahneman, 1992). Specifically, individuals tend to interpret each gain or loss relative to a *reference point* (Kahneman, Knetsch, and Thaler, 1990; Barberis, Huang, Santos, et al., 2001), such that a loss affects a value function more than an equivalent gain (Tversky and 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, and 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 innovation. While a related stream of literature on effort provision (Hannan, Hoffman, and Moser, 2005; Hossain and List, 2012; Armantier and Boly, 2015) and team productivity (Dickinson, 2001; Hong, Hossain, and List, 2015) has found that loss aversion plays a significant role in determining individual's decision to expend resources to achieve a goal, the extent to which loss aversion affects individuals' willingness to persist at innovation remains unexplored.

In summary, we investigate the intersection of two topics: 1) innovation problems in which there is little (or no) evidence about the probability of success, and 2) the behavioral tendency to avoid losses. As we will show in Section 3, 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.

2 Experimental Design

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

Our research question called for the development of an environment in which the structure, rules, and incentives would: 1) lead each subject to develop a multiple-model view of the environment; 2) lead each subject to develop *"foresight"* into a superior solution to the game under one of those models; 3) avoid revealing information about whether the foresight would work to maintain uncertainty; 4) allow researchers to manipulate the structure of incentives to induce loss aversion without otherwise changing the expected earnings from the game; and 5) 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 Maze Game."[1]

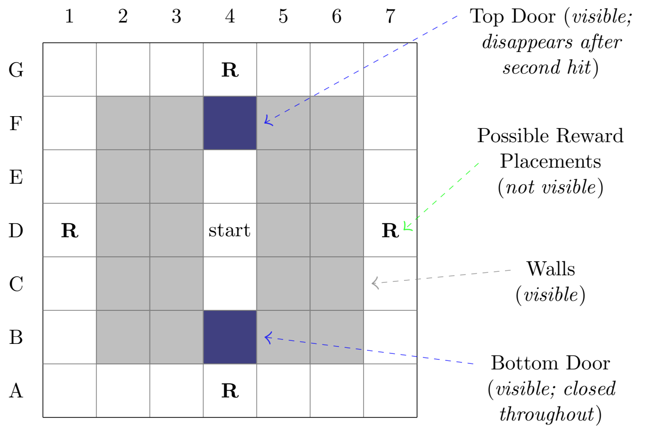
2.1 Rules of the Environment

The Maze Game is played on a 7×7 grid, as illustrated in Figure 1. At the start of the game, participants are placed in the center cell (marked "start") 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 upon each move.

---

[1] Designing an environment to vary incentives independently of loss aversion is a difficult problem and we believe that our solution is a significant new contribution to the field. Other researchers have identified the problem – for example, see footnote 10 in (Elfenbein, Knott, and Croson, 2016): "We chose to shift the RL parameters by +130 to avoid problems associated with loss aversion..." – but this 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. We will open-source the code and online platform for running this game and environment upon publication.

Figure 1: Maze Setup



Notes: Walls are light gray. Doors are dark blue (B4 & F4). Subjects cannot backtrack and cannot hit walls. Subjects can hit a door from any side; if they do, they jump to the start position (D4). The top door (F4) is opened on the second hit and then disappears for the rest of the experiment. The bottom door (B4) remains closed for the entire experiment. Potential reward placements are marked with an **R** (G4, D1, D7, A4). Neither potential nor actual reward placements are displayed to subjects. A cell flashes green when a subject reaches a reward. Grid coordinates (A-G/1-7) are not visible. Appendix B provides step-by-step screen-shots of a sample game.

There are four types of cells that appear on the grid: empty cells (not shaded, in white); walls (shaded in grey); doors (shaded in blue); and rewards (shaded in green once discovered). While we describe grey 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," and subjects are left on their own to infer what they will about the environment (see Appendix A for Game Instructions). Subjects can move through empty cells one move at a time. Subjects cannot move into walls, nor can they 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. Appendix B provides a step-by-step series of screen-shots of a sample game. The bottom door (B4) remains closed at all times throughout the experiment – that is, no matter how many times a subject hits the bottom door, the bottom door always flashes red, and the subject always returns to the start position.[2]

---

[2] We considered using an alternative configuration in which the bottom door would open after perhaps 5, 10, or

The spatial distribution and frequency of rewards is 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 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 – see 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. Because the layout of both the maze (Figure 1) and 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* (i.e., go up and find a reward with relative certainty) and *exploration* (i.e., go down and attempt to open the bottom door – an innovation that might lead to superior outcomes).

Given the layout and rules of the environment described above, we then vary the structure of incentives within the experiment in two ways: first, we vary the placement of rewards to vary the relative benefit of exploration; and second, we vary the direction of earnings to vary exposure to losses. Again, we vary each factor such that a change in one dimension does not affect expected payoffs resulting from differences in the other dimension. In Section 2.2 we describe the configuration of two probability distributions for the positioning of rewards (*Baseline* vs. *Breakthrough*), and in Section 2.3 we describe two earning policies for the avoidance or inducement of *loss aversion* (*Gains* vs. *Losses*).

---

15 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.

2.2 Placement of Rewards: *Baseline* vs. *Breakthrough*

We test two configurations for the placement of rewards: 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). Because rewards are equally dispersed around the board, there is no true incentive to explore in the *Baseline* configuration once the top door has been opened. 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. Because the *Breakthrough* configuration increases the likelihood of obtaining a reward below the bottom door, there is a clear incentive to see if the bottom door can be opened. Table 1 summarizes the two probability distributions for reward placements under *Baseline* and *Breakthrough*.

Table 1: Probability Distribution of Rewards

Location	**Baseline**	**Breakthrough**
Top (G4)	0.25	0.25
Left (D1)	0.25	0
Right (D7)	0.25	0
Bottom (A4)	0.25	0.75

Notes: Rewards are placed on the grid according to the two probability distributions.

The distributions of rewards in Table 1 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 less 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 insight to the *Baseline* vs. *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.*

2.3 Framing of Rewards: *Gains* vs. *Losses*

We test two configurations for the accumulation of earnings: First, 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), while 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, while the reward amount remains constant at \$0.06. The two incentives treatment are presented in Table 2. 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. Additionally, if one considers the expected reward per cycle – i.e., from starting at the start position to finding a reward – then, the expected rewards are equal in absolute value and only differ in sign. Specifically, the expected earning per cycle is \$0.06 in the gains treatment and -\$0.06 in the losses treatment.

Table 2: Payoff Incentives

Parameters	**Gains**	**Losses**
Starting Balance	\$1.50	\$6.50
Moves Per Game	500	500
Cost Per Move	\$0.00	\$0.01
Reward Amount	\$0.06	\$0.06
E[Earn Per Cycle]	\$0.06	\$-0.06
E[Earn Per Game]	\$4.00	\$4.00

Notes: *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. E[Earn Per Cycle] and E[Earn Per Game] are computed for the scenario in which a subject goes up through the Top door (location F4) each and every time. Appendix D provides detailed calculations for expected earnings per cycle and per game.

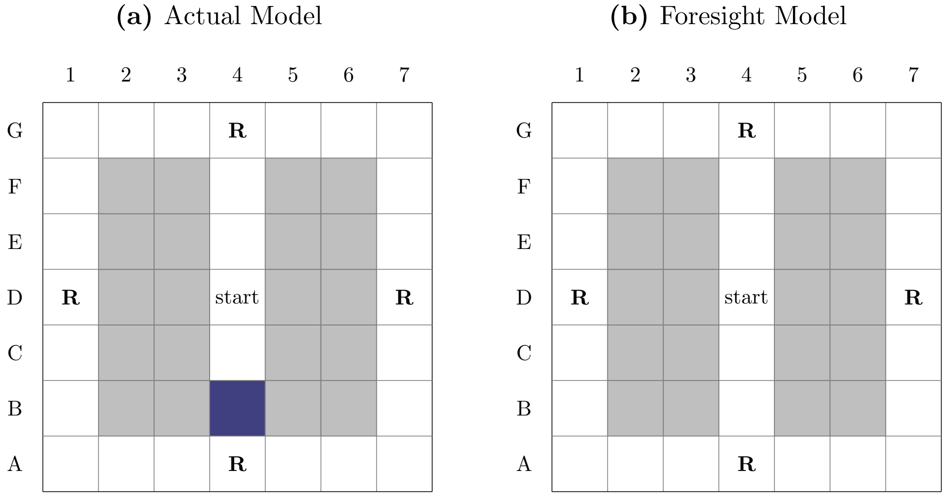
Given the structure of payoffs and expected accumulation of earnings at the end of the game, we conclude that there is no substantive difference between the *Gains* and *Losses* conditions. We summarize this key insight to the *Gains* vs. *Losses* manipulation with the following remark:

Remark 2 *The payoffs for both the Gains and Losses treatments are equivalent.*

3 Predictions

We now turn to develop predictions about the persistence of exploration within the Maze Game. In addition to the experimental manipulations described above (i.e., *Baseline* vs. *Breakthrough*, and *Gains* vs. *Losses*), outcomes in the game will also depend on whether individuals actually are (are not) loss averse, and/or on whether individuals act with (without) foresight. In other words, assumptions about another level of 2 × 2 crossed conditions will affect our predictions. The first alternative (*Loss Averse* vs. *Not Loss Averse*) has already been widely tested in the literature and we include it here for comparison. The second alternative (*Actual* vs. *Foresight*), however, entails competing views (i.e., 'models") that individuals might hold about the environment. Figure 2 illustrates these two views.

Figure 2: Models of the Environment.


(a) Actual Model

(b) Foresight Model

Notes: 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, after which point 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.

In the "Actual Model", 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," the top door is opened but subjects consider the possibility that the bottom door also may be opened through some unknown set of actions; this view is consistent with an intuition of how the rules of environment *might* be, given what happened to the top door.

While we can manipulate the placement and framing of rewards (i.e., *Baseline* vs. *Breakthrough*, and *Gains* vs. *Losses*), we cannot manipulate whether individuals are actually loss averse, or whether individuals view the Actual Model or Foresight Model to be the true model for the environment. Therefore, we pursue two analytical lines of inquiry to predict how being loss averse or not being loss averse, and how having foresight or not having foresight, affect exploration. First, we assume that subjects have full information about the environment and we derive the probability that an agent will explore the breakthrough opportunity (Section 3.1). This analysis provides an important baseline for the study, but subjects clearly do not have full information about the environment and must learn about it through repeated interaction. Therefore (and Second), we use reinforcement learning to observe how agents behave when they do *not* have full information and instead have to learn about the environment from observed experience (Section 3.2).

3.1 Full Information

To understand the role of foresight and loss aversion, we use a classic stochastic model of choice with an assumption of full information (Luce, 1959) and an assumption of loss aversion. Specifically, we use the logit specification to examine the probability that an agent will choose to explore the breakthrough opportunity:

$$p = \frac{e^{V^D}}{e^{V^D} + e^{V^U}} = \frac{1}{1 + e^{V^U - V^D}},$$

where p denotes the probability of choosing the Down action, and $V^D(V^U)$ denotes the value of going Down (Up). This model has been widely used to study individual choice across economics and operations domains (Luce, 1959; McFadden et al., 1973; Su, 2008), as well as outcomes in multi-agent strategic environments (McKelvey and Palfrey, 1995). In the case of loss aversion, we also assume the standard definition of loss aversion: $v(x) < -v(-x)$ for $x > 0$ (Kahneman and Tversky, 1979).

Given full information, Table 3 presents the expected payoffs from playing the game if a subject were to stick with one action throughout their game. We separate the expected earnings in terms of actions at the start given the two models of the environment – the *actual* model (Down, Up), and the *foresight* model (Up, Down). That is, for the actual model, we consider the expected earnings for the environment, as presented in Figure 2a, while for the foresight case, we consider the expected earnings *as if* the bottom door were open, as presented in Figure 2b.

Table 3: Incentives Under Full Information

(a) Actual Model

	Baseline		**Breakthrough**	
	Up	Down	Up	Down
Gains	2.50	0.00	2.50	0.00
Losses	−2.50	−5.00	−2.50	−5.00

(b) Foresight Model

	Baseline		**Breakthrough**	
	Up	Down	Up	Down
Gains	2.50	2.50	2.50	5.00
Losses	−2.50	−2.50	−2.50	0.00

Notes: Expected average earnings for each of the treatment combinations if participant sticks with one action throughout the game. In Panel (b), earnings under the *foresight* model are shown *as if* the bottom door were open.

In the remainder of this section, we derive predictions for our manipulations, given the four possible cases of background conditions that we cannot manipulate directly.

Case 1: Actual, No Loss Aversion.

In the first case, agents do not have foresight and are not loss-averse. In other words, agents consider *actual* payoffs, as shown in Table 3. Then, the difference between payoffs for Up and Down will be the same for gains (subscript g) and for losses (subscript l) for each of the baseline (subscript ba) and breakthrough (subscript br) treatments:

$$V^U_{ba,\,g} - V^D_{ba,\,g} = V^U_{br,\,g} - V^D_{br,\,g} = V^U_{ba,\,l} - V^D_{ba,\,l} = V^U_{ba,\,l} - V^D_{ba,\,l}.$$

Therefore,

$$p_{ba,\,g} = p_{br,\,g} = p_{ba,\,l} = p_{br,\,l}.$$

Thus, in Case 1, there should be no difference in the number of attempts to explore the breakthrough opportunity among the four treatments.

Case 2: Actual, Loss Aversion.

In the second case, agents do not have foresight and are loss-averse. Then, valuation of gains and losses differs, but there is no difference between baseline and breakthrough treatments:

$$V^U_{ba,\,g} - V^D_{ba,\,g} = V^U_{br,\,g} - V^D_{br,\,g} < V^U_{ba,\,l} - V^D_{ba,\,l} = V^U_{br,\,l} - V^D_{br,\,l}.$$

Therefore,

$$p_{ba,\,g} = p_{br,\,g} > p_{ba,\,l} = p_{br,\,l}.$$

Thus, in Case 2, there should be fewer attempts to explore the breakthrough opportunity in the losses than in the gains treatments, but an equal number of attempts when comparing baseline and breakthrough in each incentive condition.

Case 3: Foresight, No Loss Aversion.

In the third case, agents have foresight and are not loss-averse. In other words, agents consider *foresight* payoffs, as shown in Table 3. Then, the difference between payoffs for Down and Up will be the same for gains and for losses but different between the baseline and breakthrough treatments:

$$V^U_{ba,\,g} - V^D_{ba,\,g} = V^U_{ba,\,l} - V^D_{ba,\,l} > V^U_{br,\,g} - V^D_{br,\,g} = V^U_{br,\,l} - V^D_{br,\,l}.$$

Therefore,

$$p_{ba,\,g} = p_{ba,\,l} < p_{br,\,g} = p_{br,\,l}.$$

Thus, in Case 3, there should be fewer attempts to explore the breakthrough opportunity in the baseline than in the breakthrough treatments, but an equal number of attempts when comparing gains and losses in each reward placement condition.

Case 4: Foresight, Loss Aversion.

In the fourth case, agents have foresight and are loss-averse. Then, there will be a difference in valuations between the baseline and breakthrough treatments; additionally there will be a difference in valuations between gains and losses in the breakthrough treatment, but not in the baseline treatment:

$$V^U_{ba,\,g} - V^D_{ba,\,g} = V^U_{ba,\,l} - V^D_{ba,\,l} > V^U_{br,\,g} - V^D_{br,\,g} > V^U_{br,\,l} - V^D_{br,\,l}.$$

Therefore,

$$p_{ba,\,g} = p_{ba,\,l} < p_{br,\,g} < p_{br,\,l}.$$

Thus, in Case 4, there should be fewer attempts to explore the breakthrough opportunity in the baseline than in the breakthrough treatments. Furthermore, in the breakthrough treatment, there should fewer attempts to explore in the gains treatment than in the losses treatment.

To summarize, the four cases above yield distinct and testable predictions under full information with respect to the number of attempts to explore the breakthrough opportunity. In what follows, we relax the full information assumption and develop a more complicated learning model with foresight to provide simulation predictions.

3.2 Reinforcement Learning

In this section, we use a reinforcement learning approach to see how loss aversion may influence exploration behavior. Specifically, we consider Q-learning (Watkins and Dayan, 1992) to simulate the action selection and learning process. We chose to use a machine reinforcement learning approach for several reasons. First, simple reinforcement learning can be applied to the classical multi-armed bandit problem, as well as to more-complex setups such as grid-world problems akin to our environment (Sutton and Barto, 1998). Second, the reinforcement learning approach has been successfully applied to modeling learning behavior in many economic environments (see, for example, Waltman and Kaymak, 2008; Greenwald, Kannan, and Krishnan, 2010). Third, multinomial logit action selection (or Boltzman exploration strategy) is one of the most common action selection procedures in reinforcement learning (Sutton and Barto, 1998). Fourth, we can build on prior research in ways that allow us to incorporate model uncertainty into the learning model (Sutton, 1990). Although we consider a very simple implementation of reinforcement learning, the results support the intuition developed in Section 3.1 on individual choice. In particular, we show that in the presence of foresight and loss aversion, we obtain excess exploration of the option that has not worked.

3.2.1 Learning with Foresight

In order to incorporate foresight into a learning model, we consider the reinforcement learning algorithm with an additional step for planning. Specifically, we take a Q-leaning algorithm (Sutton, 1990) and add the indirect learning step presented in Figure 3. This approach uses a model of the environment, denoted by $Model(s, a)$, to generate a simulated sequence of rewards. More specifically, the model takes the current state, s, and the action taken in that state, a, and returns the next state s' and the reward obtained r.

Figure 3: Q-Learning with Foresight

Algorithm 1 Q-learning with Foresight

```
Initialize $Q(s,a)$ and $Model(s,a)$ for all $s \in S$ and $a \in A$
Do forever:
    $s \leftarrow$ current state
    $a \leftarrow Policy(s,Q)$
    Execute action $a$; observe resultant state, $s'$, and reward, $r$
    $Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma max_{a'} Q(s',a') - Q(s,a)]$
    Repeat N times:
        $s \leftarrow$ random state
        $a \leftarrow$ random action in $s$
        $s', r \leftarrow Model(s,a)$
        $Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma max_{a'} Q(s',a') - Q(s,a)]$
```

Notes: $Model(s,a)$ predicts next state and reward for a state-action pair s, a. We consider two extreme-case scenarios in which the model is either *actual* or *foresight*. $Policy(s,a)$ denotes the action selection policy. For simulations, we implement the *softmax* action selection. Since this is an undiscounted task, we will consider per-step average of rewards (i.e., $\gamma = 1$, and rewards are normalized by the number of steps it takes to get them).

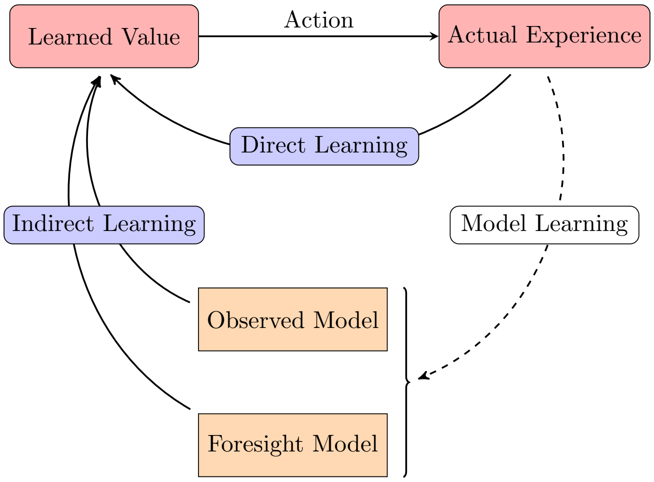
Figure 4 provides some intuition behind the Q-learning algorithm and how we extended it for model learning. The diagram illustrates two channels for learning: a direct channel and an indirect channel. The direct learning channel uses actual experience to update the value function, whereas the indirect learning channel bootstraps from a model to update the value function. The two models we consider are – an (actual) observed model and a (hypothetical) foresight model. Even though the down door never actually opens in our environment, the Foresight Model simulates experience *as if* the down door were open, making the Down strategy more attractive than an Up strategy in the *Breakthrough* treatment.[3] Although it may not be surprising that there would be more attempts on the bottom door in the *Breakthrough* treatment when accounting for Foresight, the consideration is that this approach allows us to simulate the effects of loss aversion. While one might view our setup as a two-armed bandit model (i.e., given that the environment is left/right symmetric and there is no backtracking, an individual will always be faced with the same Up/Down decision at the start of each cycle), any algorithm that learns only from actual experience (such as a bandit model) will never give any weight to the idea that the Down door could be a breakthrough

---

[3] Because the placement of rewards is such that the distance to the reward through the Top door is the same in the *Baseline* and *Breakthrough* treatments, on average, there should be no difference between *Baseline* and *Breakthrough* for an algorithm that considers only actual experience. We do note, however, that there is more variability in *Baseline*, a side-effect that was impossible to control for while maintaining the overall design of our experiment.

opportunity – that is because the bottom door never opens and thus the signals and rewards from going Down will always be the same between the *Baseline* and *Breakthrough* treatments. Furthermore, there should be no difference between the *Baseline* condition and a *Breakthrough* condition based only on actual experience, regardless of whether or not the agent is loss-averse.

Figure 4: Reinforcement Learning with a Channel for Foresight



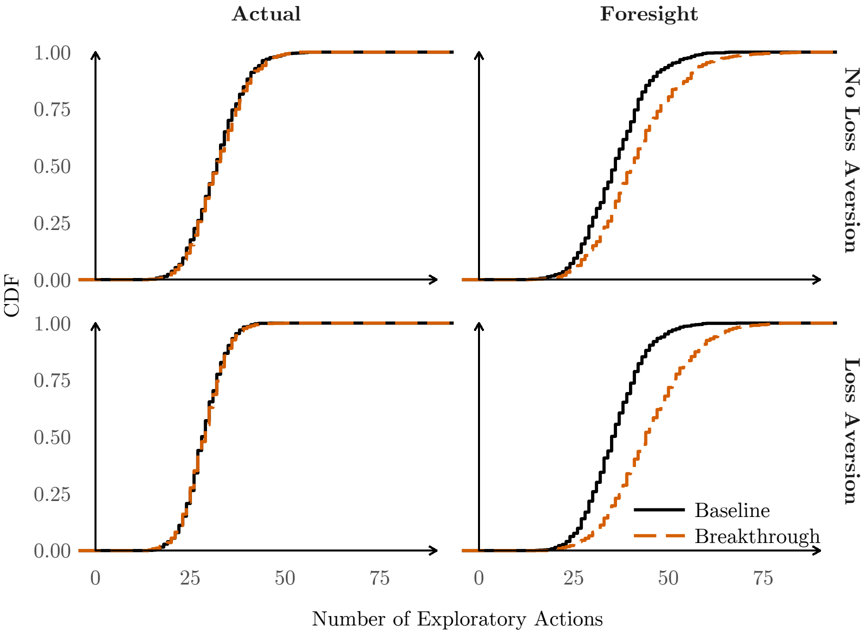
Notes: In our simulations, model learning will involve learning only about the distribution of rewards, and not about the likelihood of the foresight/actual model. A more extensive version of the model may incorporate full model learning.

Our reinforcement learning simulations, therefore, are substantially different from prior research that builds on bandit models and direct learning to explain motivations to innovate (e.g., Manso, 2011). Our simulations account for other sources of indirect learning – such as intuition, analogical reasoning, and other insights – factors that we bundle together as Foresight. We note, however, that we consider only the extreme case in our simulations – that is, we simulate indirect learning as coming either from the observed model or from the foresight model, but do not consider any selection between the models over time. An avenue for future research may be to look at how human subjects transition from the foresight to the actual model in the wake of new signals.

3.2.2 Simulation Results

The following simulations provide an intuition about the potential differences between our experimental conditions. We are interested in the directional differences in the following results, not the magnitude (which would depend on the specific parameters picked). For each of the four treatments, we simulate the learning model 1000 times over a 500-period horizon.[4] Figure 5 presents the comparison between *Baseline* and *Breakthrough* treatments, and Figure 6 presents the comparison between *Gains* versus *Losses* treatments.

Figure 5: Simulation of Baseline vs. Breakthrough for the Losses Condition



Note: Figures plot the cumulative distribution function (CDF) for the number of exploratory actions (attempts on the down door). There is no difference between *Baseline* and *Breakthrough* when a learning agent follows an *Actual Model*, regardless of whether we assume, or do not assume, loss aversion. However, the agent explores more under the *Breakthrough* condition when following a *Foresight Model*, and the effect is even more pronounced when we assume loss aversion. These figures are only for the *Losses* condition where we would expect loss aversion to matter. Complete results for all conditions are reported in Appendix C.

There are two main points to consider with respect to the results in Figure 5: First, the difference between Baseline and Breakthrough is evident under the *foresight* model, but not under the *actual* model; Second, the difference between baseline and breakthrough is greater with loss aversion than it is without loss aversion. Thus, if an agent has the foresight to consider a scenario in which the bottom door is open, then a difference between the two incentives treatments becomes evident. Indeed, the down option is more attractive if one believes that the true model of the environment

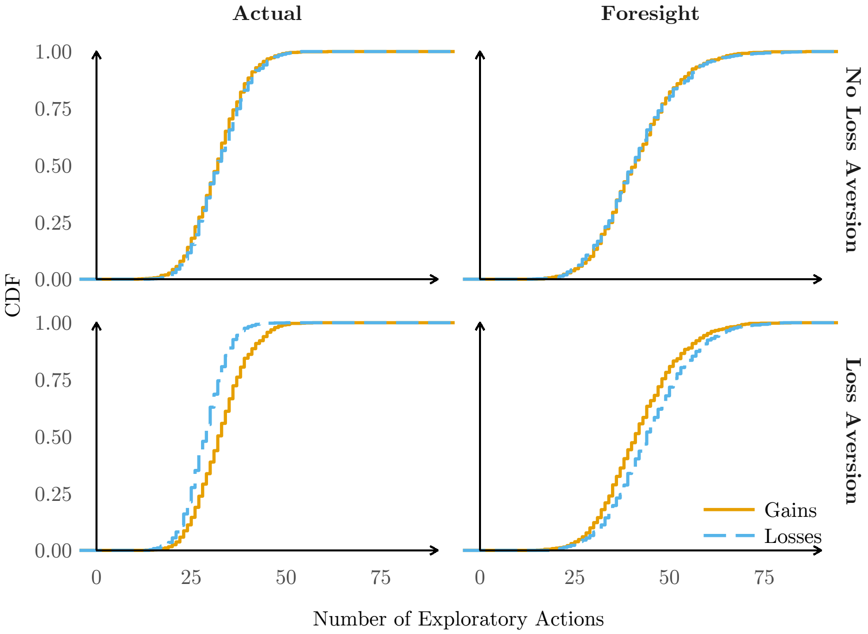
---

[4] Given the finite horizon of our experiment, maximizing earnings over the game is equivalent to maximizing per-step earnings. Therefore, for our model we formulate the reward in terms of per-step reward.

is such that the bottom door can be opened. Moreover, loss aversion makes the down door even more attractive in the breakthrough treatment than it does in the baseline treatment.

Figure 6 presents the comparison of *Gains* versus *Losses*. Here we focus on two main points: First, there is no difference between gains and losses without loss aversion (as can be seen in the top row of Figure 6). Second, loss aversion has the opposite effect on exploration, depending on whether agents consider a *Foresight* model or the *Actual* model (as can be seen in the bottom row of Figure 6). This result highlights a fundamental contrast with respect to the effect of loss aversion on innovation. As shown in the bottom row of Figure 6, the framing of Rewards as losses (and assuming loss aversion by individuals) leads to **less** exploration when agents only learn from an actual model of observed experience, but **more** exploration when agents indirectly learn from a foresight model through hypothetical conjecture.

Figure 6: Simulation of Gains vs. Losses for the Breakthrough Condition



Note: Figures plot the cumulative distribution function (CDF) for the number of exploratory actions (attempts on the down door). There is no difference between an *Actual* and *Foresight* model when the learning agent is **not** loss averse (top row). However, when the agent **is** loss averse (bottom row), then the algorithm attempts exploration **less** when following an *Actual* model, and **more** when following an *Foresight* model. These figures are only for the *Breakthrough* condition where we would expect exploration to be more prevalent. Complete results for all conditions are reported in Appendix C.

We summarize the main takeaways from Section 3 in Figure 7. In this section, we derive predictions for our experiment from both static and dynamic models of choice. Contrary to a multi-armed bandit model (or other model with only direct learning), reinforcement learning with indirect model learning predicts increased exploration when rewards are framed as *Losses* and when individuals are assume to be in fact *Loss Averse*. In the next section, we present the results of human subject experiments to test this prediction.

Figure 7: Summary of Predictions

Model

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

Notes: N denotes the number of attempts on the Bottom door 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.

4 Results

To test the above predictions, we recruited three hundred subjects from the Amazon Mechanical Turk labor market ("M-Turk") to participate in the experiment, and randomly assigned each subject into one of our four experimental conditions for the duration of the experiment (see Table 4). The M-Turk population is now widely used to recruit subjects for social science experiments (Paolacci, Chandler, and Ipeirotis, 2010; Buhrmester, Kwang, and Gosling, 2011; Horton, Rand, and Zeckhauser, 2011; Berinsky, Huber, and Lenz, 2012; Rand, 2012; Goodman, Cryder, and Cheema, 2012). Participants were told in an advertisement that they could earn up to \$4.00 as part of the experiment, and after recruitment participants where told that they would earn a base payment of two dollars (\$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) and on average the experiment lasted approximately 20 minutes. The effective average hourly earnings of \$11.58 was a relatively high compensation rate for M-Turk workers (see Horton and Chilton, 2010).

Table 4: Count of Observations

	Baseline	**Breakthrough**
Gains	50	100
Losses	50	100

Participants were restricted to M-Turk workers located within the U.S., with a "masters" status and an approval rating of at least 90% on prior work conducted at M-Turk. Experiments were conducted online on a private web server. We ran experiments between the conditions simultaneously in order to avoid potential differences in populations related to the hour of the day and/or the day of the week (i.e., any such population differences should be randomly assigned equivalently across the conditions). Table 5 presents a randomization check between the *Gains* and *Losses* conditions, as well as the educational and demographic characteristics for the sample.

Table 5: Randomization Check and Demographics by Condition

	Female	Male	Age $\geq$ 34	Stat $\geq$ 2	Econ $\geq$ 2	Biz $\geq$ 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%

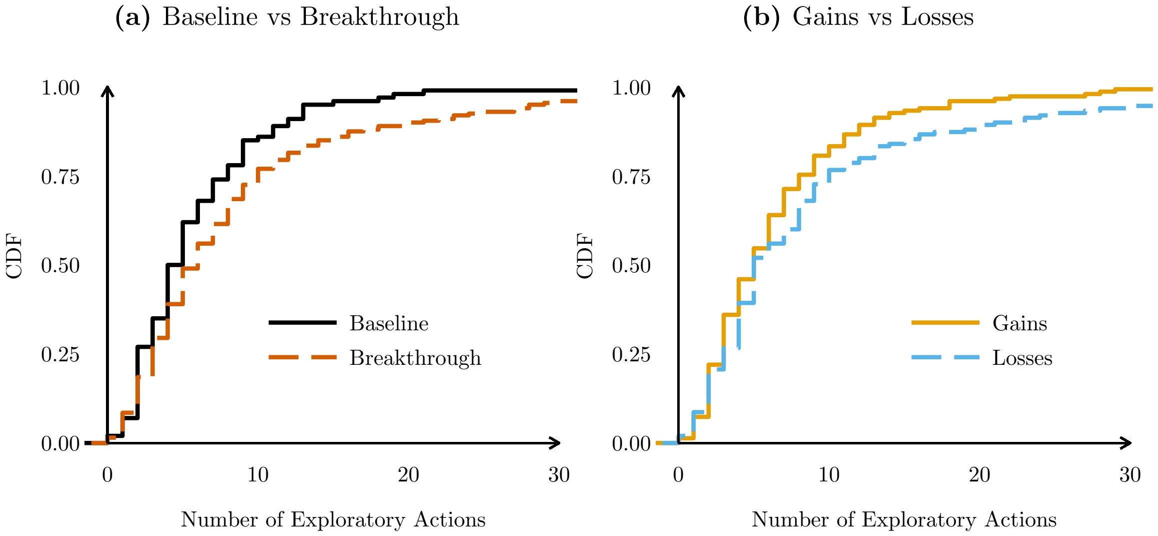
Notes: *Stats* = percentage of subjects with more than 1 course in statistics. *Econ* = percentage of subjects with more than 1 course in economics. *Biz* = percentage of subjects with more than 1 course in business. *College* = percentage of subjects with a college degree.

The rest of this section is organized as follows: First, we present the raw data on the number of exploratory actions by treatment. Second, we conduct permutation tests on the key statistics of the data. Third, we present two regression as robustness tests to control for heterogeneity in the random placement of rewards throughout the experiment.

4.1 Data

Figure 8 plots the cumulative distribution function for the number of exploratory actions (hits on the Down door) by treatment. Two key observations stand out: 1) Subjects tended to explore the unproven path more frequently in the *Breakthrough* treatment (compared to *Baseline*); and 2) Subjects tended to explore the unproven path more in the *Losses* treatment (compared to *Gains*). Thus, qualitatively, we find support for the predictions from the reinforcement learning in Section 3. We also note that greater exploration under *Breakthrough* than *Baseline* is a basic requirement for our experiment – if, upon discovering an advantageous spatial distribution in the placement of rewards, individuals did not try to take a shorter path to those rewards in the game, then the game itself would not be a valid test for exploration.[5] In the following two sections, we test the statistical significance of these findings with permutation tests and regression analysis.

Figure 8: Cumulative Distribution of Exploratory Actions



Notes: Each figure plots the cumulative distribution function (CDF) for the number of exploratory actions (attempts on the blocked, downward door).

---

[5] As a robustness check, we also ran an *Intermediate* configuration that weighted rewards toward the bottom door more than *Baseline*, but less than *Breakthrough*. As expected, effects for the *Intermediate* condition had lower magnitudes than the *Breakthrough* condition due to lower-powered incentives, and all effects were consistent with the effects reported in the paper (results available on request).

4.2 Permutation Tests

Table 6 presents the average number of exploratory actions per treatment, as well as results for statistical comparisons between conditions made through two-tailed permutation tests. Permutation tests are non-parametric randomization tests, in which the distribution of the test statistic is obtained through random permutation of labels for treatment among observations (Phipson and 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* (see Table 4). 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, 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 et al., 2004; Phipson and Smyth, 2010). The statistical significance of a comparison of means between cells in Table 6 is then marked by $>$, $\gg$, and $\ggg$, with statistical significance at 0.10, 0.05, and 0.01.

Table 6: Average Number of Exploratory Actions

	Baseline		**Breakthrough**
Gains	5.2 (0.569)	$\ll$	7.27 (0.677)
	$\wr$		⩓
Losses	6.56 (0.802)	$\ll$	10.82 (1.389)

Notes: Bootstrapped standard errors are in parentheses. $>$, $\gg$, and $\ggg$ denote significance at the 0.10, 0.05, and 0.01 levels, respectively. $\sim$ denotes no significance. p-values are determined using two-tailed permutation tests. The unit of observation is a unique subject.

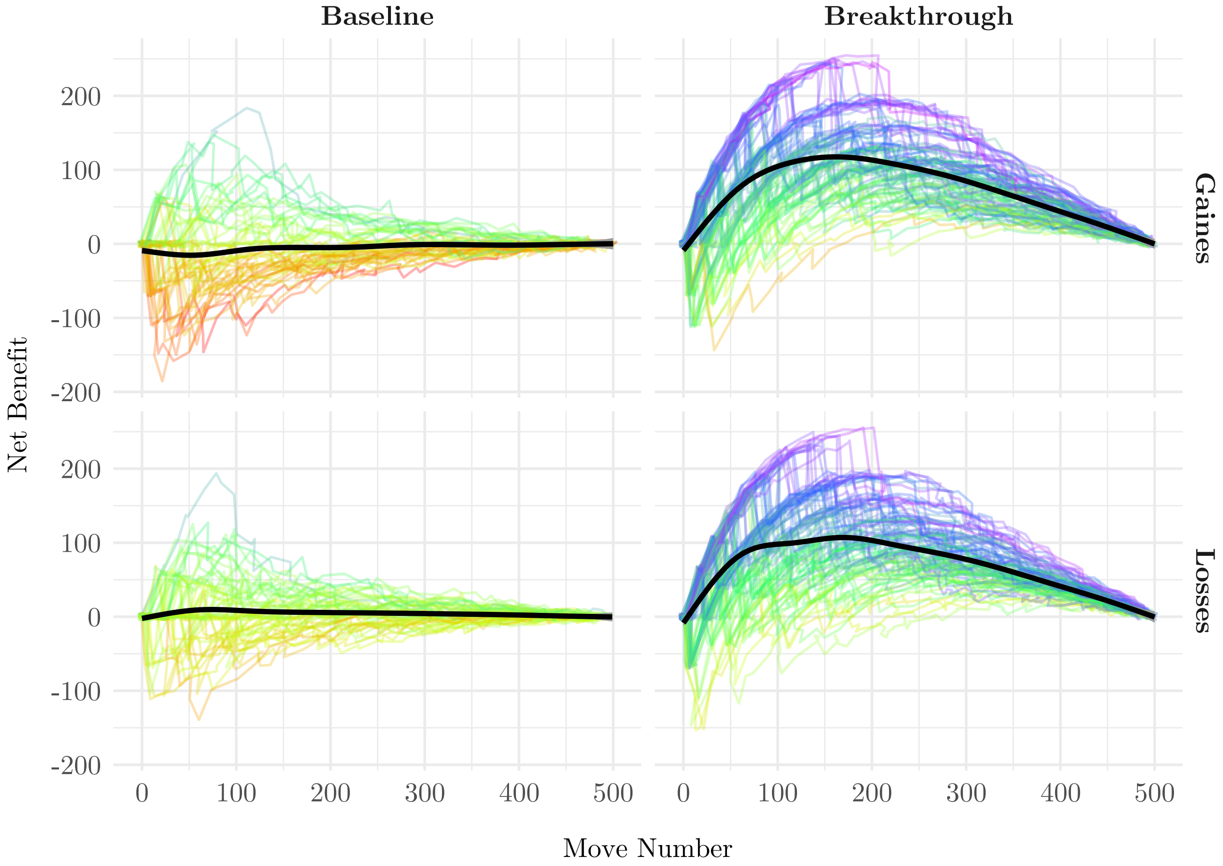
As reported in Table 6, there are two key results to consider. First, we find that subjects were significantly more likely to explore the unproven path (bottom door) in the *Breakthrough* treatment, regardless of whether they are under a condition of *Gains* or *Losses*. Although this is unsurprising, it confirms that subjects do act with foresight in the experiment, and that we therefore achieved our objective of inducing the innovation 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 (attempt 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 reinforcement learning predictions in Section 3 for loss-averse agents who act with foresight.

It is important, however, to point out that participants may choose different actions throughout the experiment, and 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 random reward placements from the baseline treatment might actually be more representative of a breakthrough treatment (and vice versa). Although it would be unlikely for that to happen, on average (and the descriptive statistics reported above suggest that it did not), our research design is "between-subjects," and so it is good practice to perform a randomization check and 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 group. Second, we construct matched samples where subjects from different treatment conditions for loss aversion are paired up and compared to each other based on the exact sequence of reward placements (i.e. "signals") that they observe.

4.3 Regression Analysis

Consider the following scenario. Each of two participants—one in the *Baselin*e treatment and the other in the *Breakthrough* treatment—finds the first reward at location A4. In other words, they receive the same first signal about the distribution of rewards. Thus, although they are assigned into two different reward distribution treatments, from their perspective, the two situations are identical. To quantify the similarity of the two scenarios, we take two approaches. In the first approach, we develop a "Net Benefit Conditional on Foresight" measure. Specifically, we find the difference between the expected reward that an agent would obtain by opening the bottom door (location B4) and the expected reward of going through the top door (location F4) for the remainder of the experiment.[6] Figure 9 presents this potential net benefit measure for each of the four treatments.

Figure 9: Net Benefit, Conditional on Foresight Model Being True



Notes: The thin, colored lines represent the expected difference for the remainder of the experiment between using the *Down* door (B4) door and the *Up* door (F4) under the *Foresight* model (i.e., *if* the Down door were open). Expected values are obtained for the realized sequence of signals by applying Bayes' rule starting with a uniform prior. Each subject therefore has his or her own "trajectory" of the expected net benefit over time of opening the bottom door. *Thick black lines* represent the average potential net benefit for each treatment.

As expected, there is substantial heterogeneity within each treatment in terms of the overall placement of rewards observed by subjects. Furthermore, there is a substantial overlap between the *Baseline* and *Breakthrough* treatments – observations that were statistical noise in a between-

---

[6] We calculated expectations based on updated probabilities using Bayes' Rule, starting with a uniform prior.

subjects permutation test. But also as expected, the average measure of the expected net benefit of opening the bottom door (depicted by the bold black lines in Figure 9) is consistent with the parameters that we set for the distribution of rewards between the treatments. Knowing the potential benefit of opening the bottom door, however, allows us to compare subjects with different sequences of signals at different points in time. It also allows us to add demographic and educational characteristics into a regression framework. Specifically, we run a logistic regression on the probability of attempting to go through the bottom door on the treatment conditions, the net benefit measure, and the demographic characteristics. Next, we determine whether the subject went down or up for each treatment and code the dependent variable one if they went down, and zero if they went up. The results for our first regression are presented in Table 7.

Table 7: Logit Model of Exploration, Given Potential Net Benefit

	(1)	(2)	(3)	(4)	(5)
Losses	1.3737**			1.3953**	1.4229**
	(0.141)			(0.150)	(0.154)
Net Benefit		1.0017**		1.0013*	1.0013*
		(0.001)		(0.001)	(0.001)
Moves			0.9956***	0.9956***	0.9955***
			(0.000)	(0.000)	(0.000)
Sex					0.8144
					(0.098)
Age					0.9932
					(0.005)
College					1.1062
					(0.128)
Statistics					1.0634
					(0.077)
Biz & Econ					0.9447*
					(0.022)
Constant	0.1709***	0.1882***	0.4798***	0.3824***	0.5482***
	(0.010)	(0.010)	(0.022)	(0.029)	(0.097)
Observations	20,089	20,089	20,089	20,089	19,439
Log likelihood	-9072	-9094	-8565	-8519	-8199

Notes: Dependent variable is *Exploration*, coded 1 when going down and 0 when going up. The model controls for the "net benefit" of exploration, conditional on *Foresight*. Coefficients are reported as odds ratios. Robust standard errors in parentheses. $+ \ p < 0.10$, $* \ p < 0.05$, $** \ p < 0.01$, and $*** \ p < 0.001$.

In Table 7, we find support for our main result from the previous section. Consistent with the permutations tests, we find a strong causal association between being in the losses condition and the likelihood of attempting to open the bottom door. Additionally, we find that participants are more likely to attempt the breakthrough option if the net benefit variable is higher, and that participants are less likely to attempt the breakthrough action over time. However, we find no effect of demographic variables. Finally, as indicated by the highly significant coefficients on $Moves$, there also is a tendency to eventually give up, as the likelihood of going down decreases on average as one progresses through the game.

For our second regression approach, we match subjects based on the realized sequence of signals that they observed throughout the game. Specifically, we build a cycle-by-cycle sequence of signals (reward locations) observed by each subject. We then matched each subject from the losses treatment to a subject from the gains treatment at the same cycle within the game and with the exact same sequence of prior observed signals. If there was more than one matching subject, then we picked one at random. If there were no matches, then we skipped the observation. Next, we determined whether the subject went down or up for each treatment, and coded the dependent variable one if they went down, and zero if they went up. Finally, we ran a logistic regression for the likelihood of exploration, clustering standard errors at the level of the subject to adjust standard errors for repeated observations.

The results are presented in Table 8. In Column 1 we find that subjects with incentives structured for loss aversion were significantly more likely to explore. In Column 2 we control for the number of signals received for that observation, and again find that the likelihood of going down decreases on average as one progresses through the game, but that the aforementioned effect of loss aversion does not change. In column 3, we subsample down to just those observations starting with the 10th signal, going through the 40th signal, on the logic that subject may spend the early part of the game learning, but give up toward the end. Indeed the treatment effect of loss aversion is stronger in Column 3, both in terms of magnitude and statistical significance.

Table 8: Logit Model of Exploration, Given Matched Samples by Signal

	(1)	(2)	(3) Subsample
Losses	1.2654*	1.2686*	1.4870**
	(0.142)	(0.143)	(0.220)
Signal		0.9814***	0.9793***
		(0.005)	(0.005)
Constant	0.2014***	0.3117***	0.2158***
	(0.009)	(0.035)	(0.029)
Observations	9,306	9,306	5,876
Log likelihood	-4452	-4402	-2321

Notes: Dependent variable is *Exploration*, coded 1 when going down and 0 when going up. Coefficients are reported as odds ratios. Robust standard errors in parentheses. $+$ $p < 0.10$, $*$ $p < 0.05$, $**$ $p < 0.01$, and $***$ $p < 0.001$.

5 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 learning model and find that it helps to explain human exploration behavior in an uncertain environment with possibility of innovation. Moreover, our results suggest that individuals explore more when they are reminded of the incremental cost of their actions, a counter-intuitive result for research on entrepreneurship and innovation, but one that directly extends findings on loss aversion and predictions from prospect theory.

Our results have implications for how incentives can be framed to increase persistence by actors when attempting to make a breakthrough innovation. There are many contexts where the expected outcome from innovation is low (*ex-ante*), but the *potential* benefit is high (*ex-post*, conditional on success). Examples include the search for new sources of alternative energy (e.g., biofuels, nuclear fusion, artificial photosynthesis), and the search for medical breakthroughs (e.g., a cure for cancer, neural interfaces, and anti-aging).[7] While the literature on *tolerance for failure* (Azoulay,

---

[7] The road to the fountain of youth is littered with failure, but a surprising amount of effort continues in that vein. Google-backed Calico Labs has a reported $1.5 billion to invest in the search for an anti-aging breakthrough.

Graff Zivin, and Manso, 2011; Manso, 2011) has emphasized the need for resource providers to tolerate early failure and reward long-term success, our behavioral findings suggest that an optimal incentive structure may need to induce individuals to experience the urgency of ongoing losses, even if (or even as) individuals are shielded from the consequences of being wrong.

Our results may help to explain certain organizational practices that attempt to balance a tolerance for failure with salient losses. Raynor (2007), for example, discusses the practice of "Requisite Uncertainty" wherein top managers commit resources to a range of highly-uncertain bets and business unit managers are assigned to execute on each of those bets. In the case of failure, a business unit manager will be reassigned within the organization, but managers nevertheless feel losses along the way through the explicit reporting of their business unit accounts. Interestingly, although research on the escalation of commitment (Staw, 1976, 1981; Staw and Ross, 1978) has been generally viewed in a negative light, there are contexts where resource providers may want to escalate commitment to a strategy on purpose. Our model of reinforcement learning with foresight "endogenizes" escalation of commitment, making higher levels of commitment an outcome of the model.

Our results may also help to explain results from other experimental studies on innovation. For example, Kagan et al. (forthcoming) find in an ideation-execution experiment that subjects persist at the ideation stage to a sub-optimal point when they are allowed to trade-off time in the execution stage with time in the ideation phase. Because time lost in the execution stage induces a loss framework, loss aversion in conjunction with foresight may lead subjects to over-persistence at something that has not worked before.

Finally, we note that there are several limitations to our study. First, there is substantial heterogeneity among our subjects in terms of the number of attempts on the bottom door. A potential avenue to explore in future research is the extent to which heterogeneity in the willingness to persist is driven by the heterogeneity in individual loss aversion, and/or how much is driven by other individual characteristics. Second, we obtained our predictions using two extreme cases. Future work could investigate the transition between the two by incorporating the likelihood that the foresight model is valid after a sequence of attempts. Third, we use simulations of the learning model to motivate our predictions. An alternative approach could be to estimate the learning model based on the data.

References

Acharya, V. V., R. P. Baghai, and K. V. Subramanian (2014): "Wrongful discharge laws and innovation," *Review of Financial Studies*, 27(1), 301–346.

Acharya, V. V., and K. V. Subramanian (2009): "Bankruptcy codes and innovation," *Review of financial studies*, 22(12), 4949–4988.

Aghion, P., J. Van Reenen, and L. Zingales (2013): "Innovation and institutional ownership," *The American Economic Review*, 103(1), 277–304.

Armantier, O., and A. Boly (2015): "Framing of incentives and effort provision," *International Economic Review*, 56(3), 917–938.

Azoulay, P., J. S. Graff Zivin, and G. Manso (2011): "Incentives and creativity: evidence from the academic life sciences," *The RAND Journal of Economics*, 42(3), 527–554.

Barberis, N., M. Huang, T. Santos, et al. (2001): "Prospect Theory and Asset Prices," *The Quarterly Journal of Economics*, 116(1), 1–53.

Berinsky, A. J., G. A. Huber, and G. S. Lenz (2012): "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk," *Political Analysis*, 20(3), 351–368.

Buhrmester, M., T. Kwang, and S. D. Gosling (2011): "Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?," *Perspectives on Psychological Science*, 6(3), 3–5.

Dickinson, D. L. (2001): "The carrot vs. the stick in work team motivation," *Experimental Economics*, 4(1), 107–124.

Ederer, F., and G. Manso (2013): "Is pay for performance detrimental to innovation?," *Management Science*, 59(7), 1496–1513.

Elfenbein, D. W., A. M. Knott, and R. Croson (2016): "Equity stakes and exit: An experimental approach to decomposing exit delay," *Strategic Management Journal*.

Ernst, M. D., et al. (2004): "Permutation methods: a basis for exact inference," *Statistical Science*, 19(4), 676–685.

Gneezy, U., S. Meier, and P. Rey-Biel (2011): "When and why incentives (don't) work to modify behavior," *The Journal of Economic Perspectives*, 25(4), 191–209.

Good, P. (2013): *Permutation tests: a practical guide to resampling methods for testing hypotheses*. Springer Science & Business Media.

Goodman, J. K., C. E. Cryder, and A. Cheema (2012): "Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical-Turk Samples," *Journal of Behavioral Decision Making*, 26, 213–224.

Greenwald, A., K. Kannan, and R. Krishnan (2010): "On evaluating information revelation policies in procurement auctions: A Markov decision process approach," *Information Systems Research*, 21(1), 15–36.

Guadalupe, M., O. Kuzmina, and C. Thomas (2012): "Innovation and foreign ownership," *The American Economic Review*, 102(7), 3594–3627.

Hannan, R. L., V. B. Hoffman, and D. V. Moser (2005): "Bonus versus penalty: does contract frame affect employee effort?," in *Experimental business research*, pp. 151–169. Springer.

Hansen, L. P., and T. J. Sargent (2001): "Robust control and model uncertainty," *The American Economic Review*, 91(2), 60–66.

Hong, F., T. Hossain, and J. A. List (2015): "Framing manipulations in contests: a natural field experiment," *Journal of Economic Behavior & Organization*, 118, 372–382.

Horton, J., and L. Chilton (2010): "The Labor Economics of Paid Crowdsourcing," in *Proceedings of the 11th ACM Conference on Electronic Commerce*, EC '10, pp. 209–218, New York, NY, USA. ACM.

Horton, J. J., D. G. Rand, and R. J. Zeckhauser (2011): "The online laboratory: conducting experiments in a real labor market," *Experimental Economics*, 14(3), 399–425.

Hossain, T., and J. A. List (2012): "The behavioralist visits the factory: Increasing productivity using simple framing manipulations," *Management Science*, 58(12), 2151–2167.

Howell, S. T. (2017): "Financing innovation: evidence from R&D grants," *The American Economic Review*, 107(4), 1136–1164.

Kagan, E., S. Leider, and W. S. Lovejoy (forthcoming): "Ideation-execution transition in product development: Experimental analysis," *Management Science*.

Kahneman, D., J. L. Knetsch, and R. H. Thaler (1990): "Experimental tests of the endowment effect and the Coase theorem," *Journal of political Economy*, pp. 1325–1348.

Kahneman, D., and A. Tversky (1979): "Prospect theory: An analysis of decision under risk," *Econometrica: Journal of the econometric society*, pp. 263–291.

Kerr, W. R., R. Nanda, and M. Rhodes-Kropf (2014): "Entrepreneurship as experimentation," *The Journal of Economic Perspectives*, 28(3), 25–48.

Knight, F. H. (1921): "Risk, uncertainty and profit," *New York: Hart, Schaffner and Marx.*

Lerner, J., and J. Wulf (2007): "Innovation and incentives: Evidence from corporate R&D," *The Review of Economics and Statistics*, 89(4), 634–644.

Lim, A. E., J. G. Shanthikumar, and Z. M. Shen (2006): "Model uncertainty, robust optimization, and learning," in *Models, Methods, and Applications for Innovative Decision Making*, pp. 66–94. INFORMS.

Luce, R. D. (1959): "Individual Choice Behavior a Theoretical Analysis," *John Wiley and sons.*

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

McFadden, D., et al. (1973): "Conditional logit analysis of qualitative choice behavior," .

McKelvey, R. D., and T. R. Palfrey (1995): "Quantal response equilibria for normal form games," *Games and economic behavior*, 10(1), 6–38.

Nanda, R., and M. Rhodes-Kropf (2016): *Innovation Policy and the Economy*vol. 16, chap. Financing Entrepreneurial Experimentation. National Bureau of Economic Research.

Owen, H., and I. Gale (2003): "Optimal design of research contests," *The American Economic Review*, 93(3), 646–671.

Paolacci, G., J. Chandler, and P. G. Ipeirotis (2010): "Running experiments on Amazon Mechanical Turk," *Judgment and Decision Making*, 5(5), 411–419.

Phipson, B., and G. K. Smyth (2010): "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).

Rand, D. G. (2012): "The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments," *Journal of Theoretical Biology*, 299, 172 – 179.

Raynor, M. E. (2007): *The strategy paradox: Why Committing to Success Leads to Failure and What to Do About It.* Currency Doubleday.

Seru, A. (2014): "Firm boundaries matter: Evidence from conglomerates and R&D activity," *Journal of Financial Economics*, 111(2), 381–405.

Staw, B. M. (1976): "Knee-deep in the big muddy: A study of escalating commitment to a chosen course of action," *Organizational behavior and human performance*, 16(1), 27–44.

——— (1981): "The Escalation of Commitment to a Course of Action," *Academy of Management Review*, 67(4), 577–587.

Staw, B. M., and J. Ross (1978): "Commitment to a policy decision: A multi-theoretical perspective," *Administrative Science Quarterly*, pp. 40–64.

Su, X. (2008): "Bounded rationality in newsvendor models," *Manufacturing & Service Operations Management*, 10(4), 566–589.

Sutton, R. S. (1990): "Integrated architectures for learning, planning, and reacting based on approximating dynamic programming," in *Proceedings of the seventh international conference on machine learning*, pp. 216–224.

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

Tian, X., and T. Y. Wang (2014): "Tolerance for failure and corporate innovation," *Review of Financial Studies*, 27(1), 211–255.

Tversky, A., and D. Kahneman (1991): "Loss aversion in riskless choice: A reference-dependent model," *The quarterly journal of economics*, pp. 1039–1061.

——— (1992): "Advances in prospect theory: Cumulative representation of uncertainty," *Journal of Risk and uncertainty*, 5(4), 297–323.

Waltman, L., and U. Kaymak (2008): "Q-learning agents in a Cournot oligopoly model," *Journal of Economic Dynamics and Control*, 32(10), 3275–3293.

Watkins, C. J., and P. Dayan (1992): "Q-learning," *Machine learning*, 8(3-4), 279–292.

Appendix A: Experimental Instructions

Today's experiment will last about 30 minutes, although the time may vary based on the speed of the Internet. Every participant will earn a base payment of two dollars ($2.00). You also may earn a substantial bonus depending on decisions you make within the game.

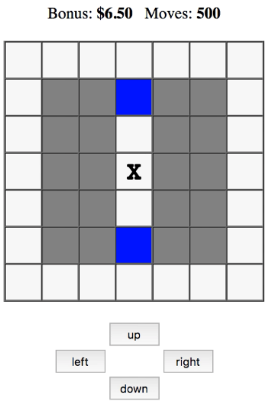
In the experiment, you will navigate through a maze and collect rewards. You can observe your location, your point balance, and the remaining number of moves available to you in the game. You will start the game with 500 moves. You must complete all 500 moves in order to be paid a bonus. Rules for your 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.

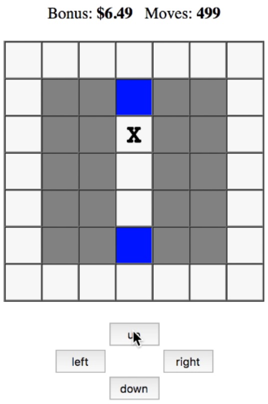
At the end of the experiment, your point balance will be converted into a bonus payment at the rate of 100 points per dollar. The bonus payment is in addition to the base payment of two dollars ($2.00) you receive for participating in the experiment. For example, the following point balances at the end of the game would be converted into the following bonus payments:

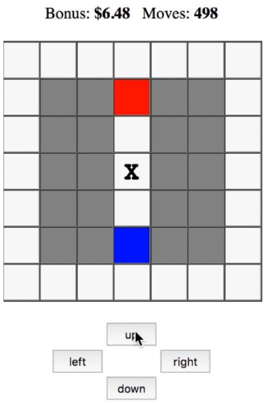
- 100 points = $1.00
- 200 points = $2.00
- 300 points = $3.00
- 400 points = $4.00

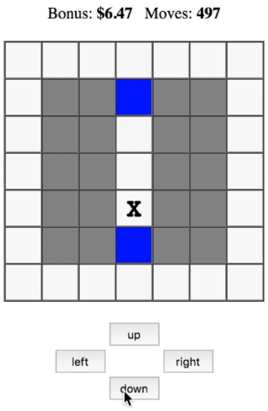
You will receive a confirmation code at the end of the experiment. **You must submit the confirmation code to receive payment.** If you are unable to complete the experiment, please exit the game and indicate the nature of the problem. Please allow a day or two for the processing of your bonus.

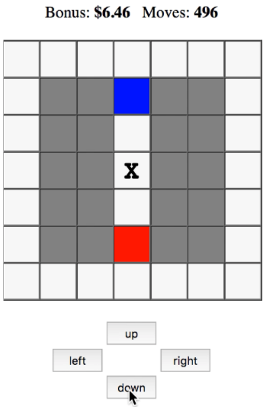
Appendix B: Screen-Shots from a Simulated Game (First 31 Moves)

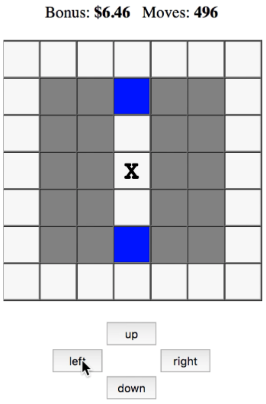


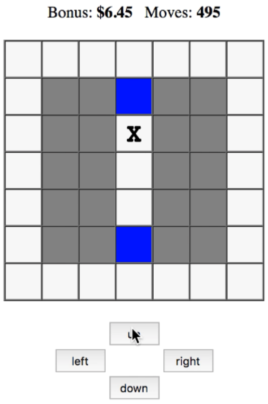


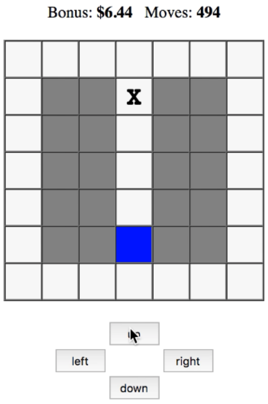


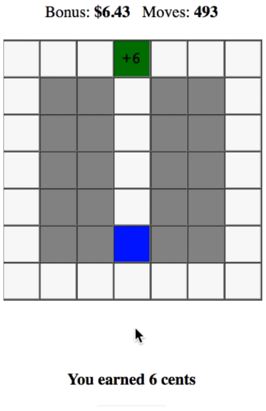




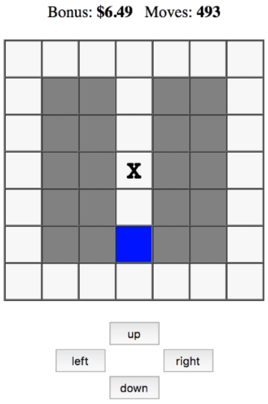


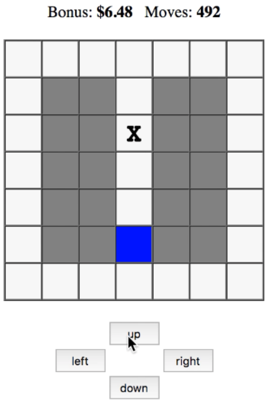


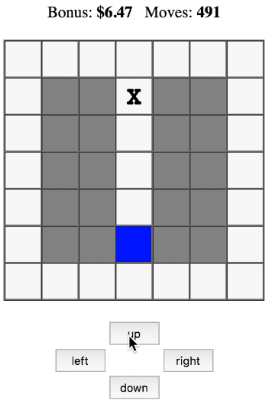


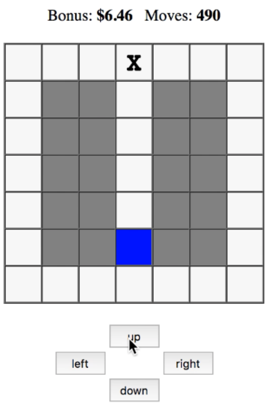


You earned 6 cents

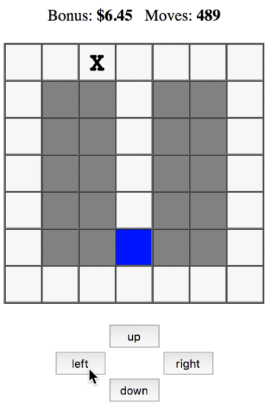




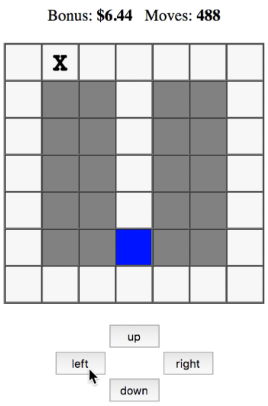




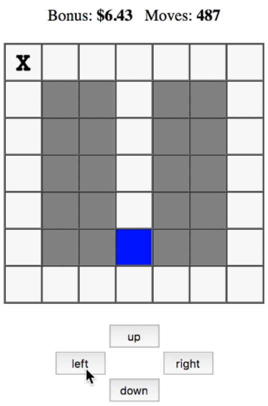
Bonus: **$6.46** Moves: **490**



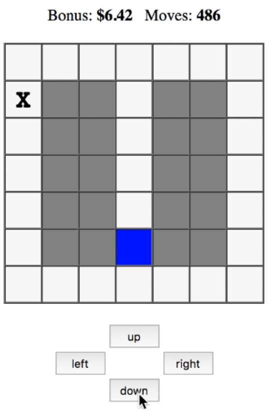
Bonus: **$6.45** Moves: **489**



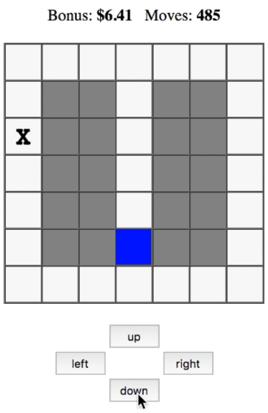
Bonus: **$6.44** Moves: **488**



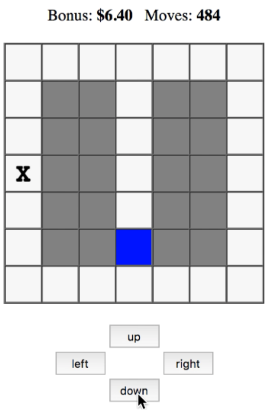
Bonus: **$6.43** Moves: **487**



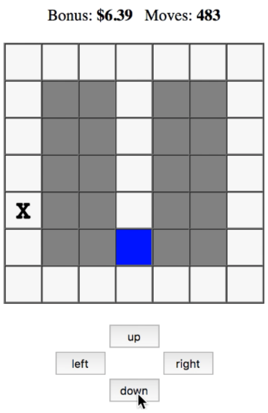
Bonus: **$6.42** Moves: **486**



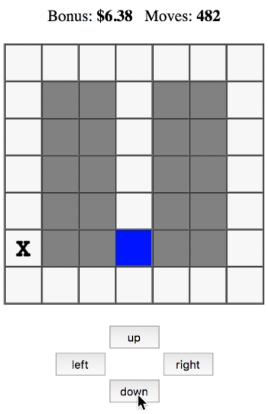
Bonus: **$6.41** Moves: **485**



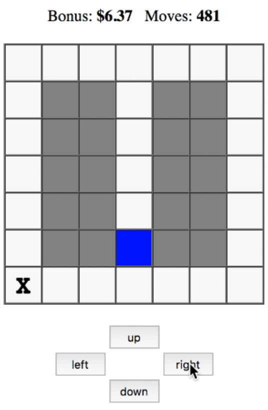
Bonus: **$6.40** Moves: **484**



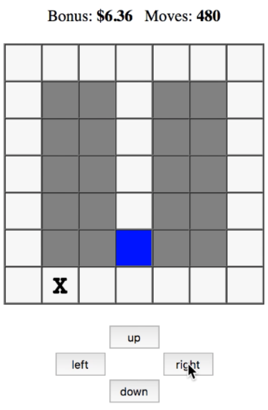
Bonus: **$6.39** Moves: **483**



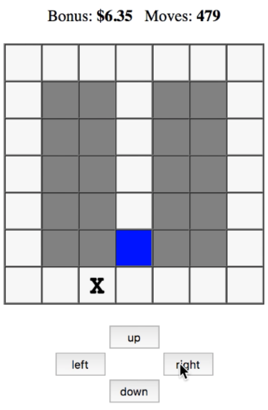
Bonus: **$6.38** Moves: **482**



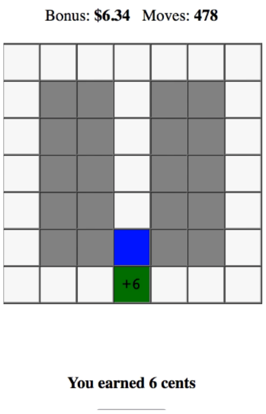
Bonus: **$6.37** Moves: **481**



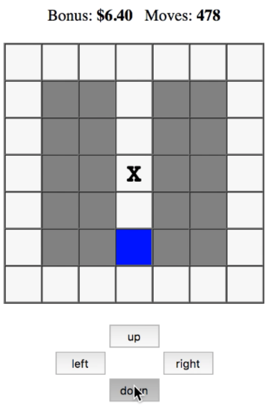
Bonus: **$6.36** Moves: **480**

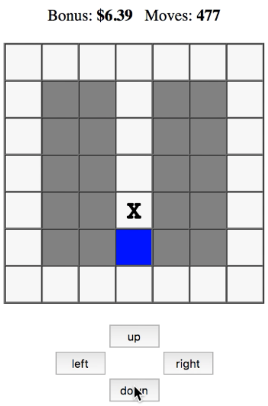


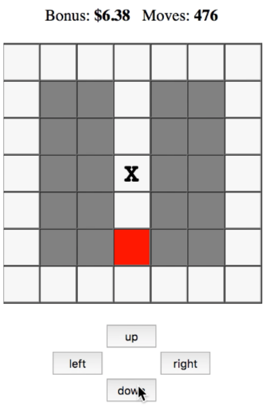
Bonus: **$6.35** Moves: **479**

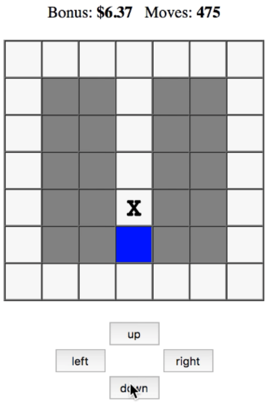

Bonus: **$6.34** Moves: **478**

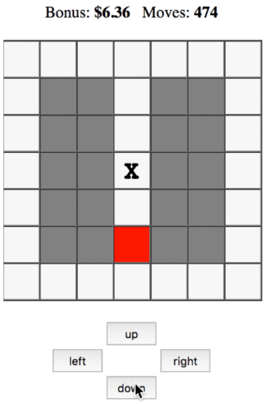
You earned 6 cents

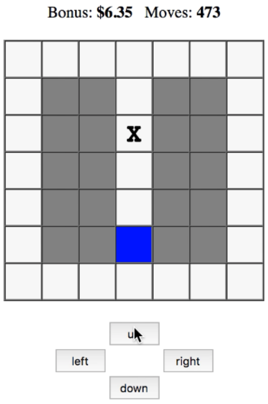

Bonus: **$6.40** Moves: **478**

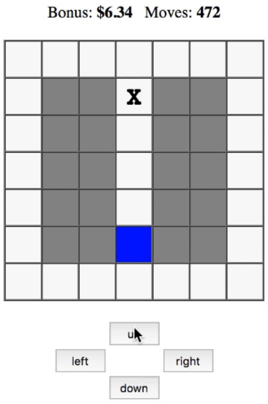

Bonus: **$6.39** Moves: **477**

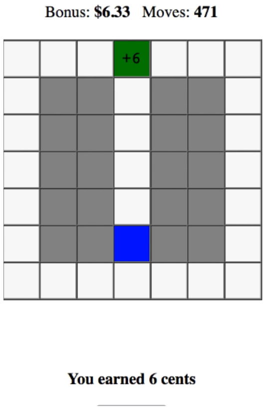

Bonus: **$6.38** Moves: **476**


Bonus: **$6.37** Moves: **475**

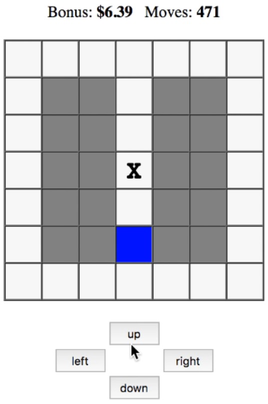

Bonus: **$6.36** Moves: **474**

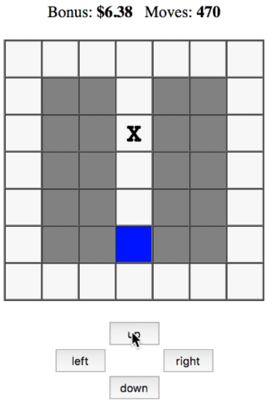

Bonus: **$6.35** Moves: **473**

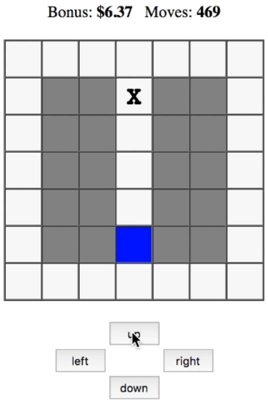

Bonus: **$6.34** Moves: **472**


Bonus: **$6.33** Moves: **471**

You earned 6 cents

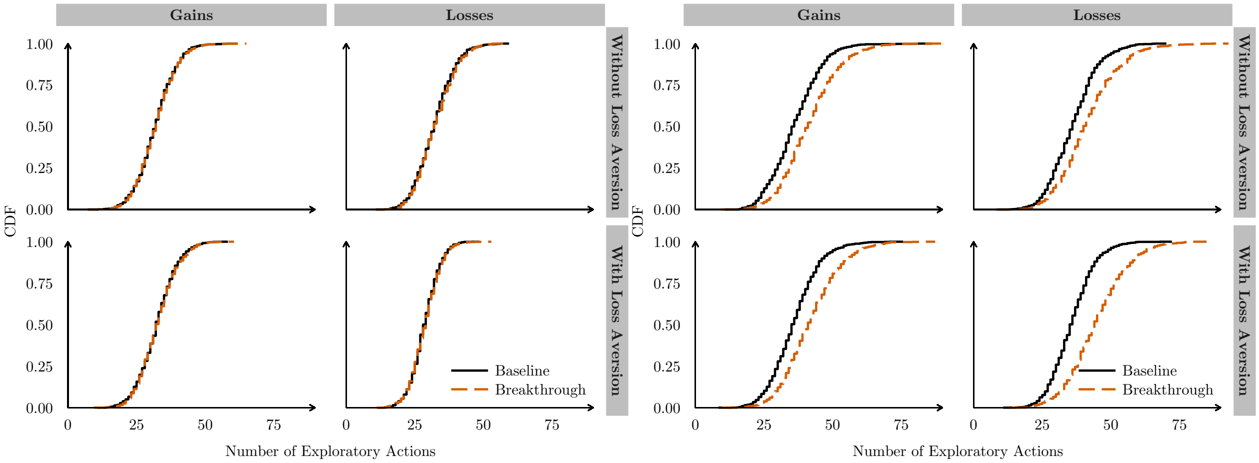

Bonus: **$6.39** Moves: **471**


Bonus: **$6.38** Moves: **470**


Bonus: **$6.37** Moves: **469**

Appendix C: Supporting Figures for Reinforcement Learning and Simulation Results

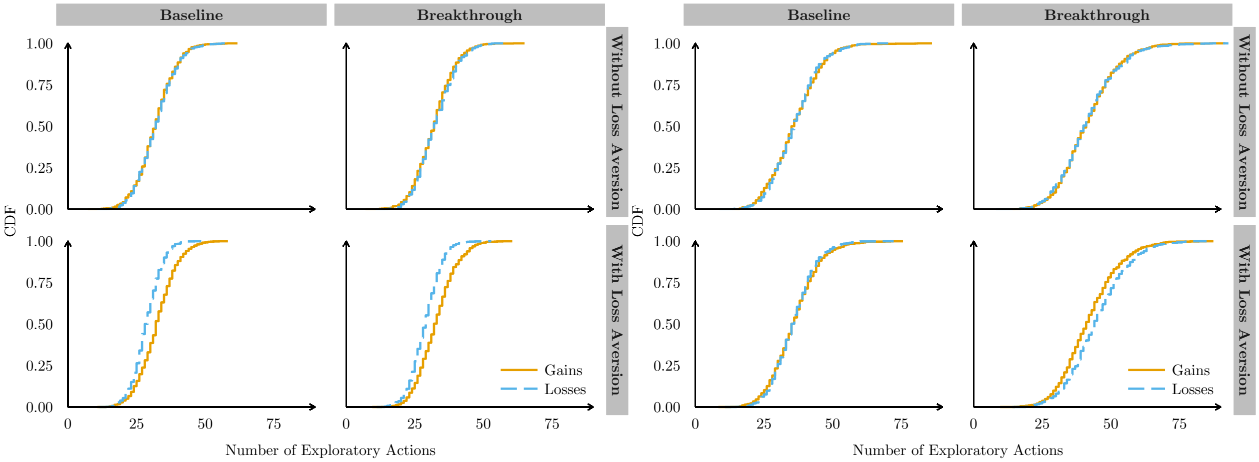
Figure 10: Simulation Comparison – Baseline vs. Breakthrough



(a) Actual Model **(b)** Foresight Model

Notes: **(a)** *actual* model leads to no difference between baseline and breakthrough, regardless of payoffs and loss aversion; **(b)** *foresight* model leads to more attempts at breakthrough innovation.

Figure 11: Simulation Comparison – Gains vs. Losses



(a) Actual Model **(b)** Foresight Model

Notes: **(a)** *actual* model with loss aversion leads the reinforcement learning agent to not attempt the breakthrough option under losses; **(b)** *foresight* model with loss aversion leads the reinforcement learning agent to attempt the breakthrough option more often under losses.