# The Streetlight Effect in Data-Driven Exploration *

Johannes Hoelzemann
University of Vienna

Gustavo Manso
UC Berkeley

Abhishek Nagaraj
UC Berkeley

Matteo Tranchero
UC Berkeley

September 29, 2022

**Abstract**

We consider settings such as innovation-oriented R&D and entrepreneurship where agents must explore across different projects with varying but uncertain payoffs. How does providing partial data on project payoffs affect individual performance and social welfare? While data can typically reduce uncertainty and improve welfare, we present a simple theoretical framework where data provision can *decrease* group and individual payoffs. We predict that when data shines a light on sufficiently attractive (but not optimal) projects, it can crowd-out exploration activity, lowering individual and group payoffs as compared to when no data is provided. We test our theory in an online lab experiment and show that data provision on the true value of one project can hurt individual payoffs by 12% and reduce the group's likelihood of discovering the optimal outcome by 48%. Our results provide a theoretical and empirical examination of the streetlight effect, outlining the conditions under which data leads agents to look under the lamppost rather than engage in socially beneficial exploration.

**JEL Classification:** C73, C92, D81, D83.
**Keywords:** Data; Innovation; Exploration and exploitation; Streetlight effect; Social learning; Dynamic public-good problem; Laboratory experiment.

*The experiment was conducted online via the Toronto Experimental Economics Laboratory in September 2021 and March 2022, under The University of Toronto Research Ethics Board 083 approval #00038482. Pre-registration information for the experiment is available at https://osf.io/zs2mu/. We are grateful to Lee Fleming, Ramana Nanda, and seminar participants at the NBER Productivity seminar, the SIE workshop, the MOM workshop at Harvard Business School, the SMS Special Conference at Bocconi, the Psych & Econ Lunch seminar at UC Berkeley for their suggestions. We thank Adit Jain for his excellent help in designing the experimental platform. Authors listed in alphabetical order. All authors contributed equally. Corresponding author: nagaraj@berkeley.edu.

# 1 Introduction

In a broad range of contexts, decision-makers must engage in strategic experimentation where they explore across multiple high-stakes choices under significant uncertainty. In such settings, we study how the provision of data that clarifies uncertainty about some options affects exploration outcomes. We are motivated by the parable of the streetlight effect, where information provision leads agents to search based on reasons of data availability rather than broader relevance or policy importance. Such an effect starkly contrasts with the view that (accurate) data provision can only be helpful because it reduces uncertainty and guides exploration making search more efficient. Our paper tries to reconcile both points of view by theoretically and experimentally studying how the streetlight effect might emerge in exploratory search among rational agents and outlining the conditions under which data hampers rather than spurs individual and social outcomes.

Our research question is motivated by two observations. First is the longstanding importance of settings where agents engage in strategic experimentation among risky options (Camuffo et al., 2022). Examples include venture capitalists evaluating different startups, pharmaceutical firms deciding between different drug candidates, retailers picking from a set of candidate store locations, mineral exploration firms choosing among different early-stage targets, and so on. The second observation is the arrival of the big data, artificial intelligence (AI), machine learning, and analytics revolution that is fundamentally reshaping how agents engage in risky exploration (Cockburn et al., 2019). These technologies can be seen as playing a "triaging" role, where they provide useful predictions or assessments on the viability of different risky choices, fundamentally affecting the search process (Christin, 2020). For example, venture capitalists are using data and analytics on firm performance to decide where to invest (Kerr et al., 2014; Ewens et al., 2018), pharmaceutical firms are using genetic data to guide drug development (Kao, 2021), and exploration firms are using satellite images and machine learning techniques to decide where to look for the next mineral deposit or oil field (Nagaraj, 2022).

While case studies and anecdotal evidence point to the positive effects of information on exploration outcomes, the macroeconomic effects of such technologies seem mixed (Brynjolfsson et al., 2021). In particular, informed commentators worry about how a more data-driven search process could manifest the streetlight effect, leading to suboptimal outcomes for both individual performance and social welfare. Our paper speaks to the burgeoning literature on the role of data and predictive analytics in shaping exploratory search and highlights the conditions under which it might harm rather than help performance and welfare.

We develop our theory based on strategic multi-armed bandit models, which have become canonical to study strategic experimentation. In our simple theoretical framework, agents choose among risky projects over two periods and both actions and payoffs are perfectly observable. Risky projects can be either low, medium or high value, but their quality can only be learned by exploring them. In each period, the decision-maker either optimally exploits the information they already have or they decide to invest in exploration in order to generate new data (for themselves and for others). As such, we model the idea that innovation and social learning are often times the work of pioneers, who, by bearing the costs of experimenting with new avenues, create informational spillovers for others. In this setup, we examine how the presence of a streetlight, i.e., data on the value of one opportunity, can shape exploration outcomes. Our key result is that the effects of data provision depend crucially on the type of project illuminated: data that sheds light on the *medium* value opportunity can reduce individual and group payoffs relative to not having any data at all, while data on low and high value opportunities are likely to benefit agents and society.

The intuition behind this result is as follows. In exploratory search, it is individually rational for agents to choose to invest in medium value opportunities highlighted by the data, whenever the value of a certain medium project is higher than the expected value of other risky projects. As the expected value of the risky projects is the same across players, and players can observe each other's actions and payoffs, there is a positive informational externality associated with a player's exploration decision. This gives rise to a public-good problem in the form of dynamically evolving information about the agents' common state of the world. In other words, agents' exploration decisions provide information beneficial to others beyond the private benefit they might receive from exploration, but it is privately suboptimal for them to engage in exploration. In a setting where data sheds light on a medium value project, all rational agents settle on the same medium but certain value project and the other options remain unexplored. In contrast, when no data is provided, agents are likely to select different projects to explore, leading to a greater likelihood that the high value project will be discovered. This ensures that agents who chose a low value project in the first period benefit from endogenously generated data and adjust their decision accordingly in the following time period. Absent data, agents might initially make poor decisions, but are more likely to learn and choose optimally later, leading to higher overall payoffs in the long run. Note that our predictions arise in the presence of data accurately capturing the underlying ground truth and do not rely on data misleading exploration because they are faulty or inaccurate (Weick, 1988; Puranam and Swamy, 2016). Our theoretical predictions are also robust to relaxing a few key assumptions such as simultaneous vs sequential moves, costly search, etc.

While our theoretical framework raises an interesting hypothesis, it is still an open question whether agents' behavior in practice is consistent with its predictions. If individuals factor social goals over personal payoffs, or are risk loving or behave in ways not captured by our simple expected utility framework, our results might not hold in practice. Accordingly, we developed and implemented an online lab experiment to test our theoretical predictions. In the experiment, groups of players collectively engaged in a two-period game of strategic exploration. They were each presented with a choice of five options with unknown but varying payoffs drawn from a known distribution. In the first period, participants were instructed to sequentially choose one project whose value would only be revealed after everyone had made their choice. In the second period, they could see the payoffs of the project chosen in the first period by all participants as well as their own before making their new choice. Payoffs were non-rival and cumulative, that is, players earned the full sum of payoffs from their choices across the two periods irrespective of whether other players made the same choice. In this setup, participants in the baseline condition were not shown any information, while those in other conditions were shown the payoff of one (low, medium, or high value) project at the outset of the game.

Results show that data provision on the medium value project reduces individual payoffs by 12% and reduces individual likelihood of finding the optimal outcome by 64% compared to the condition without any initial data. At the group level, the chances of discovering the best outcome decrease by 48 percentage points compared to when no data is initially provided. On the other hand, data describing the low or the high value projects raises both individual and group payoffs. Overall, the patterns we document are consistent with free-riding on information because of strategic concerns. In line with our theoretical framework, the mechanism is that data on medium value projects reduces exploration activity and lowers endogenous data generation through costly individual experimentation. We carry out a battery of robustness tests that rule out other possible explanations, such as learning and attitudes towards risk, and show that participants' behavior is consistent with our theory.

This work contributes to several strands of research. First, we speak to the literature on strategic experimentation and social learning (Bolton and Harris, 1999; Keller et al., 2005; Klein and Rady, 2011; Hörner et al., 2021). The experimental work in this literature largely considers solely single-agent bandit problems without strategic interdependencies among players and thus abstracts from informational spillovers (Meyer and Shi, 1995; Banks et al., 1997; Anderson, 2012; Hudja and Woods, 2021a,b). We are aware of two exceptions that involve an experimental investigation of a strategic-experimentation problem embedded in a simplified bandit framework. First, Boyce et al. (2016) examine a setting where participants face ambiguity about the type of the risky arm as well as asymmetric opportunity costs for making risky choices. Further, in recent experimental work, Hoelzemann and Klein (2021) study a setting where an agent can learn from the current experimentation of other agents. They show how the public-good nature of information gives rise to a free-rider problem in experimentation. Using a simpler theoretical framework than the strategic multi-armed bandit models, we build on this result by showing experimentally that public data provision might worsen these tendencies, curtailing exploration and thus aggregate data generation.

Second, we add to the growing work on the nature of data and how they shape innovation (Nagaraj and Stern, 2020; Nagaraj et al., 2020; Bessen et al., 2022). This includes recent work that looks at how access to information on past innovation affects future innovation (Gross, 2019; Hegde and Luo, 2018; Furman and Stern, 2011; Furman et al., 2021). Instead of considering data as a homogeneous commodity, we show that the nature of the data itself (in particular what it does and does not highlight) shapes agents' exploration choices (Siegel, 2013). Even a simple approximation of data as being about low, medium, or high payoff outcomes is enough to illustrate the channels through which data differentially affect innovation and payoffs. Notably, our results emerge in a context where we operationalized data as instrumental information, i.e., unbiased and directly payoff-relevant. Our results could be even starker if data were imprecise or biased (Henrich et al., 2010; Cao et al., 2021). Additionally, we also propose a novel mechanism through which data might harm exploration. Our theoretical framework shows how data can cause agents to implicitly coordinate on certain – but dominated – projects and thus lower overall exploration activity, harming group and individual outcomes. This result adds to other work that has pointed to the importance of data produced during routine economic activity of firms by showing how pre-existing data might hinder the production of new data (Jones and Tonetti, 2020; Farboodi and Veldkamp, 2020).

Finally, we build on the innovation search literature and show how data might have counterintuitive effects on experimentation. Our work is based on the stylized formalization of the trade-off between exploration and exploitation originally formalized by Thompson (1933) and Robbins (1952) and further elaborated in recent work (March, 1991; Fleming, 2001; Manso, 2011; Ederer and Manso, 2013; Klein, 2016). In particular, we contribute to research studying the importance of different types of data in shaping experimentation decisions in risky environments (Ewens et al., 2018; Camuffo et al., 2020; Krieger, 2021). Our work is also related to the management and strategy literature around search. In particular, our findings echo past work suggesting how firms might "be stuck" on local optima while searching on rugged technological landscapes (Levinthal, 1997). We add to past work on the role of scientific information (Fleming and Sorenson, 2004) by showing how data and science can both be used to rule out bad combinations and find promising ones. We add to this intuition, by showing how partial data describing the landscape might exacerbate the proclivity towards exploitation and hurt innovation if they induce agents to be stuck on a "good enough" but dominated outcome.

The remainder of the paper proceeds as follows. In Section 2, we provide an overview of the theoretical framework, including a simplified formal model and a numerical illustration. Section 3 describes the experimental setup and design. Section 4 describes the main results and Section 5 concludes.

# 2 Theory

**Motivating Example** Consider a simple example that highlights the key motivation behind our theoretical setup. Multiple firms are considering alternate technologies to discover a vaccine for a new, viral infection. For example, in the case of COVID-19, firms had the choice between DNA, mRNA and traditional inactive virus techniques to develop a new vaccine (Nagy and Alhatlani, 2021). Technologies vary widely in their efficacy for the problem at hand, with most being ineffective, some being moderately promising, while only a select few (or even just one) being the most effective at containing the disease. Firms know this general pattern, but do not know, ex ante, which technologies are duds, and which ones are promising. Firms can invest in one technology at a time; so the cost of choosing the "wrong" technology is the opportunity cost of not having chosen the optimal target in that time period, without much harm to other firms making the same decision. Firms also engage in social learning – if another firm discovers a promising technology, firms can choose to invest in the same technology in the following period. In other words, firms' exploration activity generates knowledge spillovers for others.

In this setting, consider the arrival of public data that clarifies the underlying potential of a subset of technologies. For example, in the COVID-19 case, if public data suggests that the mRNA platform is effective, future research investments might skew in this direction. This might be efficient for society but it might also reduce investments in other technologies such as the DNA platform, even if the latter might hold more promise in the long run. More generally, how does data clarifying uncertainty affect firm performance and shape the chances of a breakthrough being discovered? And do these outcomes depend on which particular technology is described by the public data? These are the questions our framework tries to tackle. Our simple theoretical setup is inspired by bandit models where agents are uncertain about the payoffs associated with projects, but then can learn this information over time. We embed this idea into a strategic setting where several agents solve the same simplified bandit problem and thus can exploit information generated by the other agents.

**Setup** There are $n$ agents engaged in a search to maximize their individual payoffs and must choose between $m$ projects of initially unknown value, with $m \leq n$. Such projects can have three types of payoffs: $m_L$ of them have a low payoff $L$, $m_M$ have a medium payoff $M$, and the remaining $m_H$ have a high payoff $H$, such that $m_L, m_M, m_H > 0$, $m_L + m_M + m_H = m$ and $0 < L < M < H$. This distribution is known ex ante to agents, but they do not know the type of any specific project at the outset of the game. All agents live for two periods and are risk neutral with zero discounting. Agents cannot communicate with each other. Like the example highlighted before, this general setup represents many settings where agents are faced with options of unknown values, and where "good" projects are hard to find, but have high payoffs (Kerr et al., 2014; Manso, 2016).

**Dynamics** The game unfolds as follows. The $n$ agents sequentially choose a project in each period according to a random order. They can observe the options that players who moved earlier selected, but do not yet learn the value associated with their choice. Once all agents have selected a project, the underlying payoffs of their chosen project are revealed to all players, and period 1 concludes. In period 2, knowing payoffs associated with previously explored projects, agents repeat this process. Similar to period 1, agents choose sequentially according to the same random order. They can select a previously explored project of known value or an unexplored one, the value of which will be revealed at the end of the second period. Once all choices are made, period 2 ends. Payoffs are cumulative across the two periods, i.e., the sum of values associated with their choices over time, and projects are assumed to be non-rival. If multiple agents choose the same project, they all receive its payoff. This means that there is no penalty to choosing later in a sequence since agents have the full menu of choices. No payoff relevant information is revealed until the end of the period and payoffs are independent of others' choices. This setup mimics competitive markets where organizations engage in parallel research and development: projects do not directly compete, but the generation of information about what works and what does not is valuable for all participants in the market (Krieger, 2021). Thus, in contrast to conventional payoff externalities in public good problems, externalities are purely informational. The presence of the other agents impacts a given agent only via the information they produce over time (Hoelzemann and Klein, 2021).

**Equilibrium without Data** To set the stage, we lay out the setting in the absence of any prior information. The following proposition considers the equilibrium without data, meaning that no information about payoffs is disclosed to agents at the outset of the game.

**Proposition 1.** *The equilibrium without data involves all projects being explored in period* $1$*, and the high value project (breakthrough) being selected by all agents in period* $2$*. At least one agent achieves a breakthrough and the expected payoff to each agent is:* $\frac{m_L}{m}L + \frac{m_M}{m}M + \frac{m+m_H}{m}H$*.*

In this simple setup, agents are initially indifferent between choosing projects since they all have the same expected payoff. However, rational agents select a different project than the ones already chosen in order to reveal more information that might be useful in the second period. If each agent acts accordingly, all projects will be explored in the first period as $n \geq m$. As a result, there will be a breakthrough and thus the maximum is always uncovered. The expected payoff is the likelihood of a random draw in period $1$, and of $H$ in period $2$. To summarize, without any prior information dispersed exploration ensures a breakthrough in the first period – intended as the discovery of the maximum – and high payoffs in the second period.

**Equilibrium with Data on $L$ or $H$ Projects** We compare the setup outlined above with a setting where data is provided. That is, the payoff of one project is publicly revealed at the outset of the game. Depending on which project is disclosed, different dynamics unfold.[1] We begin with the two cases that have been studied before, namely when data either rule in the best alternative or rule out a poor option (Nelson, 1982; Fleming and Sorenson, 2004). Let $\pi_i$ denote a player's payoff where $i \in \{\emptyset, L, M, H\}$ indicates data provided and $P(H|i)$ be the conditional probability of discovering $H$ given data $i$.

**Proposition 2.** *If the underlying value of a project is revealed to be high, every agent selects the project whose value is revealed at the outset of the game. Thus, each agents' payoff is* $2H$*. If the project revealed is low value, the data rules out one dominated option. The expected payoff to each agent is* $\frac{m_L-1}{m-1}L + \frac{m_M}{m-1}M + \frac{m+m_H-1}{m-1}H$*. Taken together, we have the following two results:*

$$\text{[Payoffs]} \quad \pi_\emptyset < \pi_L < \pi_H \tag{1}$$

[1] In online Appendix A, we compare equilibrium payoffs with no data and equilibrium payoffs of revealing the output of a random project, showing that our main result also applies in that case.

$$\text{[Breakthrough]} \quad P(H|\emptyset) = P(H|L) = P(H|H). \tag{2}$$

If the $H$ project is revealed, all underlying uncertainty is resolved and each player selects that particular project in both periods. Each agent's expected payoff is maximized and equals $2H$. This highlights naturally how data can guide discovery by leading directly to the best outcome (Nagaraj, 2022). If the project revealed returns a low payoff, then the expected payoff to each agent is still strictly greater than in the absence of data. Indeed, it is no mean feat that information can help innovation also by simply ruling out low potential alternatives (Kao, 2021). Irrespective of whether an $L$ project is revealed at the outset of the game or whether no data is provided at all, there is dispersed exploration and agents always achieve a breakthrough.

**Equilibrium with Data on $M$ Projects** What is arguably more interesting, and so far understudied, is the intermediate case when the *medium value* project is revealed. In this case, there exists a non-empty parameter space where data can be detrimental to exploration and social welfare due to the *streetlight effect*. To see this, we only require the payoff from choosing $M$ to be appealing enough relative to further exploration of other projects whose underlying value is thus far unknown. Agents choose the medium value project in both periods if and only if $2M > \frac{m_L}{m-1}(L+M) + \frac{m_M-1}{m-1}(2M) + \frac{m_H}{m-1}(2H)$. This condition can be synthesized as follows:

**Assumption 1** ("Medium Project is Good Enough")**.**

$$M > \frac{m_L}{2m_H + m_L} L + \frac{2m_H}{2m_H + m_L} H \tag{3}$$

Assumption 1 ensures that selecting the medium project individually dominates searching for the high-value one. Rational agents choose it in both periods and achieve a payoff of $2M$. However, whenever $M$ is not too large relative to $L$ and $H$, we can show that – perhaps contrary to what one would intuitively expect – individual payoffs with the data are even lower than payoffs with no data. More formally, we introduce:

**Proposition 3** ("Individual Payoff with Data on Medium Project")**.** *Under Assumption 1 and if*

$$M < \frac{m_L}{2m_L + m_M + 2m_H} L + \frac{m_L + m_M + 2m_H}{2m_L + m_M + 2m_H} H, \tag{4}$$

*the following strict payoff ranking holds:* $\pi_M < \pi_\emptyset < \pi_L < \pi_H$.

*Proof.* In Equation (1) of Proposition 2, we established that $\pi_\emptyset < \pi_L < \pi_H$. We only need to show that the expected individual payoff without data dominates the expected individual payoff whenever the $M$ project is described by the data and revealed at the outset. This is true if $\frac{m_L}{m} L + \frac{m_M}{m} M + \frac{m+m_H}{m} H > 2M$, which is equivalent to the condition in the proposition. ∎

Proposition 3 illustrates that agents' payoffs are dominated whenever they receive data on $M$ by data on $L$, data on $H$, and even no data at all. This occurs because even if $M$ is "good enough" to be preferred over exploring, selecting it in period 1 precludes the possibility of generating new data that could lead to a higher payoff in period 2. Our result is an immediate consequence of the informational externality that arises in our setting because agents can learn from the experimentation of others. Despite being the payoff-maximizing choice at the individual level, if each agent chooses the medium value project, the group as a whole is forfeiting risky exploration and missing innovations that would make everyone better off in the second period. This argument is summarized in the following result:

**Proposition 4** ("Breakthrough with Data on Medium Project"). *Under Assumption 1 and if*

$$M < \frac{m_L}{2m_L + m_M + 2m_H} L + \frac{m_L + m_M + 2m_H}{2m_L + m_M + 2m_H} H, \tag{5}$$

*the following strict inequality holds:* $P(H|M) < P(H|i)$ where $i \in \{\emptyset, L, H\}$.

*Proof.* The proof directly derives from our preceding discussion. If $M$ is appealing enough, agents forfeit exploration and never achieve a breakthrough, i.e., never discover $H$. If no data is provided or $L$ is revealed ex ante, agents explore all remaining unknown options in period 1, thus always uncovering the maximum. The statement is trivially true whenever $H$ is revealed. ∎

**Simple Example** We now consider a simple example. Suppose there exist five agents and five projects, i.e., $n = m = 5$. Assume that there are three low-value projects, one medium-value project, and one high-value project. In this setting, Assumption 1 is equivalent to $M > 3L/5 + 2H/5$ and the condition in equations (4) and (5) is equivalent to $M < L/3 + 2H/3$. For instance, the combination $L = 1, M = 5, H = 10$ satisfies both assumptions. In this case, $\pi_M = 10 < \pi_\emptyset = 13.6 < \pi_L = 14.3 < \pi_H = 20$. Using this set of parameters (and others that similarly satisfy the assumptions of Propositions 3 and 4), we examine whether our predictions hold and the streetlight effect emerges in practice.

# 3 Design

While our theory raises an interesting hypothesis it is still an open question as to whether it would explain the behavior of agents in practice. In particular, it is possible some agents might be risk-loving or pro-social, thereby ignoring certain "M" payoffs, and exploring even when this information is provided, to the benefit of all. It is also possible that individuals fail to understand or calculate private payoffs or lack attention, thereby violating the key predictions of our model.

To test whether these deviations are strong enough in practice to overturn our baseline results, we created an experimental environment that mimics our theoretical setup. This experiment allows us to test our hypotheses on the effects of data on payoffs and innovation. We then conduct an online experiment in which multiple participants have to solve an exploration task mirroring our theoretical framework.

## 3.1 Experimental Procedure and Logistics

Participants were invited to either the data or no-data condition in groups of ten. All ten participants logged oto the platform remotely at a specific time. Upon arrival, participants received detailed written instructions about the experiment and watched a compulsory six minute video that reiterated the main instructions while familiarizing them with the experimental platform.[2] Participants were then required to complete a short quiz as an attention and comprehension test.

The experiment consisted of independent "rounds." Mimicking our conceptual framework, each round was composed of two periods over which player payoffs were calculated. The participants were randomly split into two groups made of five players each. These groups were randomly reshuffled every five rounds played. In total, participants played 20 rounds. At the end of the experiment, we collected some information on participants' demographic

[2]The video shown to participants in the no-data condition is here: `https://drive.google.com/file/d/1TsGs2fLIcV6XFyMTmAwuJUDnAP-rAi31` and the one shown to participants in the condition with data is here: `https://drive.google.com/file/d/1vx0F-VG1P6kQQJanO99VKaYsfz1QoNbR`

attributes and we elicited their degree of risk aversion with a monetarily incentivized and upscaled variant of the Holt & Laury task (Holt and Laury, 2002). Participants were then paid their experimental earnings from one randomly selected round plus a show-up fee of CA$ 5 and the amount earned in the lottery associated with the risk attitude elicitation task.

The experiment was programmed with the open source software oTree (Chen et al., 2016) and conducted by the Toronto Experimental Economics Laboratory (TEEL). Participants were recruited from TEEL's subject pool using ORSEE (Greiner, 2015) among undergraduate students who had participated in at most five experiments. Participation was voluntary and participants could withdraw at any point during the experiment. We ran 20 sessions with 200 participants in total and no participant was allowed to join more than one session. The age of participants ranged from 18 to 32 years, with an average of 20.04 and a standard deviation of 2.33. The experimental sessions took place in early September 2021 and early March 2022. Sessions lasted about 75 minutes, with average earnings of CA$ 26.78 and a standard deviation of CA$ 5.43. The experiment took around 50 minutes, but extra time was needed to read the instructions, watch the explanatory video, and answer the attention quiz.

## 3.2 Task Description and Implementation

As shown in Panel A of Figure 1, the experimental environment and the layout of an individual round were designed to track our theoretical framework closely. Participants take the role of an individual engaged in a hunt for precious gems. There are $m = 5$ mountains, each hiding one type of gem that can only be uncovered by exploring the mountain. There are three types of gems of varying rarity and value hidden in the mountains: three topazes ($L$), one ruby ($M$), and one diamond ($H$). The exact values of the precious stones vary across rounds but the diamonds are always worth more than the rubies and the rubies are always worth more than the topazes. The chosen parameters always satisfy equations (3) and (4) as outlined in Section 2. Participants are told the empirical frequencies, namely that there are always three topazes, one ruby and one diamond, although they do not know which mountain hides which gem. In addition to specifying the values and distributions of the gems, the interface keeps track of the period ("stage"), the round and the "block" number as participants make their way through the experiment (top right). A new block simply indicates the reshuffling of participants as new groups are being formed, which then stay together for five rounds.

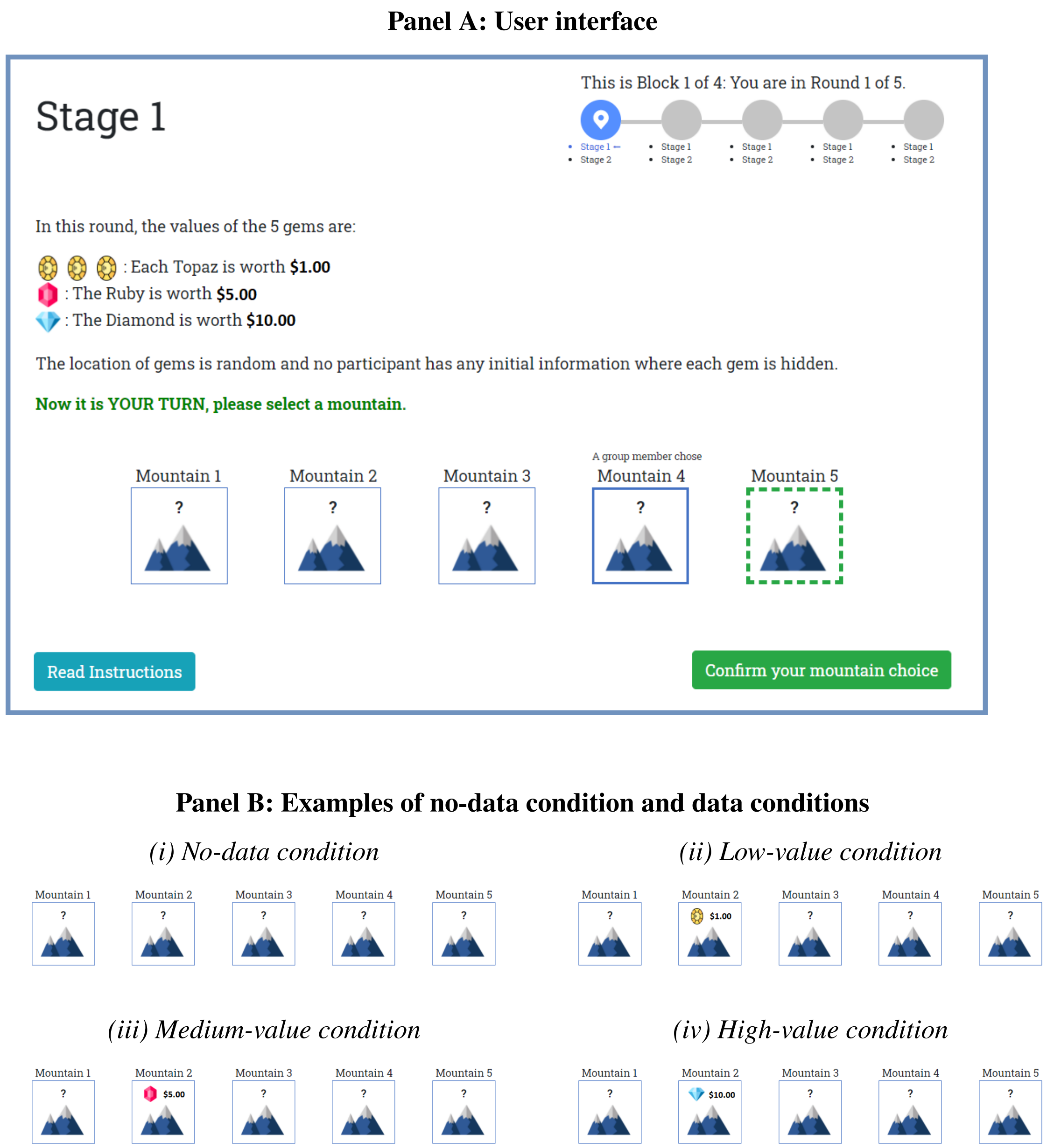



Figure 1: Experimental platform.

Note: This figure reproduces the interface seen by participants in our online experiment. Panel A shows how the experimental platform is seen by the participants in the no-data condition. In this example, Mountain 4 has been selected by some other participants, and the user has selected Mountain 5. Note that the dollar value of the gems changes in every round and it is showed on the left. Panel B exemplifies the four different conditions of the experiment. When subjects are assigned to the data condition, they see the value of the gem hidden behind one randomly chosen mountain. This could either be the medium, the high, or one of the low outcomes. The specific monetary value of the mapped mountain changes in every round and it is reported near the gem image.

In addition, participants always have access to written instructions at any point in time and could contact an experimenter via cell phone or Zoom for assistance. The objective of the game is to find the most valuable gems, since the value of the gems found directly translates into earnings in dollars.

All five players in any given round are anonymous to each other, and cannot directly interact or communicate. Participants know that their co-players change every five rounds, but they have no means to know whom they were playing with each time, since players were not identified in any way.[3] Players select which mountain to explore sequentially, based on a random order that changes every round. A dynamic instruction element on their screen turns green and indicates that it is their turn to make a choice (otherwise they must wait). None of them has any initial private information about the gems' location, which changes every round (but not between the first and second period of the same round). While waiting for their turn, players can see which mountains are being selected by their co-players. When it is their turn, players choose one mountain to explore. They can pick the same or different mountain as other players and their payoff is independent of whether or not someone else has already selected their choice. In other words, if participants overlap in their choice of mountain, each of them still receives the entire value of the gem uncovered since payoffs are non-rival.

In the no-data condition, the two periods of a round proceed as follows. In period 1, all participants sequentially choose one mountain to explore, as described above. At the end of period 1, the gems hidden in the mountains selected by the participants are revealed to all players, and each player earns the value hidden in the mountain of their choice. In period 2, players can again choose any of the same five mountains according to the same sequential order. The position of gems remains the same, but this time participants will also see the gems located in the mountains explored in period 1. Therefore, each player can either choose the same mountain of period 1 or switch to another one exploiting the new data generated by collective exploration choices. At the end of period 2, the gems contained by the mountains selected in period 2 are revealed, and their values are added to participants' round payoffs. Individual earnings for the round equal the sum of the value of the gems found in period 1 and period 2.

In the data condition, the two periods proceed exactly as in the no-data condition but one

[3]In a sense, players could only interact indirectly by choosing which option to explore. When a player selected an option, the other four group members only saw a generic "A group member chose this option," without ever identifying who made the choice. See Figure 1 for an example.

of the mountains is “mapped,” i.e., the gem hidden behind one mountain is revealed to all participants at the start of each round. Panel B of Figure1 shows the different possibilities. Figure (i) is the no-data condition where all mountains are undisclosed. Figures (ii), (iii) and (iv) represent the three possibilities where the mapped mountain happens to have a low, medium or high value (topaz, ruby or diamond), respectively. Precisely which mountain is revealed and in what order is decided by a script employing stochastic processes, as shown in Figure B.1. The data on the mapped mountain constitutes the only public information on gems’ position that participants in the data condition know before starting exploring in period 1.

In total, we collected data both at the individual and group level for 800 rounds. Out of the 800 rounds played, participants saw data on one of the low value outcomes in 320 rounds, data on the medium outcome in 156 rounds, and data on the high value outcome in 164 rounds. In the remaining 160 rounds participants did not receive any initial data on the gems’ location. This setup allows us to test whether behavior is consistent with the predictions of our theoretical framework by comparing (i) individual payoffs and (ii) the likelihood of a breakthrough, i.e., of discovering the diamond, in each of the four conditions. We present the main results at the individual level to closely track our conceptual setup, but we also show group-level outcomes to capture the interdependence between individual exploration, social welfare, and the trade-offs involved in the dynamic production of public information.

# 4 Results

## 4.1 Payoffs

Our first set of experimental results shows that data can have strong heterogeneous effects on individual outcomes in line with our predictions. Panel A of Figure 2 tests the chain of predictions in Proposition 3. For each round, we calculate the maximum possible payoff, that is the value of the diamond times two, and compute average individual payoffs as a percent of this value. This allows us to compare individual payoffs across rounds, even though the specific values of the low, medium and high value gems vary. We plot this average by the three data conditions and the no-data condition. As is apparent, providing data on the high

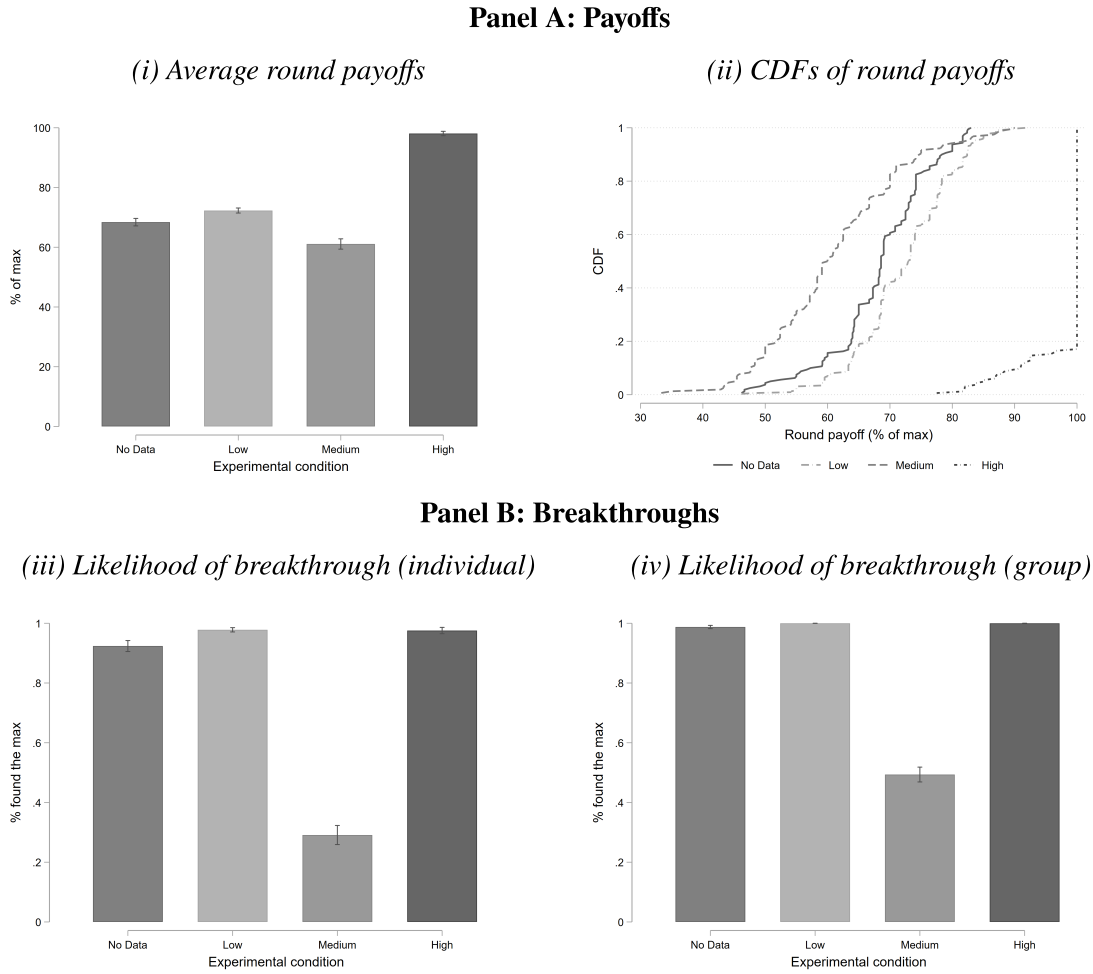



Figure 2: Round outcomes by experimental condition.

Note: Panel A reports the experimental results on the round payoffs computed as a share of the maximum possible in each round. Figure (i) shows the average collective payoffs achieved in each round by experimental condition. Figure (ii) plots the cumulative density function of round payoffs by experimental condition. Pairwise Kolmogorov–Smirnov tests of stochastic dominance confirm that the empirical distribution function of round payoffs in the M condition is strictly smaller than that of rounds without data; the opposite is true for L and H conditions. Panel B reports the experimental results on the likelihood of a breakthrough in each round. Figure (iii) shows the share of participants that found the maximum by experimental condition. Figure (iv) shows the share of rounds where the maximum was uncovered by experimental condition.

value project increases individual payoffs – in fact, average payoffs are close to 100% of the maximum value since participants can choose the diamond in both periods. Compared to the no-data condition, providing data on the payoff from one low-value project also increases average payoffs, although this effect is modest. However, the most striking finding is that providing data on the medium value project *decreases* average payoffs when compared to *all* other conditions, including whenever no data is provided at all.

Table 1, Column 1 presents regression estimates that quantify these results, showing that

Table 1: Round-level outcomes

| | Individual payoff | I(Individual found max) | I(Group found max) |
|---|---|---|---|
| High | 6.550*** | 0.052** | 0.010 |
| | (0.192) | (0.017) | (0.008) |
| Low | 0.878*** | 0.051** | 0.013 |
| | (0.147) | (0.016) | (0.008) |
| Medium | -1.635*** | -0.635*** | -0.494*** |
| | (0.272) | (0.047) | (0.062) |
| Constant | 13.700*** | 0.976*** | 1.035*** |
| | (0.209) | (0.023) | (0.030) |
| Round order FE | Yes | Yes | No |
| Block order FE | Yes | Yes | Yes |
| Payoff structure FE | Yes | Yes | Yes |
| Observations | 4000 | 4000 | 800 |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors clustered at the session level in parentheses. Estimates from OLS models. The sample in Columns 1 and 2 is at the participant-round level (5 participants $\times$ 800 rounds). The sample in Column 3 is at the group-round level (800 rounds). *Individual payoff*= participant-level round payoffs in dollars; *I(Individual found max)*:0/1=1 if the location of the maximum was found by the participant; *I(Group found max)*:0/1=1 if the location of the maximum was found by at least one participant in the round. The excluded category captured by the constant is the condition without data.

while the average participant in the no-data condition earns about \$13.7, data on the medium-value mountain reduces this payoff by \$1.64, a reduction of about 12%. Data on the high-value mountain increases payoffs by \$6.56, and data on the low value mountain increases payoffs by \$0.88. Besides being statistically significant, these differences are also large in magnitude and thus economically meaningful. Our experimental evidence is in line with Proposition 3 and shows how the streetlight effect can emerge in real world data-driven exploration tasks. Insofar as the sum of individual payoffs can measure social welfare, these results document that, under certain conditions, data provision can entail substantial societal costs. The next section highlights this result even more starkly when considering group-level outcomes.

## 4.2 Breakthroughs

Besides payoffs, the second outcome of interest is constituted by the likelihood that participants discover the high-value outcome, both individually and collectively. Following our motivating example, we are interested not only in firm payoffs for discovering key technologies alone, but also in whether *at least one* firm discovers the optimal technology. While being engaged in individual-payoff maximization, the choices of economic agents have significant spillovers on collective discovery (Rosenberg, 1992; Bolton and Harris, 1999; Keller et al., 2005). This tension is especially stark in our setting since both actions and payoffs are perfectly observable and the dynamically evolving information about technology discovery is a public good (Hoelzemann and Klein, 2021). With data being provided ex ante, if sufficiently good outcomes are known, agents can stick to them in hopes of other agents' engaging in search. The adverse effect, however, is that the optimal alternative can potentially be never discovered, which is the crucial economic insight of Proposition 4.

Panel B of Figure 2 illustrates this tension. Revealing the location of the medium outcome significantly reduces the individual chances of a breakthrough. Table 1, Column 2 estimates this effect to be a reduction in this possibility by about 64% relative to the no-data condition. In line with the intuition of previous work, innovation is highest when data directly lead to the best option, but it also increases when the available data rule out low-value alternatives. Notably, the individual decision to select the known option has aggregate consequences at the group level due to the public-good nature of the new data generated. In presence of data about the location of the medium outcome, the location of the optimum was identified in only half of the rounds played (Table 1, Column 3).

## 4.3 Mechanisms

What mechanisms can explain our results? Our theoretical framework suggests that data shapes individual experimentation choices, i.e., what we call the streetlight effect. When the projects revealed are good enough to lure participants into forfeiting further exploration, discovering the most valuable innovations is prevented and long-term payoffs are lowered. Figure 3 shows the distribution of unknown mountains selected in period 1 according to which data, if any, is provided at the outset. While exploration, defined as the likelihood that an unknown mountain is chosen in period 1, is trivially very low whenever the location of the maximum is known, comparing the other three conditions is very informative. Receiving data on a low outcome does not reduce exploration, which remains very close to the levels of the no-data condition and barely distinguishable from it (Table 2, Column 1). On the contrary, when the mountain disclosed conceals the medium value outcome, the amount of collective exploration decreases by 45% relative to the no-data condition. Moreover, Figure 3 displays a positive relationship between the amount of exploration in period 1 and the share of participants that collect the maximum reward in period 2. This is especially visible in panel (iii) for the case when the position of the $M$ outcome was revealed. When no participant

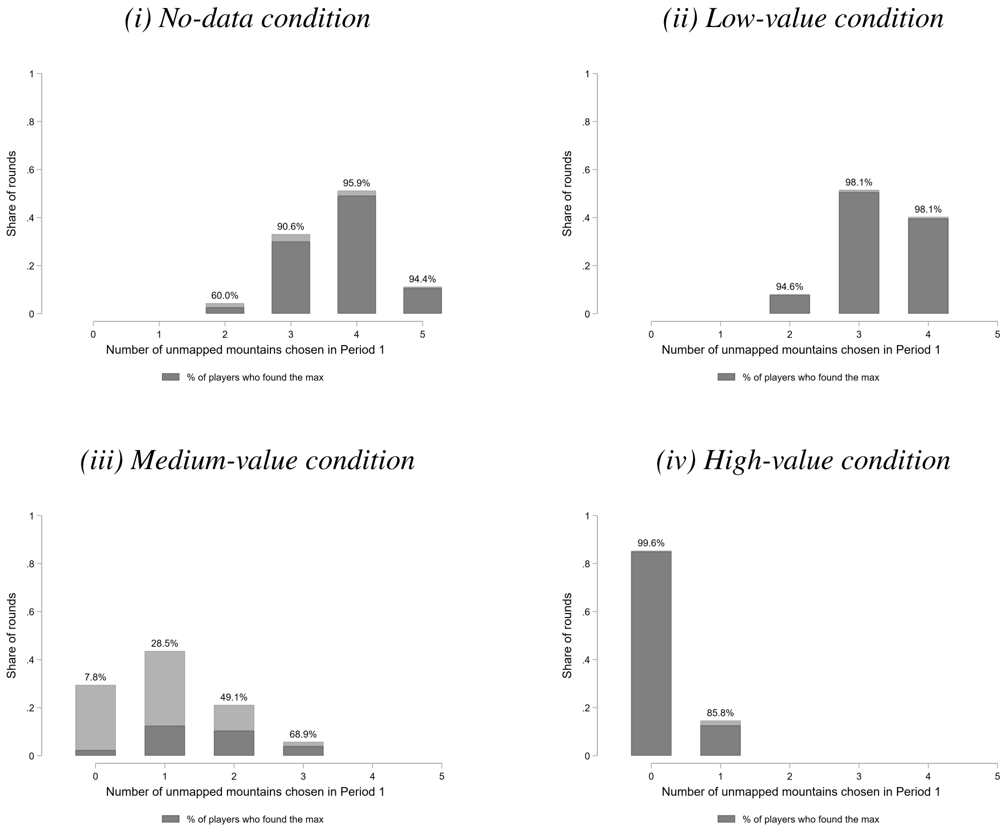



Figure 3: Distribution of the number of unknown mountains chosen in period 1 across experimental conditions, and its relationship with the discovery of the maximum.

Note: each plot represents the empirical frequencies of rounds for each possible number of unknown options chosen in period 1, showed separately by experimental condition. In each bar, the dark grey portion shows the share of participants who found the maximum payoff in period 2. This share is also written on top of each bar to ease comparisons.

ventured outside the disclosed option in period 1, the chances of a given participant finding the maximum in period 2 were only 7.8%; this number increased sixfold when two other options were explored in the first period.

These exploration dynamics are also evident when the group payoffs are divided along the two periods that constitute each round (Table 2). In period 1, data on the medium outcome increases social welfare, since participants can revert to this sure option and avoid the potential failures entailed by risky experimentation. However, the situation completely reverses in period 2: the short term gains from selecting the “good enough" option early on are more than offset by the cost of not uncovering the maximum. In the no-data and low-value cases, the endogenous data generation via experimentation allows everyone to select the best outcome in the second period. This is rarely the case when the medium

Table 2: Analysis of the mechanisms

| | Exploration | Individual payoff | | I(Individual found max) | | I(Group found max) | |
|---|---|---|---|---|---|---|---|
| | Round | Period 1 | Period 2 | Period 1 | Period 2 | Period 1 | Period 2 |
| High | -75.320*** | 6.348*** | 0.201 | 0.785*** | 0.052** | 0.297*** | 0.010 |
| | (2.968) | (0.131) | (0.108) | (0.010) | (0.017) | (0.023) | (0.008) |
| Low | 6.473* | 0.599*** | 0.280** | 0.051*** | 0.054** | 0.121*** | 0.013 |
| | (2.370) | (0.116) | (0.092) | (0.005) | (0.016) | (0.028) | (0.008) |
| Medium | -38.890*** | 1.595*** | -3.230*** | -0.133*** | -0.632*** | -0.479*** | -0.494*** |
| | (4.028) | (0.145) | (0.208) | (0.010) | (0.047) | (0.045) | (0.062) |
| Constant | 86.290*** | 3.844*** | 9.845*** | 0.191*** | 0.968*** | 0.808*** | 1.035*** |
| | (3.035) | (0.166) | (0.111) | (0.013) | (0.021) | (0.037) | (0.030) |
| Round order FE | No | Yes | Yes | Yes | Yes | No | No |
| Block order FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Payoff structure FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 800 | 4000 | 4000 | 4000 | 4000 | 800 | 800 |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors clustered at the session level in parentheses.

Estimates from OLS models. The sample in Column 1 is at the group-round level (800 rounds). The sample in Columns 2, 3, 4, 5 is at the participant-period level (5 participants $\times$ 800 periods of each type). The sample in Columns 6 and 7 is at the group-period level (800 periods of each type). *Exploration*= share of unknown mountains explored in the round; *Individual payoff*= participant-level period payoffs in dollars; *I(Individual found max)*:0/1=1 if the location of the maximum was found by the participant in the period; *I(Group found max)*:0/1=1 if the location of the maximum was found by at least one participant in the period. The excluded category captured by the constant is the condition without data.

outcome was disclosed ex ante, as we highlighted in Figure 2. Indeed, Table 2 shows that the lack of exploration early on in the game translates to a lower probability of locating the maximum, which in turn prevents its exploitation in the second period of the game. This is a direct demonstration of the streetlight effect in action: data might tilt the balance between exploration and exploitation and hurt social welfare by leaving participants stuck on a suboptimal outcome.

## 4.4 Robustness Tests

**Risk Aversion** Given the payoff structure, the individual expected payoff from choosing an unknown option when the medium outcome is revealed is dominated by the certain option. If agents are risk neutral, as we assumed in our theoretical framework, they would always prefer the ex ante disclosed option. However, it could be that risk loving agents prefer to explore, hence preventing the streetlight effect from arising. We explore this possibility using the measures of individual attitudes towards risk that we collected with an incentivized variant of the Holt & Laury task at the end of the experiment (Holt and Laury, 2002). In Table B.1, we document that risk attitudes are not associated with exploration choices when the medium outcome is known ex ante. These results suggest that given the payoff structure implied by equations (3) and (4), participants in our experiment do not have risk preferences extreme enough to offset the negative effect of data provision.

**Learning** While the evidence presented supports our theoretical framework, we investigate whether the results could be driven by an incomplete understanding of various aspects of the experiment. Over time, as participants repeatedly play variants of the game with different payoffs and mountain locations, they could learn that sticking to the medium outcome hurts their individual (and thus collective) payoffs. Figure B.2 shows this is not the case. Recall that every five rounds, participants are randomly reshuffled and new groups are formed, and this procedure is repeated four times. Our results on payoffs and discovery hold for each of the four "blocks" of five rounds each. Despite playing a total of twenty rounds, participants behaved consistently and replicated our main results over time, without changing their propensity to select the disclosed options over time (Figure B.4). This rules out the possibility that our results are due to limited familiarity with the experimental setup or that they would vanish as participants learn the game's dynamics (perhaps due to the fact that our design and game are relatively simple to understand).

**Correlates of Exploration Choices when Medium-Value Project is Revealed** Even if risk preferences and learning of the game dynamics do not seem to drive our results, we still see that sometimes there are participants who decide to explore an undisclosed option even when the medium value is revealed. Table B.2 explores which individual characteristics correlate with the decision to forfeit the medium outcome in period 1. After controlling for age, gender, and the specific payoff structure of the round in question, we find that difficulty in understanding the instructions (as proxied by being an English native speaker) does not seem to play a role in our setting. Neither the round number nor the order in which the participant chose within the round are associated with the decision to select an unknown option. The only variable that appears significant is the (standardized) number of incorrect answers to the attention quizzes that followed the instructions. Given the simplicity of our experiment design, this suggests that participants who forfeited the medium option in period 1 were possibly not paying enough attention when making their choices.

**Qualitative Evidence** At the end of the experiment, we asked participants to briefly describe the decision-making process they adopted and how they thought the other players were making their choices. Participants' responses suggested that they grasped the game's dynamics, and the overwhelming majority described following the profit-maximizing reasoning that underlies our conceptual framework and thus our propositions. Only a handful of participants declared to choose randomly, to sometimes explore an unknown option out of boredom, or to take occasional gambles. Notably, a few players understood the public good nature of their decisions, and tried to escape the free-riding dynamics arising when the medium outcome was revealed. In this case, they realized that unless other players adopted a similar strategy and without explicit coordination, the best option would be to stick to the medium-value mountain. In the words of one participant: *"If the revealed gem was a ruby, I would consider other players' choices (if they were to choose another mountain I might consider also choosing another mountain), but in most cases choosing the ruby twice gives a higher payoff."*[4]

# 5 Conclusion

Our paper develops a theoretical and empirical framework to understand the effects of data on decision-making under uncertainty. We argued that in a group setting, when multiple agents are engaged in learning about the payoffs from different projects of varying quality, providing information about "medium" quality projects can lower payoffs in the long run. Our experiment validates this prediction, since we found that participants earned about 12% less when they had data about a medium outcome than when they had no data at all. Knowing which option harbors the medium outcome improves payoffs in the short run, but it reduces the likelihood that the maximum will be discovered and, therefore, it lowers overall payoffs. To wit, we find that the likelihood that the optimum was collectively discovered was almost half as low when data on the medium outcome was provided.

While our work considers a simplified theoretical framework, the basic intuition could generalize to relaxing a few key assumptions. First, in our setup, agents wait for their

[4]The fact that participants do indeed understand the informational externality associated with their actions is consistent with the findings in Hoelzemann and Klein (2021). In particular, process data gathered using eye-trackers supports the qualitative evidence cited above, because participants carefully monitor other participants whenever informational externalities are present.

turn and can see past moves before choosing a project to explore, thus inducing implicit coordination. While this assumption models many markets where agents arrive sequentially and can see the entire pattern of past investments, it does not capture scenarios where multiple agents invest simultaneously. Our setup could easily be extended to a situation that involves simultaneous investment without coordination. Second, we assumed that data on one project is given at zero cost. Extensions could consider the cost of providing such data, and examine the parameters of "optimal" data provision. Such an analysis would allow for data provision costs as well as the possibility of mapping more than one project.

Third, in our experiment, we explicitly revealed a mountain of a specific value, implying that the policy-maker is aware of the potential payoffs before deciding which ones to reveal. This is clearly a simplification and in Appendix A, we present a simple modification of our theoretical model showing that the general dynamics work even when the policy-maker provides data on a project at random. Note that this caveat does not apply to the lab experiment since from the participants' point of view, we might as well have been revealing mountains randomly. Finally, we assumed non-rivalry between participants such that agents' payoffs are independent of others' choices; that is, if multiple agents choose the same project, they all receive the same payoff. We posit that this assumption applies to many settings with technological or knowledge spillovers where agents can learn from each other even if they are not competing directly in the product market (Bloom et al., 2013; Krieger, 2021). Our setup could also be modified to include rivalry by assuming that agents who move later receive only a fraction of the payoffs as compared to agents who choose early. If the degree of rivalry, defined as the payoff penalty applied to agents who move subsequently is small, the $M$-option could still be attractive for later agents, thereby maintaining our baseline predictions. If there is a strong degree of rivalry, it would weaken our results. Investigating the effect of rivalry on the streetlight effect is an interesting direction for future research.

In sum, there is no doubt that the data revolution has the potential to dramatically lower uncertainty and boost investment in risky exploration. However, our work highlights the limits of this logic. When data points to lucrative, but ultimately less-than-ideal projects, they have the potential to do more harm than good by causing agents to herd investment activity and reduce risky exploration. A prime example comes from the recent history of research in treatments for Alzheimer's disease. Promising early data led to the widespread belief in the amyloid-$\beta$ hypothesis, according to which the accumulation of the peptide amyloid-$\beta$ is the main cause of Alzheimer's. The entire field reorganized around this finding, to the extent that other approaches and ideas were often criticized and lacked funding (Makin, 2018). However, more recent findings suggest that these results might be similar to the "medium" condition we highlighted, where it was a promising route for individual scientists, but ultimately for the field, is unlikely to lead to a breakthrough that could treat the condition (Kametani and Hasegawa, 2018). If the results of early experiments were kept secret, firms and researchers might have taken very different approaches, potentially increasing the chances of a breakthrough.

Our paper thus provides support for practices such as "skunkworks," where firms deliberately prevent the diffusion of initial findings of their R&D among their business units, and highlights the role of concealing intermediate information about a project unless it can be confirmed that the project represents a high rather than a medium-value opportunity (Boudreau and Lakhani, 2015). It also highlights the value of "boom" periods in business cycles, when firms are perhaps more risk-loving and invest in more risky projects that can have high societal value (Nanda and Rhodes-Kropf, 2013). More generally, our paper provides behavioral evidence of a manifestation of the streetlight effect in exploration, and highlights the conditions under which it is likely to materialize and when its effects are particularly detrimental to both individual and social welfare. Future work should seek to understand how our theoretical ideas might shape exploration patterns in empirical settings.

# References

Anderson, C. M. (2012): "Ambiguity aversion in multi-armed bandit problems," *Theory and Decision*, 72, 15–33.

Banks, J., M. Olson, and D. Porter (1997): "An experimental analysis of the bandit problem," *Economic Theory*, 10, 55–77.

Bessen, J., S. M. Impink, L. Reichensperger, and R. Seamans (2022): "The role of data for AI startup growth," *Research Policy*, 51, 104513.

Bloom, N., M. Schankerman, and J. Van Reenen (2013): "Identifying technology spillovers and product market rivalry," *Econometrica*, 81, 1347–1393.

Bolton, P. and C. Harris (1999): "Strategic experimentation," *Econometrica*, 67, 349–374.

Boudreau, K. J. and K. R. Lakhani (2015): "'Open' disclosure of innovations, incentives and follow-on reuse: Theory on processes of cumulative innovation and a field experiment in computational biology," *Research Policy*, 44, 4–19.

Boyce, J. R., D. M. Bruner, and M. McKee (2016): "Strategic experimentation in the lab," *Managerial and Decision Economics*, 37, 375–391.

Brynjolfsson, E., D. Rock, and C. Syverson (2021): "The productivity J-curve: How intangibles complement general purpose technologies," *American Economic Journal: Macroeconomics*, 13, 333–72.

Camuffo, A., A. Cordova, A. Gambardella, and C. Spina (2020): "A scientific approach to entrepreneurial decision making: Evidence from a randomized control trial," *Management Science*, 66, 564–586.

Camuffo, A., A. Gambardella, F. Maccheroni, M. Marinacci, and A. Pignataro (2022): "Microfoundations of Low-Frequency High-Impact Decisions," .

Cao, R., R. M. Koning, and R. Nanda (2021): "Biased sampling of early users and the direction of startup innovation," *NBER working paper #28882*.

Chen, D. L., M. Schonger, and C. Wickens (2016): "oTree—An open-source platform for laboratory, online, and field experiments," *Journal of Behavioral and Experimental Finance*, 9, 88–97.

Christin, A. (2020): "What data can do: A typology of mechanisms," *International Journal of Communication*, 14, 20.

Cockburn, I. M., R. Henderson, and S. Stern (2019): "The impact of artificial intelligence on innovation," *The Economics of Artificial Intelligence: An Agenda*, 115–152.

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

Ewens, M., R. Nanda, and M. Rhodes-Kropf (2018): "Cost of experimentation and the evolution of venture capital," *Journal of Financial Economics*, 128, 422–442.

Farboodi, M. and L. Veldkamp (2020): "Long-run growth of financial data technology," *American Economic Review*, 110, 2485–2523.

Fleming, L. (2001): "Recombinant uncertainty in technological search," *Management Science*, 47, 117–132.

Fleming, L. and O. Sorenson (2004): "Science as a map in technological search," *Strategic Management Journal*, 25, 909–928.

Furman, J. L., M. Nagler, and M. Watzinger (2021): "Disclosure and subsequent innovation: Evidence from the patent depository library program," *American Economic Journal: Economic Policy*, 13, 239–70.

FURMAN, J. L. AND S. STERN (2011): "Climbing atop the shoulders of giants: The impact of institutions on cumulative research," *American Economic Review*, 101, 1933–63.

GREINER, B. (2015): "Subject pool recruitment procedures: organizing experiments with ORSEE," *Journal of the Economic Science Association*, 1, 114–125.

GROSS, D. P. (2019): *The consequences of invention secrecy: Evidence from the USPTO Patent Secrecy Program in World War II*, National Bureau of Economic Research.

HEGDE, D. AND H. LUO (2018): "Patent publication and the market for ideas," *Management Science*, 64, 652–672.

HENRICH, J., S. J. HEINE, AND A. NORENZAYAN (2010): "Most people are not WEIRD," *Nature*, 466, 29–29.

HOELZEMANN, J. AND N. KLEIN (2021): "Bandits in the lab," *Quantitative Economics*, 12, 1021–1051.

HOLT, C. A. AND S. K. LAURY (2002): "Risk aversion and incentive effects," *American Economic Review*, 92, 1644–1655.

HÖRNER, J., N. KLEIN, AND S. RADY (2021): "Overcoming free-riding in bandit games," *The Review of Economic Studies*, forthcoming.

HUDJA, S. AND D. WOODS (2021a): "Exploration versus exploitation: A laboratory test of the single-agent exponential bandit model," *Mimeo*.

——— (2021b): "An exploratory analysis of the multi-armed bandit problem," *Mimeo*.

JONES, C. I. AND C. TONETTI (2020): "Non-rivalry and the economics of data," *American Economic Review*, 110, 2819–2858.

KAMETANI, F. AND M. HASEGAWA (2018): "Reconsideration of amyloid hypothesis and tau hypothesis in Alzheimer's disease," *Frontiers in Neuroscience*, 25.

KAO, J. (2021): "Charted territory: Evidence from mapping the cancer genome and R&D decisions in the pharmaceutical industry," *Mimeo*.

KELLER, G., S. RADY, AND M. CRIPPS (2005): "Strategic experimentation with exponential bandits," *Econometrica*, 73, 39–68.

KERR, W. R., R. NANDA, AND M. RHODES-KROPF (2014): "Entrepreneurship as experimentation," *Journal of Economic Perspectives*, 28, 25–48.

KLEIN, N. (2016): "The importance of being honest," *Theoretical Economics*, 11, 773–811.

KLEIN, N. AND S. RADY (2011): "Negatively correlated bandits," *The Review of Economic Studies*, 78, 693–732.

KRIEGER, J. L. (2021): "Trials and terminations: Learning from competitors' R&D failures," *Management Science*, 67, 5525–5548.

LEVINTHAL, D. A. (1997): "Adaptation on rugged landscapes," *Management Science*, 43, 934–950.

MAKIN, S. (2018): "The amyloid hypothesis on trial," *Nature*, 559, S4–S4.

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

——— (2016): "Experimentation and the returns to entrepreneurship," *The Review of Financial Studies*, 29, 2319–2340.

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

Meyer, R. J. and Y. Shi (1995): "Sequential choice under ambiguity: Intuitive solutions to the armed-bandit problem," *Management Science*, 41, 817–834.

Nagaraj, A. (2022): "The private impact of public data: Landsat satellite maps increased gold discoveries and encouraged entry," *Management Science*, 68, 564–582.

Nagaraj, A., E. Shears, and M. de Vaan (2020): "Improving data access democratizes and diversifies science," *Proceedings of the National Academy of Sciences*, 117, 23490–23498.

Nagaraj, A. and S. Stern (2020): "The economics of maps," *Journal of Economic Perspectives*, 34, 196–221.

Nagy, A. and B. Alhatlani (2021): "An overview of current COVID-19 vaccine platforms," *Computational and structural biotechnology journal*, 19, 2508–2517.

Nanda, R. and M. Rhodes-Kropf (2013): "Investment cycles and startup innovation," *Journal of Financial Economics*, 110, 403–418.

Nelson, R. R. (1982): "The role of knowledge in R&D efficiency," *Quarterly Journal of Economics*, 97, 453–470.

Puranam, P. and M. Swamy (2016): "How initial representations shape coupled learning processes," *Organization Science*, 27, 323–335.

Robbins, H. (1952): "Some aspects of the sequential design of experiments," *Bulletin of the American Mathematical Society*, 58, 527–535.

Rosenberg, N. (1992): "Economic experiments," *Industrial and Corporate Change*, 1, 181–203.

Siegel, E. (2013): *Predictive analytics: The power to predict who will click, buy, lie, or die*, John Wiley & Sons.

Thompson, W. R. (1933): "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples," *Biometrika*, 25, 285–294.

Weick, K. E. (1988): "Enacted sensemaking in crisis situations," *Journal of Management Studies*, 25, 305–317.

# The Streetlight Effect in Data-Driven Exploration

## Online Appendix: Additional Results

Johannes Hoelzemann
University of Vienna

Gustavo Manso
UC Berkeley

Abhishek Nagaraj
UC Berkeley

Matteo Tranchero
UC Berkeley

September 29, 2022

## A Revealing the Quality of a Random Project

In this section, we consider social welfare when the quality of a random project is revealed. Under Assumption 1 and using the results of Propositions 2 and 3, the payoff with data is equal to:

$$\frac{m_L}{m}\left(\frac{m_L-1}{m-1}L+\frac{m_M}{m-1}M+\frac{m+m_H-1}{m-1}H\right)+\frac{m_M}{m}\left(2M\right)+\frac{m_H}{m}\left(2H\right).$$

After some algebra, we can show that the payoff with no data is higher than the payoff with data iff:

$$\frac{m_L(m-m_L)}{(m-1)m}L-\frac{m_M(m-1+m_L)}{(m-1)m}M+\frac{(m_H(m_M-m_L)+m_M(m-1-m_H))}{(m-1)m}H>0$$

This result proves that there exist a space of parameters such that our results hold even when the project revealed is chosen randomly.

**Example** We now consider a simple example in which the payoff with no data is higher than the payoff with data. Consider a situation with ten agents and ten projects such that

$n = m = 10$. Assume that there are two low value projects, seven medium value projects, and one high value project. In this setting, Assumption 1 is equivalent to $M > L/2 + H/2$. If we assume $L = 3$, $M = 8$, and $H = 12.5$, Assumption 1 is satisfied and the payoff without data is higher than the payoff with data by $2.2$.

# B Additional Figures and Tables

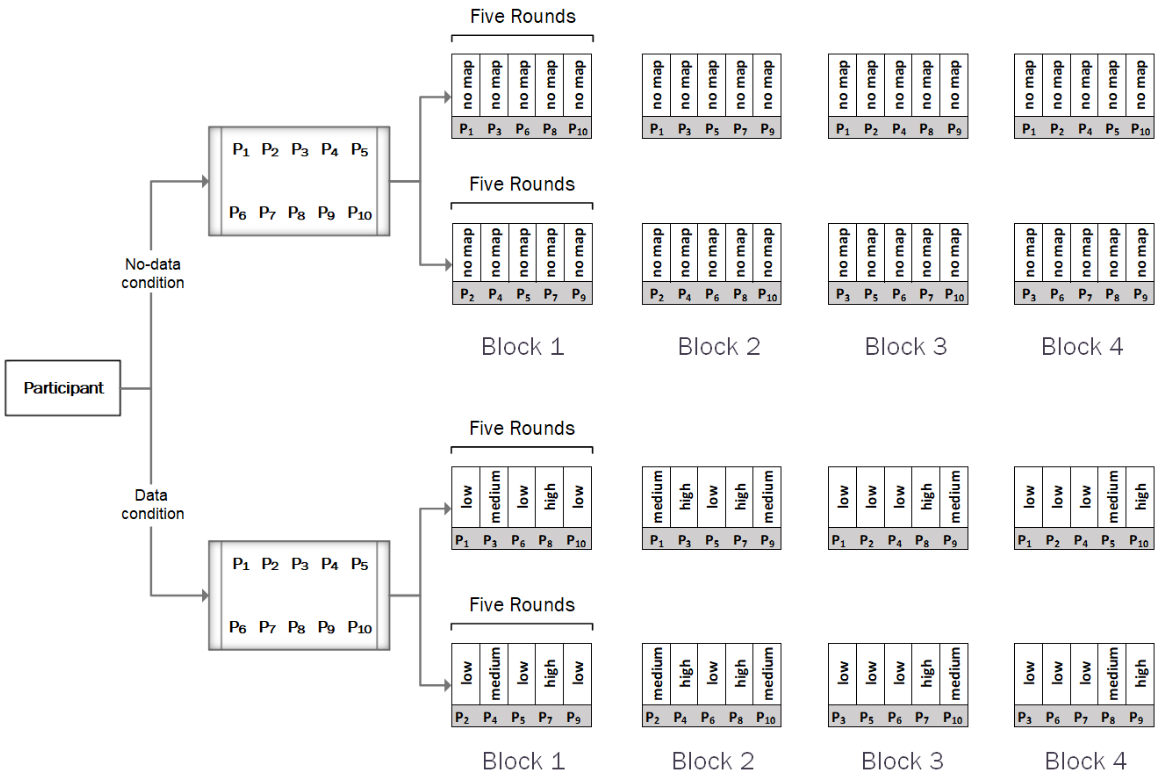



Figure B.1: Flowchart of the experimental setup.

Note: This figure provides an overview of the experiment. When participants join, they are assigned either to a data or to a no-data condition. The experiments begins when a total of ten players are assigned to the same experimental set. Then two groups of five people are randomly drawn to play the first block of five rounds. At the end of the block, the composition of the two groups is randomly reshuffled. This procedure is repeated for four times. The order of blocks seen by participants in different experimental sessions is random.

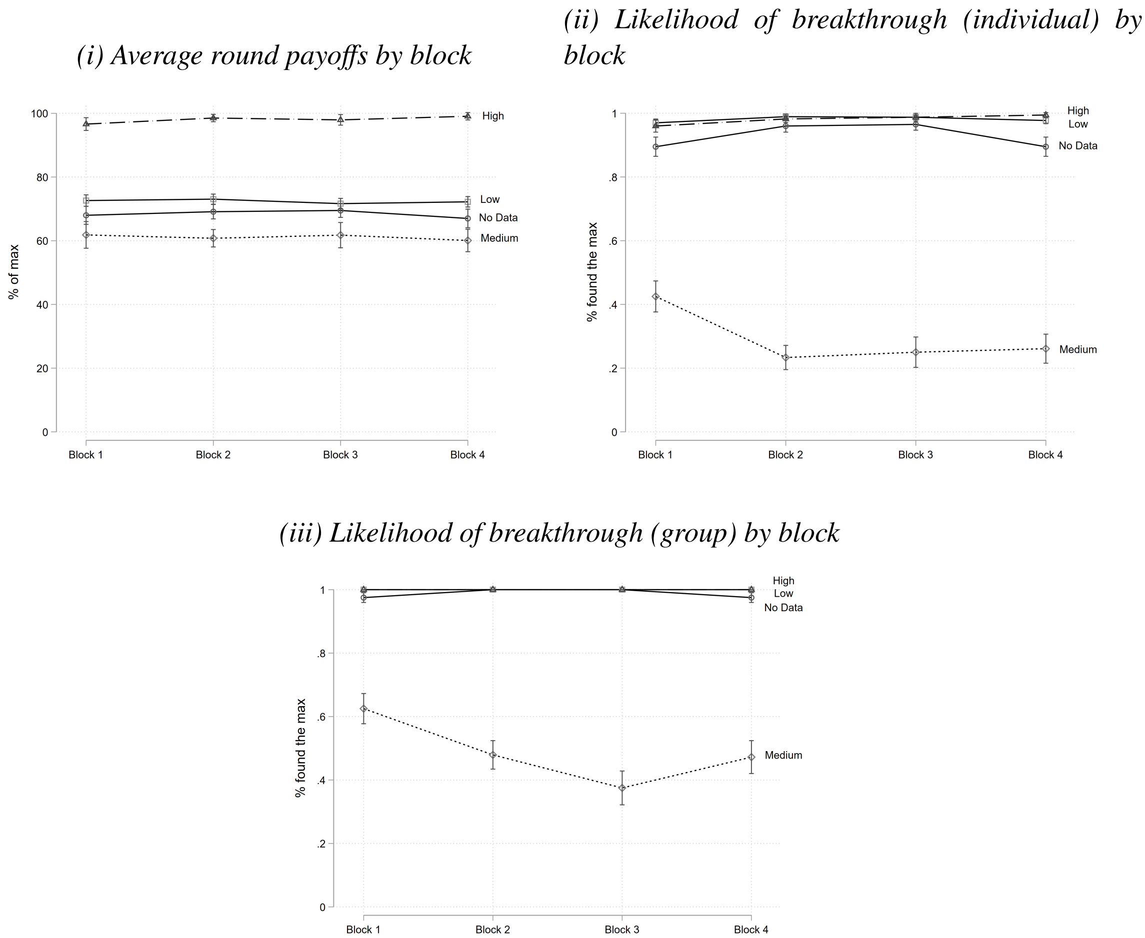
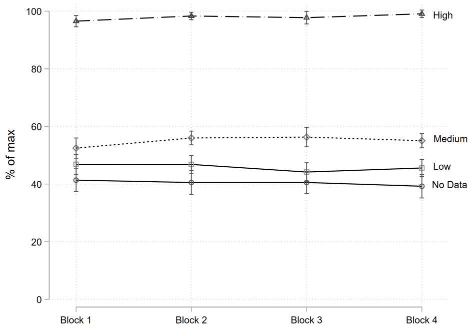



Figure B.2: Robustness of the main results over time.

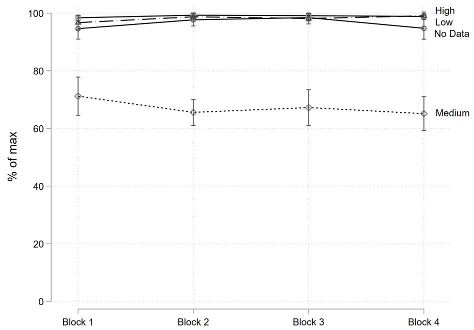
Note: The figures depict the impact of data on group outcomes as the experimental session progresses. Figure (i) shows for each block of 5 rounds the average group payoffs divided by experimental condition. Payoffs are reported as a share of the maximum available in each round. Figure (ii) shows for each block the share of participants who found the location of the maximum divided by experimental condition. Figure (iii) shows for each block the share of rounds where the maximum was found divided by experimental condition.

**Panel A: Payoffs**

*(i) Payoffs in period 1*

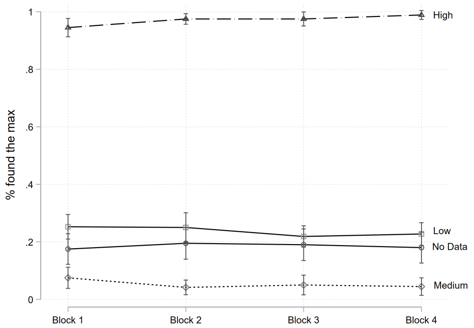



*(ii) Payoffs in period 2*




**Panel B: Breakthroughs**

*(iii) Likelihood of breakthrough in period 1*




*(iv) Likelihood of breakthrough in period 2*

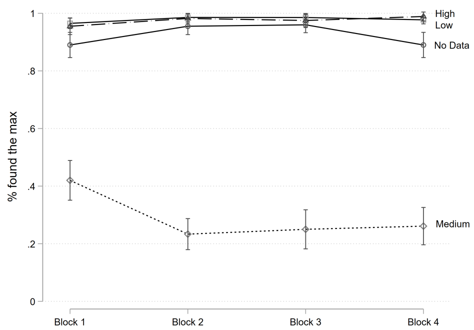



Figure B.3: Outcomes over time and by period of the game.

Note: Panel A reports the experimental results on the period payoffs computed as a share of the maximum possible in each period. Figure (i) shows the average collective payoffs achieved in period 1 by experimental condition and over time. Figure (ii) shows the average collective payoffs achieved in period 2 by experimental condition and over time. Panel B reports the experimental results on the likelihood of a breakthrough in each round. Figure (iii) shows the share of participants that found the maximum in period 1 by experimental condition and over time. Figure (iv) shows the share of participants that found the maximum in period 2 by experimental condition and over time.

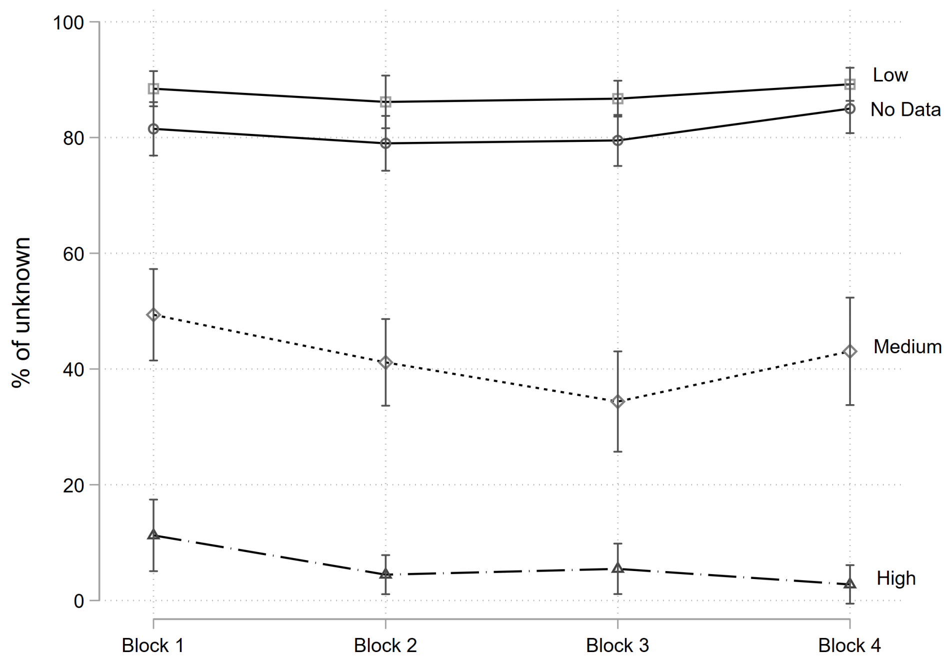



Figure B.4: Average number of mountains explored by experimental condition.

Note: The figure shows for each block of five rounds the impact of data on exploration choices divided by experimental condition. The number of mountains explored is reported as a share of the unknown mountains in each round to account for the fact that rounds without data have one more unknown option.

Table B.1: Risk aversion and decision not to choose the known outcome when medium is revealed

| | I(Not chose Medium in Period 1) | | |
|---|---|---|---|
| Risk aversion (std) | 0.003 | | |
| | (0.022) | | |
| Top quartile risk aversion | | 0.016 | |
| | | (0.048) | |
| Bottom quartile risk aversion | | | -0.050 |
| | | | (0.050) |
| Constant | 0.398** | 0.390** | 0.419** |
| | (0.128) | (0.129) | (0.127) |
| Age FE | Yes | Yes | Yes |
| Gender FE | Yes | Yes | Yes |
| Payoff structure FE | Yes | Yes | Yes |
| Observations | 780 | 780 | 780 |

* * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors clustered at the session level in parentheses

Round-participant level observations, estimates from OLS models. The sample includes all the individual observations for the 156 rounds where the medium value was revealed. *I(Not chose medium in period 1)*:0/1=1 if the player did not choose the medium value in period 1. *Risk aversion (std)* = standardized measure of individual risk aversion (Holt and Laury, 2002); *Top quartile risk aversion*:0/1=1 if the participant is in the top quartile of the risk aversion distribution in our sample; *Bottom quartile risk aversion*:0/1=1 if the participant is in the bottom quartile of the risk aversion distribution in our sample.

Table B.2: Correlates of the decision not to choose the known outcome when medium is revealed

| | I(Not chose medium in period 1) | | | |
|---|---|---|---|---|
| English native | 0.065 | | | |
| | (0.054) | | | |
| Wrong quizzes (std) | | $0.061^{*}$ | | |
| | | (0.025) | | |
| Round number | | | -0.004 | |
| | | | (0.003) | |
| Order of choice | | | | 0.0209 |
| | | | | (0.012) |
| Age FE | Yes | Yes | Yes | Yes |
| Gender FE | Yes | Yes | Yes | Yes |
| Payoff structure FE | Yes | Yes | Yes | Yes |
| Observations | 780 | 780 | 780 | 780 |

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$. Standard errors clustered at the session level in parentheses

Round-player level observations, estimates from OLS models. The sample includes all the individual observations for the 156 rounds where the medium value was revealed. *I(Not chose medium in period 1)*:0/1=1 if the player did not choose the medium value in period 1. *English native*:0/1=1 if the participant is a native English speaker based on her reported nationality; *Wrong quizzes* = standardized number of wrong answers to the initial comprehension test; *Round number* = progressive order in which the rounds were played in the experimental session; *Order of choice* = random sequential order in which the player chose in that round.

# The Streetlight Effect in Data-Driven Exploration

## Online Appendix: Experimental Interface

Johannes Hoelzemann
University of Vienna

Gustavo Manso
UC Berkeley

Abhishek Nagaraj
UC Berkeley

Matteo Tranchero
UC Berkeley

September 29, 2022

# Instructions

## General Information

Welcome. This is an experiment in the economics of decision-making. If you pay close attention to these instructions, you can earn a significant amount of money paid to you at the end of the experiment via Interac e-transfer.

To participate in this online experiment, you will need to use Chrome or Safari on your notebook or personal computer (other browsers and mobile phones are not supported). If you are using a browser or device that is not supported, please copy the experiment link, open one of these supported browsers on a notebook or pc and paste the link into the address bar.

Your computer screen will display useful information. Remember that the information on your computer screen is PRIVATE. To ensure best results for yourself and accurate data for the experimenters, please DO NOT COMMUNICATE or interact with other people on other media at any point during the experiment. If you have any questions, or need assistance of any kind, please call +1-647-606-6469 or use Zoom anytime during the experiment and one of the experimenters will help you privately. We expect the entire experiment to take up to 70 minutes to complete.

Following these instructions, you will be asked to make some choices. There are no correct choices. Your choices depend on your preferences and beliefs, so different participants will usually make different choices. You will be paid according to your choices, so read these instructions carefully and think before you decide.

## The Basic Idea

There are 5 mountains and each of them hides one type of gem, which can only be found by exploring the mountain.


There are 3 types of gems hidden in the 5 mountains:


The exact values of the topazes, rubies, and diamonds vary across rounds but the diamonds are always worth more than the rubies and the rubies are always worth more than the topazes:


You choose which mountains to explore and the value of the gems you find are your earnings in dollars.

## Location of Gems

In each round, there are:

• 3 mountains containing topazes 

• 1 mountain containing rubies 

• 1 mountain containing diamonds 

However, which mountains contain which gems is unknown. At the beginning of each round, each type of gem is randomly assigned to a different mountain, so any gem could be hidden in any mountain. No participant has any initial information in Stage 1 on the location of gems.

## How Participants Choose Mountains

In each round, participants choose which mountain to explore. The choice does not happen simultaneously, but participants choose sequentially, one after the other, according to a random order that changes every round. You can choose to explore any mountain you wish. If you choose the same mountain chosen by other participants, each of you will receive the gem's value uncovered. Similarly, if someone else chooses the same mountain that you previously chose, you will still receive the full gem's value (and so will the other participant(s) that chose it).
To repeat, no participant has any initial information in Stage 1 on the location of gems.

## Each Round Has 2 Stages

A round consists of 2 stages. At the beginning of a new round, gems are randomly allocated to the five different mountains. The position of gems will **not** be reset between the two stages in a round.
In Stage 1, all participants sequentially choose one mountain to explore. Before choosing a mountain, you will see which mountains have been selected by the other participants in your group who chose before you. You can choose the same mountain or a different mountain.
At the end of Stage 1, the gems hidden in each mountain selected by all participants in Stage 1 are revealed, and you earn the value of the gem hidden in the mountain you chose.
In Stage 2, you can again choose any of the same five mountains; that is you can either choose the same mountain of Stage 1 or switch to another one. The position of gems remains the same as in Stage 1, but this time you will also see the gems located in the mountains revealed in Stage 1.
At the end of Stage 2, the gems hidden in each mountain selected by all participants in Stage 2 are revealed, and you earn the value of the gem hidden in the mountain you chose in Stage 2. You will also see your total earnings for the round which equals the sum of the value of the gem you found in Stage 1 and the value of the gem you found in Stage 2.

## Game Structure

The game is divided into 4 blocks, each made of 5 rounds, with each round encompassing the two stages described above. At the beginning of each block, you will be randomly assigned to a new group of 5 participants, with whom you will play for the entire block (5 rounds in total). After the block is complete, you will be randomly assigned to a new group of 5 participants. Again, you will play for 5 rounds. This procedure will be repeated 4 times in total.
You will be reminded of this information in the top-right corner of your screen, as in the example below:

## Payment

**Fixed Participation Fee:** You will earn a participation fee of $5.00 for participating in this experiment.

**Additional Payment and Random Round:** One round will be randomly selected for payment at the end of the experiment. You will be paid and your earnings in that round as described above. Any of the 20 rounds (4 blocks with 5 rounds each) could be the one selected, so you should treat each round as if it will be the one determining your payment.
This protocol of determining payments suggests that you should choose in each round as if it is the only round that determines your payment as the dollar value of the gems you select will directly translate into your earnings.

**Survey and Payment:** In addition to the participation fee and the payment for the randomly selected round, you will perform a small task at the very end of the experiment, and your earnings from this task will be paid to you.

You will be informed of your payment and the round chosen for payment at the end of the experiment. Finally, after completing the experiment you will be paid electronically via Interac e-transfer with the e-mail address you entered on the previous page.

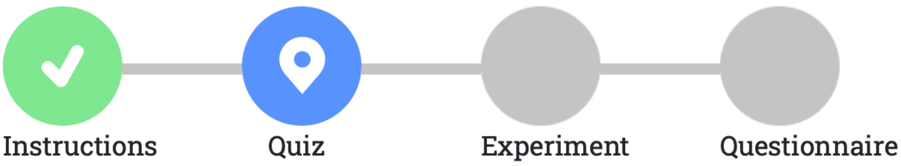
## Frequently Asked Questions

**Q1:** Is this some kind of psychology experiment with an agenda you haven't told us?
**A:** No, it is an economics experiment. If we do anything deceptive or don’t pay you as described, then you can complain to the University of Toronto Research Ethics Board and we will be in serious trouble. These instructions are meant to clarify how you earn money and our interest is in seeing how people make decisions.

**Q2:** Is there a "correct" or "wrong" choice of action? Is this kind of a test?
**A:** No, your optimal choice depends on your preferences and beliefs and different people may hold different beliefs.

Next

This button will be activated after 290 seconds. Please take your time to read through the instructions.




You have successfully finished reading the instructions.

The quiz, consisting of 8 questions in total, follows.

Next

# Quiz Time!

Please mark the following statements as correct/incorrect:

"Q1: In each round, you will select two mountains (one in Stage 1, and one in Stage 2) and collect the gem that they hide. You can choose the same mountain in both stages, or change after Stage 1."

- Correct
- Incorrect

Read Instructions

Next

# Quiz Time!

Please mark the following statements as correct/incorrect:

"Q2: If more than one player selects the same mountain, the value of the gem will be split among all the participants who chose it."

- Correct
- Incorrect

Read Instructions

Next

# Quiz Time!

Please mark the following statements as correct/incorrect:

"Q3: At the beginning of a new round, the gems are reshuffled and randomly allocated to a different mountain."

- Correct
- Incorrect

Read Instructions

Next

# Quiz Time!

Please mark the following statements as correct/incorrect:

"Q4: No group member has any private initial information in Stage 1 on the location of gems."

- Correct
- Incorrect

Read Instructions

Next

# Quiz Time!

Please mark the following statements as correct/incorrect:

"Q5: The position of gems will not be reset between the two stages of a round."

- Correct
- Incorrect

Read Instructions

Next

# Quiz Time!

Please mark the following statements as correct/incorrect:

"Q6: All group members select the mountains simultaneously."

- Correct
- Incorrect

Read Instructions

Next

# Quiz Time!

Please mark the following statements as correct/incorrect:

"Q7: If another group member chose a mountain before you, you cannot choose it again."

- Correct
- Incorrect

Read Instructions

Next

# Quiz Time!

Please mark the following statements as correct/incorrect:

"Q8: At the end of the experiment, one round will be randomly selected for payment."

- Correct
- Incorrect

Read Instructions

Next

Instructions Quiz Experiment Questionnaire

You have successfully finished the quiz.

The experiment follows: When you are ready please click "Next" to start the experiment.

Next

## Start of Block 1

This is Block 1 of 4 and each Block consists of 5 Rounds.

You have been randomly assigned to a **new** group of 5 participants.

Next

## Start of Round 1

You are now in Round 1 of 5 and each Round consists of 2 Stages.

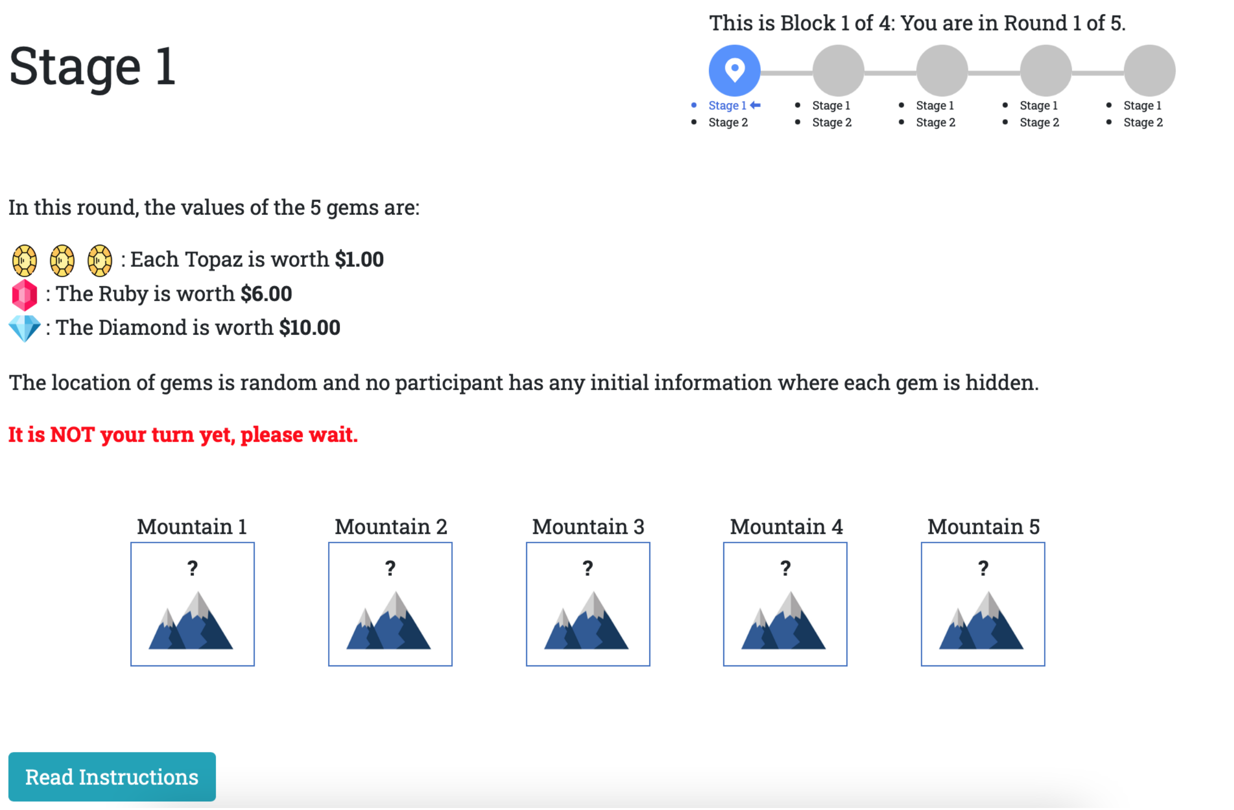
The computer reshuffled the gems. Each gem is randomly allocated to a different mountain.

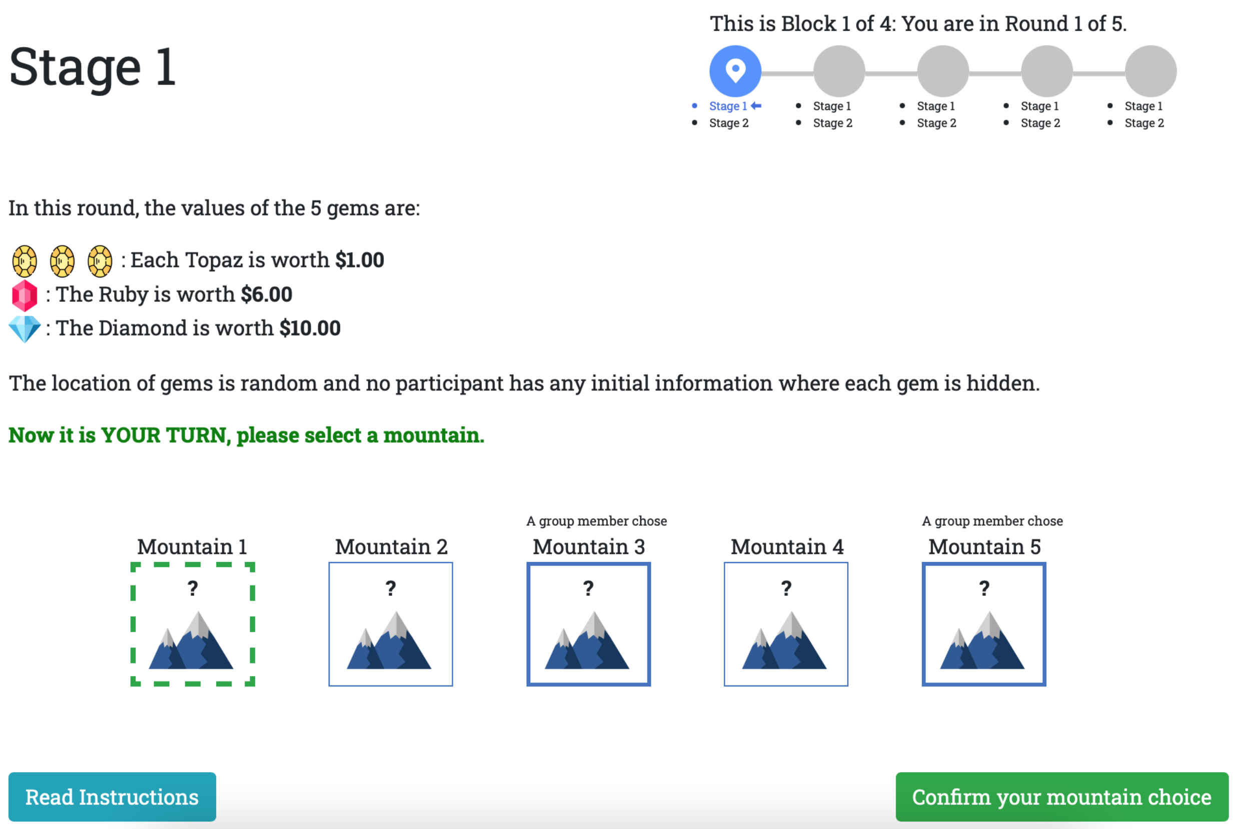
No participant has any initial information on the location of gems.

In this round, the values of the 5 gems are:

: Each Topaz is worth **$1.00**
: The Ruby is worth **$6.00**
: The Diamond is worth **$10.00**

Next


Stage 1
This is Block 1 of 4: You are in Round 1 of 5.
Stage 1
Stage 2
Stage 1
Stage 2
Stage 1
Stage 2
Stage 1
Stage 2
Stage 1
Stage 2
In this round, the values of the 5 gems are:
: Each Topaz is worth $1.00
: The Ruby is worth $6.00
: The Diamond is worth $10.00
The location of gems is random and no participant has any initial information where each gem is hidden.
It is NOT your turn yet, please wait.
Mountain 1
?
Mountain 2
?
Mountain 3
?
Mountain 4
?
Mountain 5
?
Read Instructions


Stage 1
This is Block 1 of 4: You are in Round 1 of 5.
Stage 1
Stage 2
Stage 1
Stage 2
Stage 1
Stage 2
Stage 1
Stage 2
Stage 1
Stage 2
In this round, the values of the 5 gems are:
: Each Topaz is worth $1.00
: The Ruby is worth $6.00
: The Diamond is worth $10.00
The location of gems is random and no participant has any initial information where each gem is hidden.
Now it is YOUR TURN, please select a mountain.
Mountain 1
?
Mountain 2
?
A group member chose
Mountain 3
?
Mountain 4
?
A group member chose
Mountain 5
?
Read Instructions
Confirm your mountain choice

# Stage 1: Earnings

You selected Mountain 1 and found a Topaz. Thus, you earned **$1.00** from your choice.

All discovered gems and their locations are highlighted below.
These will also be displayed in Stage 2 when you make your next choice.

Click "Next" to proceed to the next stage.


Read Instructions

Next

# Stage 2

This is Block 1 of 4: You are in Round 1 of 5.

- Stage 1 ✔
- Stage 2 ⬅

- Stage 1
- Stage 2

- Stage 1
- Stage 2

- Stage 1
- Stage 2

- Stage 1
- Stage 2

In this round, the values of the 5 gems are:

: Each Topaz is worth **$1.00**
: The Ruby is worth **$6.00**
: The Diamond is worth **$10.00**

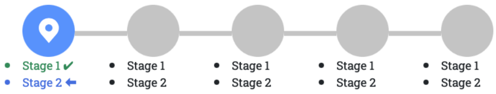
**Now it is YOUR TURN, please select a mountain.**


Read Instructions

Confirm your mountain choice

# Stage 2

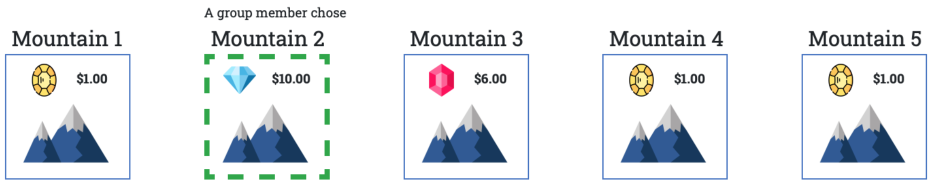
This is Block 1 of 4: You are in Round 1 of 5.




In this round, the values of the 5 gems are:

: Each Topaz is worth **$1.00**
: The Ruby is worth **$6.00**
: The Diamond is worth **$10.00**

**Now it is YOUR TURN, please select a mountain.**




Read Instructions

Confirm your mountain choice

# Stage 2: Earnings

You selected Mountain 2 and found a . Thus, you earned **$10.00** from your choice.

Your total earnings from both stages in this round are **$1.00 + $10.00 = $11.00**

All discovered gems and their locations in both Stages are highlighted below.

Please click "Next" to proceed to the next round.


Read Instructions

Next

Instructions Quiz Experiment Questionnaire

You have successfully finished the main part of the experiment.

A brief questionnaire together with a short task follows: When you are ready please click "Next" to start the experiment.

Next

## Please answer the following questions

Your answers will be kept confidential and will not affect your earnings for today's experiment.

Please state your age:

Please state your gender:

---------

Please state your student type:

---------

Please state your country of origin:

Please state your degree and field of study:

Please briefly explain, in your own words, the rules of today's experiment:

Please briefly describe how you reached your decisions in this experiment:

Please briefly describe how, in your opinion, other participants reached their decisions in this experiment:

Next

# Instructions

Thank you for your participation so far. In the last task of the experiment, you will earn an additional reward based on a set of 10 choice problems.

**How does it work?**

The Choice: You will be asked to choose between two options, **"Option A"** and **"Option B"** where:

- "Option A" always pays `$4.00` with probability `p` and `$3.20` otherwise.
- "Option B" always pays `$7.70` with probability `p` and `$0.20` otherwise.

Repeated Choices:

- You will be asked to make a choice between "Option A" and "Option B" not once, but **ten** times where `p` will increase from `10%` to `100%`, `10%` at a time.

  For example, the first choice will have `p=10%` and you will choose whether you prefer "Option A" (`$4.00` with a `10%` chance or `$3.20` otherwise) or "Option B" (`$7.70` with a `10%` chance or `$0.20` otherwise).

- Each successive choice will increase `p` by `10` percentage points until the last choice where "Option A" will pay `$4.00` with certainty, and "Option B" will pay `$7.70` with certainty.

Note: Once you switch from choosing "Option A" to "Option B", it makes sense that you will continue to choose "Option B" in all consecutive choice problems. For example, if you prefer "Option B" when `p=80%`, then it makes sense to prefer "Option B" when `p=90%` and when `p=100%`, since "Option B" is even more attractive in these choice problems.

Therefore, we have designed the interface so that you must either (a) **always** choose "Option A" or "Option B" for all 10 choice problems or (b) if you **switch** to "Option B" for a given probability `p`, then you must choose "Option B" for all the following choices as well.

You can adjust your choices as many times as you wish. When you are ready to submit your choices, you can click on the "Next" button at the bottom of the page.

**Payment**

The computer will randomly select one of the 10 choice problems and pay you according to your choice in that problem where the computer will decide the outcome based on the value of `p`.

Next

# Please Choose Between "Option A" and "Option B" on Every Line

| | Option A | Option B | |
|---|---|---|---|
|  | **$4.00** with a chance of 10%, **$3.20** otherwise | **$7.70** with a chance of 10%, **$0.20** otherwise |  |
|  | **$4.00** with a chance of 20%, **$3.20** otherwise | **$7.70** with a chance of 20%, **$0.20** otherwise |  |
|  | **$4.00** with a chance of 30%, **$3.20** otherwise | **$7.70** with a chance of 30%, **$0.20** otherwise |  |
|  | **$4.00** with a chance of 40%, **$3.20** otherwise | **$7.70** with a chance of 40%, **$0.20** otherwise |  |
|  | **$4.00** with a chance of 50%, **$3.20** otherwise | **$7.70** with a chance of 50%, **$0.20** otherwise |  |
|  | **$4.00** with a chance of 60%, **$3.20** otherwise | **$7.70** with a chance of 60%, **$0.20** otherwise |  |
|  | **$4.00** with a chance of 70%, **$3.20** otherwise | **$7.70** with a chance of 70%, **$0.20** otherwise |  |
|  | **$4.00** with a chance of 80%, **$3.20** otherwise | **$7.70** with a chance of 80%, **$0.20** otherwise |  |
|  | **$4.00** with a chance of 90%, **$3.20** otherwise | **$7.70** with a chance of 90%, **$0.20** otherwise |  |
|  | **$4.00** with a chance of 100%, **$3.20** otherwise | **$7.70** with a chance of 100%, **$0.20** otherwise |  |

Read Instructions

Next

# Thank you for participating in this experiment!

Your payoffs for this experiment are as follows:

Main Experiment:

- **Round 1** of **Block 1** was randomly selected for payment.
- In Stage 1, you found a  and received **$1.00** and in Stage 2, you found a  and received **$10.00**
- Thus, your total payoff is $1.00 + $10.00 = **$11.00**

Last Task of Experiment:

- The following choice problem was randomly selected:

| Option A | | Option B |
|---|---|---|
| $4.00 with a probability of 60%,<br>$3.20 otherwise | ○ ◉ | $7.70 with a probability of 60%,<br>$0.20 otherwise |

- As indicated above, you chose **Option B**. The computer drew a random number to determine your payoff according to the chances specified.
- Your payoff is **$7.70**

Participation Fee:

- You earned a fee of **$5.00**

In total, you earned **$11.00 + $7.70 + $5.00 = $23.70** from your choices.

You will receive your payment as an Interac e-transfer. If you encounter any problems, please contact Johannes Hoelzemann at j.hoelzemann@utoronto.ca or +1-647-606-6469.