Strategic Experimentation and Information Design in Dynamic Contests: An Experimental Study

Yan Chen* Mohamed Mostagir† Iman Yeckehzaare‡§

Abstract

Many real-world innovation contests and R&D races have an end goal that might be infeasible. Participants learn about feasibility from their own experimentation and also from observing the progress (or lack thereof) of their competitors. If participants incorrectly learn that the goal is infeasible, they quit the contest and abandon an innovation that could have been achieved. In this paper, we design a novel real-effort experiment to show that a contest designer can avert this undesirable outcome through her choice of information mechanism. By allowing participants to monitor each other's progress, either fully or partially, she significantly increases the chances that the innovation is obtained when it is indeed feasible and when the common prior belief about infeasibility is high. We show that competitor behavior and contest outcomes are sensitive to the timing at which information is released, and we discuss how different information mechanisms affect the designer's payoffs and participants' earnings. Our results provide the first experimental test of the role of information in environments that combine strategic experimentation and dynamic competition, and offer concrete guidelines for how practitioners and applied researchers should select a contest's information mechanism in order to maximize the chances of innovation.

Keywords: information design, dynamic contest, laboratory experiment

JEL Classification Numbers: C73, C91, D83

* School of Information, University of Michigan, 105 South State Street, Ann Arbor, MI 48109.
† Ross School of Business, University of Michigan.
‡School of Information, University of Michigan, 105 South State Street, Ann Arbor, MI 48109.
§We thank Paul Hur, Sarah Jabbour, and Victor Zeng for excellent research assistance.

1 Introduction

Contests are increasingly used as a mechanism for solving innovation problems. These problems are typically characterized by a degree of uncertainty regarding the feasibility of their end goal, i.e. there is a possibility that the goal is unachievable despite competitors putting effort towards reaching it. The Netflix Prize and the Heritage Prize offer examples of such competitions, with contrasting outcomes. For the Netflix Prize, participants were asked to develop a recommendation algorithm that improves on Netflix's in-house algorithm by a margin of at least 10% RMSE (root-mean square error). Netflix acknowledged this goal uncertainty in their contest announcement, stating "We suspect the 10% improvement is pretty tough, but we also think there is a good chance it can be achieved. It may take months; it might take years." Despite this uncertainty, the goal was achieved and the prize was awarded in 2009 (Koren, 2009). By contrast, the Heritage Prize offered $3 million dollars for a solution that uses patient data to predict hospital readmission rates, with a pre-specified target prediction error. "Can we identify earlier those most at risk and ensure they get the treatment they need? We believe that the answer is 'yes'." The contest ended after two years when none of the approximately 13,000 participants was able to achieve that goal (El Emam et al., 2012).

This uncertainty about an underlying state of the world is a defining characteristic of the strategic experimentation and multi-armed bandit literature (see Hörner and Skrzypacz (2016)), where agents learn about the state from their own experimentation and possibly from the experimentation of others. The possibility that the goal is infeasible introduces an extra risk factor that participants incorporate into their effort decisions. Innovation might fail to happen either because the goal is actually infeasible, or because players become pessimistic about the prospects of feasibility and drop out of the contest, abandoning a feasible innovation in the process. Understanding the role of this *infeasibility risk* on participation and effort is therefore important for designing contests that maximize the chances of obtaining the innovation whenever it is indeed feasible.

While there exists a theoretical literature that studies this infeasibility risk in environments that combine strategic experimentation *and* competition, e.g. Choi (1991); Malueg and Tsutsui (1997); Halac et al. (2017); Bimpikis et al. (2019), there is no experimental study that examines how players respond in such settings. Our paper closes this gap through a novel experimental design that captures the essential elements from the theoretical literature, tests how they affect subjects' behavior, and derives the corresponding implications for contest designers.

Our experiment models a winner-takes-all contest with two sequential stages. A contestant must finish both stages to win the prize, but there is a chance that the first stage cannot be completed (and hence no one can collect the prize). Following Choi (1991), we interpret the first stage as representing a *breakthrough* discovery, whereas the second stage represents a final discovery or an innovation that can only be made if the breakthrough discovery is achieved. In that sense, a breakthrough discovery resembles a "proof-of-concept," and signals to others that the underlying task is possible. In the words of Dasgupta and Maskin (1987): "...to know that someone has solved a problem is to know a great deal: specifically, that the problem is solvable." This phenomenon is observed in a variety of domains.

For example, the Twin Prime Conjecture witnessed a flurry of activity once a breakthrough was made in 2013 (Nature, 2013). Notably, the current best solution uses a method that is different from the one used in the breakthrough, suggesting that these breakthroughs not only serve as methodological advances, but also provide a feasibility signal that adjusts the risk level downwards for those who are interested in working on the problem. Similarly, while the aforementioned Netflix Prize was for breaking the 10% improvement barrier, there was also an intermediate milestone and associated prize for breaking the 5% barrier, which roughly correspond to the two stages of our experiment.

The downside to breakthrough discoveries is when they *do not* happen, or more accurately, when they do not happen fast enough. This is an issue in the larger strategic experimentation literature where agents learn from their own outcomes and also from the outcomes of others. Observing that no one has made a breakthrough makes players more pessimistic about feasibility compared to when they learn only from their own experimentation. As a result, players' beliefs about feasibility may sharply decline and they can stop experimenting and quit the contest. Because of this, several papers have suggested that obscuring this information channel or shutting it down altogether – by having players not observe each other's outcomes – can alleviate this effect and ultimately lead to an increase in the likelihood of a breakthrough in case the stage is indeed feasible (e.g. Halac et al. (2017) and Bimpikis et al. (2019) in a competitive multi-armed bandit setting, and Kremer et al. (2014); Papanastasiou et al. (2018) in an explore/exploit multi-armed bandit setting).

Our focus is on the information structure of the contest, which determines what players know about each other and when they know it. The competitive setting we consider introduces an extra effect to learning about breakthroughs: while they provide encouragement about feasibility, they also discourage players who learn that they now trail their opponent and are therefore less likely to win. Which of these effects dominate depends on the level of risk in the environment and the information mechanism that the contest designer implements.

We test three information mechanisms that mirror those often found in practice and studied in previous literature. The *Daylight* mechanism is equivalent to a real-time leaderboard where the status of competitors is immediately updated, so that players always know where they are with respect to their opponents. The Heritage Prize used such a mechanism. By contrast, the *Darkness* mechanism is equivalent to a design with a single grand prize, where the only information revealed is whether the contest is over because someone reached the goal. TopCoder, the world's largest competitive software development portal, offers the use of this information mechanism (Archak, 2010). A design that is in between these two extremes is a *Silent Period* design. This design commits to revealing information only after a certain time has elapsed. For example, while the award for the aforementioned Netflix prize was for breaking the 10% improvement barrier, an intermediate prize for breaking the 5% barrier was scheduled to be announced and awarded one year after the start date of the competition. Thus the contest starts out with a known and fixed future date where the status of the competition –specifically, whether a breakthrough was achieved– is broadcast to all players, provided that the entire contest has not been completed by someone before then.

Our goal is to study the interaction between infeasibility risk and the contest information mechanism to experimentally answer the following questions: as a function of the risk level in the environment, 1) how do players respond to the different information mechanisms, 2) which mechanisms maximize the designer's payoff by increasing the likelihood that breakthroughs and innovations are made, and 3) how do the different mechanisms affect participants' earnings?

Contribution and Results: Our paper contributes to the experimental literature at the intersection of R&D incentives, contest design, and strategic experimentation and statistical learning. Infeasibility risk is a distinguishing factor of many R&D environments, however, to our knowledge, there exists no other experimental work in this setting.

Our first contribution is a novel experimental design that embeds infeasibility risk in a dynamic setup by using the knapsack problem.[1] The problem has been argued to serve as a good proxy for innovation and intellectual discovery (e.g. Arthur and Polak (2006); Youn et al. (2015)) and has been successfully used to model such tasks in previous laboratory experiments, e.g. Meloso et al. (2009). We incorporate it into our design by making the first stage of our contest a knapsack problem whose parameters are tuned in such a way that the problem may or may not have a solution. The knapsack is particularly suited to our purposes because it has no known algorithmic solution aside from brute force search. This means that its feasibility cannot be easily determined based on a player's own experimentation, which increases the value of information from learning about other players' outcomes. We further motivate our choice of the knapsack problem as well as compare it to other tasks from the literature in Section 4.

Our second contribution is a characterization that is useful for applied researchers, practitioners, and further modeling work in this area. We answer the questions posed above by showing the effects of the different information mechanisms as a function of infeasibility risk in the environment. In particular, when that risk level is high, there is a strong chance that players abandon the search for the innovation when the goal is actually feasible. We show that in this case, mechanisms that allow players to observe each other, i.e., the Daylight and Silent Period mechanisms, significantly increase the chances that the innovation is reached, i.e., we find no evidence in our setup that supports the theoretical findings that shutting down the information channel about opponents' (lack of) progress helps with increasing the chances of a breakthrough.

The effectiveness of the Silent Period design has been theorized in the literature (see Hörner and Skrzypacz (2016)), but without a characterization of its optimal duration. Our experiment empirically tests several durations to show that performance is quite sensitive to how that duration is set. Specifically, we show that not all silent periods are created equal: when the duration is chosen appropriately, the mechanism significantly outperforms Darkness and Daylight by boosting the chances of breakthroughs and innovations. However, a silent period that is "too long" delivers the worst outcomes for the designer. This suggests that when infeasibility risk is high, the designer should be careful about which mechanism to implement based on the knowledge she has about the environment and

[1] The knapsack problem is a combinatorial optimization problem with the goal of finding, in a set of items of given values and weights, the subset of items with the highest total value, subject to a total weight constraint (Murawski and Bossaerts, 2016).

the task at hand, and Section 5.2 offers concrete guidelines about how that choice should be made. Finally, Section 5.3 provides an overview of how welfare and players' earnings are affected by the information mechanism used.

The rest of the paper is organized as follows. We review the related literature in Section 2. Section 3 summarize the common features from the theoretical literature upon which we base our hypotheses. Section 4 describes the design of our experiment and Section 5 presents our findings. Section 6 concludes the paper by commenting on current limitations and avenues for future work.

2 Literature Review

Our study combines three elements in innovation contests: strategic experimentation, dynamic competition, and information design. The literature at the intersection of these topics is relatively recent. On the theory side, an early work is that of Choi (1991), who studies the strategic experimentation and competition aspect in a two-stage model similar to the one in our paper, but does not study the information aspect. Halac et al. (2017), Bimpikis et al. (2019), and Mihm and Schlapp (2019) are three recent papers that study the role o f information in a setting that resembles that of Choi (1991). Halac et al. (2017) show that when players can split the contest prize, providing no information about the status of the contest increases the chances of a breakthrough. Bimpikis et al. (2019) use a different award structure yet corroborate the finding that not revealing information increases the chances of breakthroughs, and that partial revelation might be useful when the designer is not only interested in reaching the innovation but also having it happen as quickly as possible.

With a slightly different model, Mihm and Schlapp (2019) study a similar two-stage contest like the one modeled in the above papers, with one of their main conclusions echoing the finding that full information revelation hinders innovation. In the words of the authors, "contests looking for breakthrough innovation should rely solely on private feedback," where private feedback in their setting is similar to contestants learning only from their own experimentation, which is equivalent to the Darkness mechanism in our paper.

As noted, there is no experimental work that combines the three elements studied in the above papers. All the related literature we review below combine at most two of those three elements.

Competition: Our work is related to the stream of research on dynamic races, as exemplified by the seminal theoretical work of Harris and Vickers (1987) and the experiments in Zizzo (2002) and Mago et al. (2013). Competitors race through a series of stages and receive full feedback regarding the outcome at the end of each stage, and these papers characterize how effort changes in terms of the gap between players. This line of research only considers the competition aspect and does not incorporate strategic experimentation or the role of information.

Competition and information provision: The next strand of related literature combines competition and information provision to study the effects of feedback on contestant behavior. Feedback in this case provides information about how players compare with their competitors. This area spans empirical, theoretical, and experimental work. On the empirical side, Bothner et al. (2007); Genakos and Pagliero (2012) and Gauriot and Page (2019) use data from different sports competitions to study how players respond to feedback by changing their risk taking behavior (e.g. take more risky actions if they are behind or if they are only slightly leading). Similarly, Genakos et al. (2015) document the effect of feedback in sports competition and show that the leader exerts less effort.

On the experimental side, Casas-Arce and Martinez-Jerez (2009) show that interim performance feedback signals to subjects how skilled they are relative to their competitors, and therefore can make both leaders and laggards exert less effort when they realize the gap is large. Chaudoin and Woon (2018) experiment with a setting where subjects obtain information about how their competitor played in the past. Ederer and Fehr (2007) find that agents discount feedback from a principal when there is a chance the feedback is not truthful.

In addition to the experimental literature based on abstract tasks, our study shares similarities with experiments based on real effort tasks. In an experiment examining the effect of feedback on the racing choices of high school runners, Fershtman and Gneezy (2011) find that continuous feedback on opponent performance increases the likelihood a runner will quit compared to the case when this information is not available. In one study using piece-rate and rank-order tournaments, Eriksson et al. (2009a) study the impact of relative performance feedback on subject performance on an addition task under the same three information conditions we study in this paper. While they find no effect under the piece-rate scheme, they find positive peer effects in the rank-order tournament, that is, those who trail remain in the contest and leaders sustain their effort. Conversely, Gill and Prowse (2012) find that when subjects play sequentially, disappointment aversion might make the follower reduce her efforts.

Strategic experimentation: Our study differs from the work above in that we incorporate the strategic experimentation aspect into the environment. This means that information not only teaches subjects about their opponents, but also about the underlying bandit problem. Bolton and Harris (1999) were the first to study this problem in a non-competitive setting, and Keller et al. (2005) build on their work to develop the workhorse model used in most of the strategic experimentation literature. There is a scarcity of experimental work on this topic until recently, such as Hoelzemann and Klein (2018) and Boyce et al. (2016), who study whether subjects play according to the equilibrium predictions in Keller et al. (2005). We are not aware of other experimental work that studies strategic experimentation in the presence of competition as we do in this paper.

Contest design: Within the large experimental literature on contests (Dechenaux et al., 2015; Sheremeta, 2018), our study contributes to the literature on the design of innovation contests. In this domain, Taylor (1995) analyzes a multi-period tournament game where there is no pre-specified goal, but an award is given out to the agent who obtains the best outcome. A common question in this literature is how many agents to allow into the contest. Terwiesch and Xu (2008) provide a theoretical categorization of when the benefits of increased participation mitigate the negative effect on average participant effort that result from increased competition. Boosey et al. (2017, 2019, 2020) examine how subjects respond when the number of participants is uncertain or unknown, and how efforts change when subjects are informed about the number of participants in a contest with endogenous entry.

In an empirical study, Jeppesen and Lakhani (2010) examine the effect of uncertainty on effort using projects on the innovation contest platform, InnoCentive, and corroborate the predictions of Terwiesch and Xu (2008) that open participation and participant diversity have a positive effect on solving highly uncertain problems. A subsequent field experiment on TopCoder (Boudreau and Lakhani, 2015) studies the effects of similar information disclosure regimes as we consider, namely, intermediate disclosure where solutions are posted immediately after they are submitted, final disclosure where no information on other contestants' progress is released until the end of the contest, and a mixed regime where participants compete under final disclosure in the first half and intermediate disclosure in the second half of the contest. However, their problem is known to be solvable and can be improved incrementally, and the prize is given to the contestant with the best solution which is evaluated by a panel of judges. Similarly, Wooten and Ulrich (2017) report a field experiment on logo design contest under three feedback conditions: no feedback, random feedback on an idea, and directed feedback on a contestant's submitted entry. The latter is positively associated with agent participation and the average quality of submitted entries.

Lastly, while our setup examines a competitive setting in an uncertain environment, the role of feedback about relative performance was also studied as a mechanism to bolster productivity in organizations. Song et al. (2017) show that providing public feedback about top performers in a hospital has a measurable positive effect on worker productivity. In a similar vein, recent work by Cadsby et al. (2019) examines effects of private versus public feedback about relative ranking on worker and student performance, and how these effects are moderated by risk attitudes.

3 Theoretical Framework

In this section, we outline our set of hypotheses based on a simplified version of several theoretical models in the literature (Choi, 1991; Keller et al., 2005; Halac et al., 2017; Bimpikis et al., 2019). In some of these models, agents have to be incentivized (through payments) to share information about their progress. For example, in a risky environment, an agent who has made progress might not share this information in order for her competitor to become more pessimistic about feasibility and quit. These incentive requirements add a layer of complexity that we simplify away in our experiment, since we observe everything that subjects do on their computers. This allows us to focus on the effects of the different information mechanisms without worrying about these incentive issues.

Setup: We consider an innovation contest with two identical agents, 1 and 2, competing for a prize, R. The contest consists of two *sequential* stages. Completing Stage A represents a breakthrough discovery, and Stage B represents a final discovery —the innovation goal— that can only happen if a

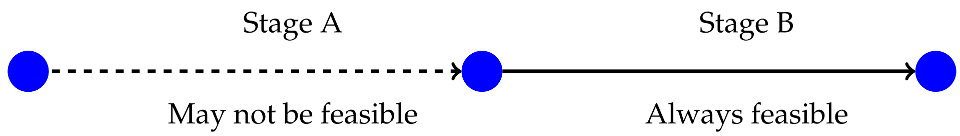

Figure 1: A dynamic contest with two stages, A and B. A breakthrough discovery happens if Stage A is completed, and the innovation goal is achieved if someone completes Stage B, which can only happen if a breakthrough has been made.

breakthrough is made. The winner, if there is one, is the agent who finishes both stages first. Stage A may or may not be feasible, while Stage B is always feasible conditional on Stage A having been completed. Thus the prize cannot be collected if Stage A is infeasible. Agents have a common prior q_A on the infeasibility of Stage A, i.e. $q_A = Pr(\text{Stage A is infeasible})$. Thus q_A serves as a measure of risk for players; As q_A increases, the level of risk that players take on increases. In our experiment, we will vary q_A to take values 0 (no risk), 20% (mild risk) and 40% (high risk).

The strategic experimentation literature assumes that discoveries and breakthroughs follow a stochastic arrival process, e.g., a Poisson process (Keller et al., 2005), whose rate is modulated by the amount of effort agents exert (Choi, 1991; Halac et al., 2017; Bimpikis et al., 2019). Similar to these papers, it is possible, with probability q_A, that the arrival rate of Stage A is zero, in which case no amount of effort that the agents exert can lead to a breakthrough. Information, if it is provided, lets players learn their relative positions (which stage each of them is in), and therefore makes them update their beliefs about the feasibility of Stage A. For example, learning that a competitor has completed Stage A immediately resolves any infeasibility risk, whereas if no information is provided a player who is still in Stage A has to update her belief using only her own experience.

Information: The information mechanisms we study are **Daylight**, **Darkness**, and **Silent Period**. In the Daylight mechanism, an agent is informed as soon as her opponent completes Stage A, while under Darkness no information is revealed unless one of the agents has completed the entire contest. Silent Period is an intermediate mechanism where no information is revealed prior to a certain pre-announced time.

Players know in advance that Stage B is always feasible, thus, our focus in on the interplay that occurs within Stage A between the risk level (measured by q_A) and the information provision mechanism. Specifically, we are interested in how players react to different risk levels in the presence of competition, and which information mechanisms make them more likely to give up and quit Stage A or persevere in the face of infeasibility risk and achieve a breakthrough (when Stage A is indeed feasible). Note that while most of the action happens in Stage A, Stage B is still necessary for two reasons: first, when the contest consists of a single stage, a player in the Darkness treatment can still learn from her opponent's experience, just by virtue of the fact that the contest is not over (i.e., the opponent has not been able to complete the stage yet).[2] Second, Stage B allows us to capture the

[2] An exception can be found in Halac et al. (2017), who examine the Darkness design but allow play to continue even after some players have already achieved the goal.

encouragement and competition effects by examining whether players in Stage A choose to continue competing or exit upon receiving news that the stage is indeed feasible (good news) but that they are also now trailing behind their opponent (bad news).

Designer Objective: The designer is interested in maximizing the chances that the final discovery is made and that this happens in the shortest possible time. In particular, assuming the utility from discovery is U, then the designer gets discounted payoff $e^{-r\tau}U$ for some discount rate r and time of discovery τ. Similar to Halac et al. (2017), in the absence of discounting, this objective reduces to maximizing the probability that the innovation goal is reached.

This objective is closely tied to agents' quitting behavior. If both agents quit, then clearly the designer's payoff is zero. Having both agents exert effort in the early, risky stage doubles the rate of the underlying arrival process and increases the chances of a breakthrough (assuming independent discoveries and that the stage is indeed feasible).

The results below —drawn from several papers in the theory literature that considers this or similar models— describe how players behave in the absence of risk (Proposition 1), how risk changes this behavior (Proposition 2), and how the different information mechanisms modify these effects (Proposition 3).

The following hypotheses is based on results from Section 5 in Bimpikis et al. (2019) and Section 4 in Choi (1991):

Proposition 1 (Environments Without Risk)**.** *In the absence of risk, i.e., when $q_A = 0$:*

(i) Under the Daylight mechanism, both players continue playing in the contest until one player finishes Stage A. At this point, the trailing player quits the contest.

(ii) Under the Darkness mechanism, no player quits and the contest continues until someone achieves the goal and completes the contest.

The above proposition implies that, when there is no risk in the environment, Stage A will be finished regardless of the information mechanism used. In such an environment, learning that an opponent has made a breakthrough has no upside since players already know that Stage A is feasible, and the player trailing her opponent quits when she learns that she is behind. This will later serve as the basis for Hypothesis 1.

On the other hand, the next result follows from the first two propositions of Bimpikis et al. (2019) and Halac et al. (2017), and shows that in the presence of risk, a breakthrough (and consequently, innovation) is not always going to happen — even when the stage is actually feasible. Players become pessimistic about feasibility and quit, and this effect is stronger the higher the prior belief q_A is. This serves as the basis of Hypothesis 2.

Proposition 2 (Effect of Risk)**.** *When $q_A > 0$, contestants follow a cutoff experimentation policy under both the Daylight and Darkness mechanisms: contestants exert full effort until a specific time (different for each* *mechanism), at which point, they quit if a breakthrough has not happened. Further, the likelihood of quitting Stage A increases with the risk level,* q_A.

The next statement compares the effects of information mechanisms in the presence of risk, and is based on Propositions 3, 5, and 6 in Bimpikis et al. (2019). The first part states the familiar result that "no news is bad news": when players do not experience a breakthrough and also see their opponents not making a breakthrough, they become pessimistic about feasibility and quit the contest. This effect is weakened (and players quit less often) when they only learn from their own experimentation. Therefore, when Stage A is feasible, we should expect to see more players making breakthroughs under the Darkness mechanism compared to the Daylight mechanism. When the risk level q_A is high enough, the Silent Period mechanism outperforms both of these mechanisms and maximizes the chances that the innovation is obtained. this forms the basis of Hypothesis 3.

Proposition 3 (Effect of Different Information Mechanisms)**.** *In the presence of risk:*

$i)$ *The probability that the first stage is completed is higher under the Darkness mechanism.*

$ii)$ *There is a value* $\bar{q}$ *such that if* $q_A > \bar{q}$*, the Silent Period design with an appropriately chosen duration outperforms both the Daylight and Darkness mechanisms: that is, it increases the chances a player breaks through to Stage B and, conditional on someone reaching Stage B, it increases the chances that both players remain in the contest in Stage B.*

The theoretical framework outlined in this section forms the basis for our experimental design and the corresponding hypotheses.

4 Experimental Design

We use a 3×3 factorial design to investigate the effects of information provision mechanisms and infeasibility risk on contestant behavior. The information provision factor represents the type of mechanism that provides information about competitor progress, with possible mechanisms of **Darkness**, **Daylight**, and **Silent Period**, whereas the risk factor represents the different levels of risk about the feasibility of the problem being solved, with possible values of **No**, **Mild**, and **High** risk.

Task Selection. Infeasibility risk is a defining characteristic of many R&D and strategic experimentation settings. This aspect is unique to our experiment and therefore requires a novel task compared to prior research. Consider the following tasks that have been used in the dynamic contests experimental literature: adding randomly-generated numbers (Eriksson et al., 2009b), choosing a number to indicate level of effort (Ederer and Fehr, 2007), physically running a race (Fershtman and Gneezy, 2011), moving a slider to a correct location on a screen (Gill and Prowse, 2012), etc. These tasks may vary in difficulty but their feasibility is never in question. While appropriate for the environments modeled in these papers, they are not suitable for a setting that requires varying levels of risk about task feasibility, which is the core issue at the heart of our paper.

To this end, we utilize the knapsack problem in our design. The problem has been successfully used as a proxy for an innovation task in previous laboratory experiments to compare whether patents or markets are better information aggregators for tasks requiring creativity (Meloso et al., 2009). The knapsack problem is especially well suited to our purposes because it has no known algorithmic solution aside from brute force search, and therefore its feasibility cannot be easily verified based on a player's own experience, which increases the value of information from learning about other players' outcomes.

We use the knapsack problem in our design as follows: we provide subjects with an instance of the problem consisting of a weight constraint C and a set of items, with each item having a non-negative weight and value. We ask subjects to find a subset of these items that fit in the knapsack and has total value greater than or equal to a target y. The problem is *infeasible* if no such subset exists. We determine whether a problem is feasible or not by tuning the target y, as we describe in the next section. Figure 2 shows the interface and an example of a feasible problem.

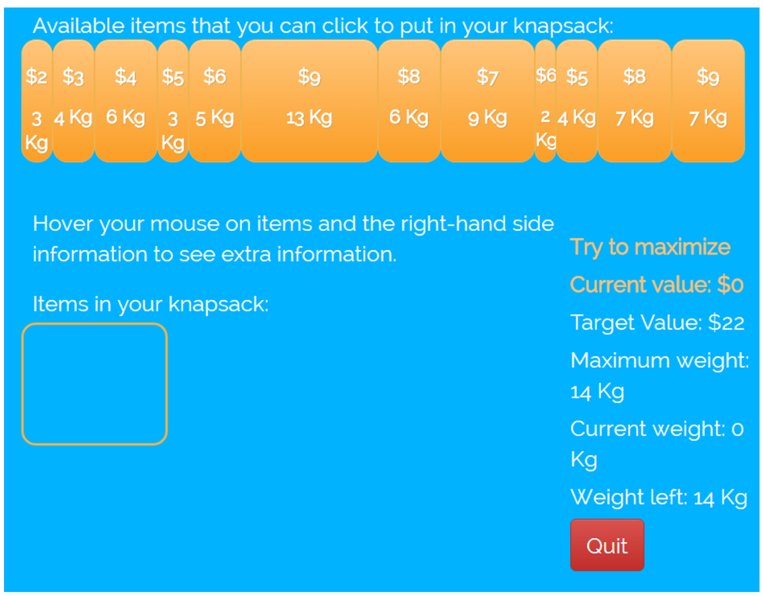

Figure 2: Interface for the knapsack problem. Items and values are shown at the top, with item sizes commensurate with their weights. Subjects click on items to insert them into or out of the knapsack (whose size is also commensurate with its capacity). The right side of the interface summarizes the parameters of the problem and dynamically updates the value and weight as items are dragged into or out of the knapsack. Subjects have the option to quit at any point by pressing the Quit button.

4.1 Treatments

Our experiment implements a 3 (information provision mechanisms) $\times 3$ (infeasibility risk levels) factorial design to evaluate the performance of the three information provision mechanisms under different risk levels. In what follows, we explain our design choices.

Information Provision Mechanisms (between-subject): As mentioned, we implement three information provision mechanisms. However, each subject experiences only one of three mechanisms to avoid cognitive overload. In market and mechanism design experiments, it is often common to use a between-subject design across mechanisms to avoid information overload and confusion (Kagel and Roth, 2000; Chen and Sönmez, 2006; Cason et al., 2020).

- In the Daylight treatment, subjects receive information on the progress of their opponent at all times, so that both contestants know exactly which stage the other one is in throughout the contest duration (including if the opponent has quit).

- In the Darkness treatment, subjects receive no information about their opponent unless one of them finishes both stages and wins the contest.

- In the Silent Period treatment, subjects are informed at the beginning of the contest that they will learn about their positions only after a certain pre-specified amount of time has elapsed (provided neither subject has finished the entire contest nor both subjects have quit prior to the announcement time).

  As theory does not provide sufficient guidance on the optimal revelation time, we rely on experimental data to inform the optimal design of the Silent Period mechanism, in the spirit of an engineering approach treating the laboratory as a wind tunnel (Roth, 2002). To examine the effect of different revelation times on the choices that players make, we select a few times for each game in the Silent Period and randomize among these selected times whenever contestants play that particular game. To obtain our set of possible revelation times, we conduct two pilot sessions under the Darkness treatment. In this treatment, we construct an empirical distribution for each game of the time it takes the first player to complete Stage A in the Darkness pilot sessions. We then choose times at the $50^{th}, 60^{th}, 70^{th}$, and 80^{th} percentiles of that distribution. Thus, for each game i, we have a set of four possible revelation times T_i, corresponding to the values described above. A time is chosen randomly from this set whenever this game is played in the Silent Period treatment.

Infeasibility risk (within-subject): For each information provision mechanism, we implement three infeasibility risk levels within subject, i.e., for each contest, each subject experiences the three risk levels with equal probability. The choice of a within-subject design along the risk factor is based on two considerations. First, using a blocked random assignment on each subject increases statistical power without substantially increasing their cognitive load or confusion (Friedman and Sunder (1994), Chapter 3), as varying probabilities is easy to explain to the subjects. A secondary benefit of a within-subject design is that it reduces payment variance across subjects. The three infeasibility risk levels include:

- No risk ($q_A = 0\%$): This is a benchmark case where it is always possible to finish Stage A (and consequently, the contest);

- Mild risk ($q_A = 20\%$): This risk level represents a mild chance that Stage A is infeasible;
- High risk ($q_A = 40\%$): This risk level represents a high chance that Stage A is infeasible.

To randomize the risk treatments, each contest is equally likely (with probability $1/3$) to be no, mild, or high risk. In the treatment with no risk, the target value for the first game of a contest is the optimal solution to the knapsack problem, i.e. it is not possible to fit items worth more than this value in the knapsack. This means that subjects play the games exactly as written in Appendix C. If the contest is under the mild risk treatment, we set its parameters as follows: a number in $(0, 1)$ is generated uniformly at random prior to the contest, and if the number is less than or equal to 0.2, then we make the first game of the contest infeasible. Subjects are presented with a game whose target value is slightly-inflated above the optimal solution;[3] if the number is above 0.2 then the game is feasible and the target value is again equal to the optimal solution. Subjects play the game knowing that they are in the first scenario (game is infeasible) with probability 20% or in the second scenario (game is feasible) with probability 80%. A similar procedure is used for the high risk treatment, but with a threshold of 0.4 instead of 0.2.

In the corresponding part of the Experimental Instruction (Appendix B), we tell the subjects, "before you start the contest, you will be told that the chance Game 1 is feasible is 100% (definitely feasible), 80%, or 60%. Before each contest, you will see one of the following sentences displayed on the screen:"[4]

"For the next contest, the chance that the first game is feasible is 100%."

"For the next contest, the chance that the first game is feasible is 80%."

"For the next contest, the chance that the first game is feasible is 60%."

Contests where Stage A is infeasible cannot be completed since it is not possible for any player to reach Stage B. In all treatments, since Stage B is always feasible, a subject who completes Stage A no longer has any residual risk about the feasibility of the contest. The structure of the contest is summarized in Figure 3, which is also the diagram shown to subjects when they receive the instructions for this phase of the experiment.

4.2 Experimental Procedure

Each session of our experiment consists of 12 subjects, who are students at a large public university. Subject characteristics are summarized in Table 8 in Appendix A. After completing a training phase to become familiar with the task, each subject participates in three contests. A subject completes the

[3]How we increment the target depends on the values in the problem. For example, in Contest 1 Game 1 in Appendix C, the values of all items end in either 0 or 5, and hence the modified infeasible target should also end in one of these values since otherwise it is clear that it is not possible to fit that amount into the knapsack.

[4]Note that we frame the infeasibility risk in terms of chances of feasibility in the instructions, which facilitates subject comprehension.

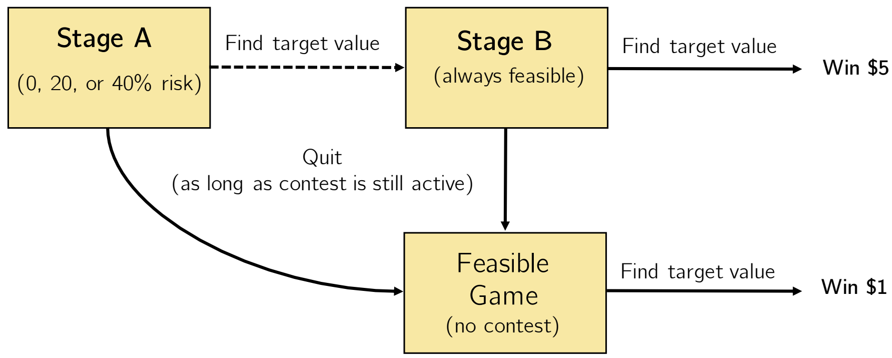

Figure 3: Contest Progress Diagram. The dotted arrow between Stage A and Stage B indicates that progress may not always be possible, due to the chance that Stage A can be infeasible (in the mild and high uncertainty treatments). A player who finishes both stages earns $5, while a player who quits at any time while the contest is active, i.e. the opponent has not completed Stage B, gets to play a non-competitive consolation game that earns them $1.

session after completing both the training and contest phases. We describe the details of the training and contest phases next.

Training Phase. Before the contest begins, subjects are individually presented with an identical sequence of six training games of increasing difficulty. There is no direct measure of how difficult an instance of a (feasible) knapsack problem is, but a recent framework for modeling its complexity is provided in Franco et al. (2018). We use one of the measures defined in Meloso et al. (2009).[5] Each game is a knapsack problem with a given target value y. The games in the training phase are always feasible and subjects are informed of this fact in the beginning. The training games are designed to achieve two goals: to familiarize subjects with the knapsack problems and to measure each subject's skill level. Subjects are paid $1 per successfully completed game, which is announced before the training phase starts.

In a given game, if a subject is unable to fit the target value into the knapsack, they can skip that game and receive no payment for it. Subjects have a total of 30 minutes to complete the training games. To represent each player's skill level, let s_{ij} be the score of player i in training game j, where $s_{ij} = 1$ if the subject finds the subset of items that achieve the target value and $s_{ij} = 0$ otherwise. Let t_{ij} be the time it takes player i to solve training game j, then the (normalized) skill of player i is measured according to the following formula :

$$skill(i) = \sum_{j=1}^{6} \frac{\frac{s_{ij}}{t_{ij}} - \min_{i'} \frac{s_{i'j}}{t_{i'j}}}{\max_{i'} \frac{s_{i'j}}{t_{i'j}} - \min_{i'} \frac{s_{i'j}}{t_{i'j}}} \in [0, 6].$$

[5]Under this measure, a problem's difficulty is a function of the binary representation of the *input size* (where the input is the knapsack capacity and the number of items), which is a proxy for the the amount of information to keep track of in each step of computation.

Statistic	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Skill	2.649	0.935	0.702	1.984	3.292	5.382

Table 1: Distribution of Skills

i.e., subjects' skills take into account whether they solved the problem and how long it took them to do so (with lower times indicating higher skills). Summary statistics about skills are in Table 1.

Once the training phase is complete, subjects are presented with the instructions for the contest phase and given a short multiple-choice quiz that ensures they understand these instructions. Subjects receive 20 cents for each correct answer. They can earn up to $1 for answering all quiz questions correctly. The contest phase then starts.

Contest Phase. After the training phase is completed, subjects are divided into three groups of four subjects each. Each group consists of subjects with adjacent skills. Thus, the first group contains the four subjects with the lowest skills, the second group the next four subjects and so on. We group subjects of similar skills together to approximate the assumption of identical skills found in most of the theory literature. Subjects are informed that they will be competing against players of similar skills prior to starting the contest. Specifically, we instruct the subjects, "Based on your performance in the training phase, you will be placed in a group of 4 players with similar skill levels." (see Appendix B for instructions).

Within each group, subjects participate in contests in a round-robin fashion, so that each subject plays three contests, one against each member of their group. Therefore, we observe a total of 6 contests in a group of four contestants and 18 contests per session. The problems that players face in each contest are given in Appendix C. We selected these problems from Meloso et al. (2009) based on the criterion that they are of similar difficulty. The problems are presented in the same order, but the risk treatments are randomized (within-subject) for each of the three contests. This means that a problem may or may not look exactly like the problem in the appendix (recall that we modify problems when they are infeasible). In the subsequent analysis, we mitigate potential order effects by controlling for problem difficulty and game fixed effects.

Each contest consists of two stages. Similar to Section 3, we refer to the first stage as Stage A and to the second stage as Stage B. Each stage consists of a single knapsack problem and is completed if a contestant finds the target value y. Stage B cannot be started unless Stage A is finished. The contestant who first completes Stage B wins a prize of $5 while her opponent receives nothing.

Subjects can choose to quit the contest at any point while it is active (i.e., as long as no one has completed Stage B), and exercise an outside option. In this option they solve a (feasible and non-competitive) knapsack problem similar to the training games for a prize of $1. A player who does not quit and continues to play until her opponent has completed both stages receives $0. Thus the three possible outcomes and payoffs for any player are: i) win the contest and receive $5, ii) quit while the

Information Provision	Infeasibility Risk 0%	20%	40%	Total no. of contests (feasible)
Daylight	59 (59)	57 (49)	58 (37)	174 (145)
Darkness	57 (57)	57 (48)	60 (34)	174 (139)
Silent Period	60 (60)	60 (44)	60 (35)	180 (139)
Total	176 (176)	174 (141)	178 (106)	528 (423)

Notes: Our dataset has a total of 352 subjects participating in 3 contests each, for a total of 1056 observations under the 9 treatment conditions. The two numbers in each treatment represent the total number of contests in that treatment and, in parentheses, the number of contests that were feasible in that treatment.

Table 2: Features of Experimental Conditions

contest is active and get \$1 for completing the feasible game or \$0 otherwise, or iii) neither win nor quit, and receive \$0 once the other contestant has completed the entire contest.

Table 2 summarizes the features of experimental conditions. For each of the three information provision mechanisms, we conduct 10 independent sessions at the Behavioral Economics and Cognition Experimental Lab at a large public university in the United States. As mentioned, each session consists of 12 subjects.[6] No subject participates in more than one session. This design gives us a total of 30 independent sessions and 352 distinct subjects.[7]

Our subjects were recruited using ORSEE (Greiner, 2015) as well as a separate subject pool from the business school at the university. Sample demographics by treatment can be found in Table 8 in Appendix A. The last column reports the p-value of joint orthogonality tests, which indicates that our randomization leads to a balanced assignment in most observable characteristics ($p > 0.05$) except for the Daylight treatment having a higher proportion of undergraduates ($p = 0.038$) . The sessions comprise a total of 528 contests, including 174 Daylight, 174 Darkness, and 180 Silent Period contests. Of these contests, 423 are feasible, with 145 (139, 139) under the Daylight (Darkness, Silent Period) condition.

At the end of the experiment, each subject fills out a demographic and strategy survey on the computer and is then paid in private. Each experimental session lasts approximately 53 minutes. The average payment is \$16.65, including a \$5 show-up fee. The experiment is programmed in Python (Django framework), HTML, CSS and JavaScript.

The experimental instructions and the list of knapsack games are included in Appendices B and C, respectively. The software is archived on Github.[8] Data are available from the authors upon request.

[6] Two sessions had 11 subjects show up, which meant it was only possible to form two groups of size 4 each. In both of these sessions, the experimenter joined the remaining three subjects in order to get the number up to four. We discarded the data from the two groups where the experimenter participated from our analysis.

[7] Before starting the sessions, we ran two pilot sessions to test the software and the interface. Data from the pilot sessions are not included in the analysis, except for a robustness check in Appendix A.

[8] The authors will provide a link to the software hosted on Github prior to publication. The link cannot be provided during the review process as it will violate anonymity.

	All	No Risk			Mild Risk			High Risk		
		Darkness	Daylight	Silent Period	Darkness	Daylight	Silent Period	Darkness	Daylight	Silent Period
Quit	0.235	0.096	0.254	0.142	0.219	0.306	0.284	0.368	0.365	0.186
	(0.424)	(0.297)	(0.437)	(0.350)	(0.416)	(0.463)	(0.454)	(0.486)	(0.485)	(0.392)
Breakthrough	0.905	0.965	0.966	0.95	0.896	0.898	0.909	0.706	0.865	0.886
	(0.292)	(0.186)	(0.183)	(0.220)	(0.309)	(0.306)	(0.291)	(0.462)	(0.347)	(0.323)
Contest Completed	0.803	0.912	0.814	0.917	0.833	0.735	0.727	0.676	0.730	0.771
	(0.397)	(0.285)	(0.393)	(0.279)	(0.377)	(0.446)	(0.451)	(0.475)	(0.450)	(0.426)
Contest Duration (seconds)	311.397	313.423	348.683	263.794	292.431	372.820	401.988	276.656	328.469	340.142
	(287.828)	(283.812)	(316.102)	(217.333)	(255.068)	(359.862)	(273.558)	(162.804)	(299.584)	(300.399)
Time Until Quit (seconds)	210.234	170.153	217.111	283.257	223.370	222.460	242.457	155.879	174.492	200.147
	(198.012	(180.608)	(209.152)	(207.364)	(228.464)	(219.458)	(194.943)	(145.218)	(204.018)	(154.406)
Number of Observations	846	114	118	120	96	98	88	68	74	70
Number of Contests	423	57	59	60	48	49	44	34	37	35

Notes: This table reports means and standard deviations (in parentheses) for the variables used in the main analysis. For individual-level outcomes (Quit and Time Until Quit), each observation is a subject playing a feasible first stage game. For contest-level outcomes, each observation is a contest.

Table 3: Summary Statistics of Main Outcome Variables

5 Results

We now report our experimental results. Recall that we are interested in how infeasibility risk and competition influence individual behavior and how the different information mechanisms modulate this behavior. What makes subjects quit or continue to exert effort? Which mechanisms should the designer choose as a function of the risk in the environment? To address questions on individual behavior (Section 5.1), *we focus on contests where Stage A is feasible* (but subjects may be uncertain about this fact), since if Stage A is infeasible then there is only one possible outcome: no breakthroughs are made and both subjects eventually quit. Table 3 presents the mean and standard deviation of key outcome variables by experimental condition for the 846 feasible Stage A games in our dataset.

We describe how subjects play under the different treatments in Section 5.1. We then use these results to explore the effects of the different mechanisms on the designer's objective in Section 5.2, and discuss welfare and subject earnings in Section 5.3.

5.1 Individual Behavior

We consider two outcome variables: **Quit** is a subject-level binary variable that is equal to 1 if the subject quits Stage A and 0 otherwise. **Breakthrough** is a stage-level binary variable that is equal to 1 if Stage A is completed by at least one subject and 0 otherwise. These two variables allow us to capture the fact that a similar number of agents quitting across treatments does not necessarily imply that the number of breakthroughs are also similar across these treatments.[9]

Table 4 reports logistic specifications with the two outcome variables described above and independent variables that include treatment dummies and their interactions (with Darkness and No Risk being the omitted categories for these variables throughout the paper). Recall that we implemented four versions of the Silent Period mechanism based on our pilot data under Darkness, with revelation

[9]For example, it is possible that Treatments T_1 and T_2 have the same number of subjects quitting but that Treatment T_1 has twice as many breakthroughs as T_2.

time set to the 50^{th}, 60^{th}, 70^{th} or 80^{th} percentile of the empirical distribution of the time it takes the first player to complete Stage A for each knapsack game. In the Silent Period treatment, one of the four revelation times is randomly selected for that game. As each revelation time represents a version of the Silent Period mechanism, we bundle them into two different classes, with the 50^{th} and 60^{th} percentiles representing a short silent period, and the 70^{th} and 80^{th} percentiles representing a long one. We include separate dummies depending on whether the duration of the silent period is short (**Short SP**) or long (**Long SP**). For completeness, the results when using a single dummy that incorporates all silent period durations are included in Table 10 in Appendix A.

We capture the effect of competition by including a dummy variable **Trail** that is equal to 1 if a subject is trailing her opponent (i.e. opponent is in Stage B while subject is still in Stage A), and we interact this variable with the information mechanisms to understand how competition and information combine to affect outcomes. We also add control variables that capture the fixed effects of the specific game being played, the order of the game, and subject random effects. In all specifications, robust standard errors are clustered at the group level, as each group of four contestants interacts only within its group during a session.

We start by discussing the quitting behavior of subjects as a function of competition, risk, and the information mechanism used. Idiosyncratic factors aside, that subjects quit when they feel they will not be able to achieve the goal and collect the prize for the contest. The hypotheses in Section 3 suggest that this can happen for two reasons, either a) competition: subjects see that they trail their opponent (i.e. the opponent has already completed Stage A) and therefore their own chances of winning are slimmer, or b) risk: subjects think the stage is infeasible and therefore there is no reason to continue playing. We would like to tease out which effect dominates as we vary the risk level and information mechanism used.

5.1.1 Competition

A natural benchmark is to start with contests where subjects face no infeasibility risk, so that we can isolate the interaction between information and competition. The following hypothesis, based on Proposition 1, summarizes what we expect to see in this scenario, and is reminiscent of the discouragement effects documented in the literature (see Section 2).

Hypothesis 1 (Effect on Information Mechanisms in the Absence of Risk)**.** In the absence of risk, and *as long as no player has advanced*, information mechanisms have no effect on quitting or breakthroughs. However, the Daylight treatment increases the likelihood that those who are trailing their opponent will quit the contest.

Our first result supports this prediction.

Result 1. In the absence of risk, breakthroughs are equally likely under all three information mechanisms. However, subjects who trail their opponents are on average 42 percentage points (p.p.) more likely to quit under Daylight.

	Dependent variable:		
	Quit (Subject-Level)	Quit (Subject-Level)	Breakthrough (Stage-Level)
	(1)	(2)	(3)
Daylight	0.198^{**}	-0.001	0.002
	(0.079)	(0.077)	(0.080)
Short SP	0.124	0.077	-0.103
	(0.109)	(0.111)	(0.097)
Long SP	0.008	-0.070	0.017
	(0.098)	(0.082)	(0.099)
Mild Risk	0.148^{*}	0.136^{*}	-0.102
	(0.081)	(0.075)	(0.088)
High Risk	0.294^{***}	0.268^{***}	-0.296^{***}
	(0.084)	(0.080)	(0.110)
Mild Risk × Daylight	-0.099	-0.077	-0.001
	(0.070)	(0.074)	(0.100)
Mild Risk × Short SP	-0.0173	-0.015	0.056
	(0.112)	(0.106)	(0.059)
Mild Risk × Long SP	0.013	0.025	-0.005
	(0.119)	(0.113)	(0.101)
High Risk × Daylight	-0.143^{***}	-0.113^{*}	0.060
	(0.0519)	(0.061)	(0.059)
High Risk × Short SP	-0.243^{***}	-0.242^{***}	0.116^{***}
	(0.022)	(0.047)	(0.019)
High Risk × Long SP	-0.035	-0.012	0.092
	(0.123)	(0.123)	(0.121)
Trail		-0.088	
		(0.060)	
Trail × Daylight		0.425^{***}	
		(0.098)	
Trail × Short SP		0.097	
		(0.140)	
Trail × Long SP		0.210	
		(0.167)	
Baseline (Darkness + No Risk)	0.107^{***}	0.13^{***}	0.96^{***}
	(0.013)	(0.02)	(0.002)
No. of observations	846	846	432

Notes: Treatment effects on quitting (subject-level) and breakthroughs (stage-level). The omitted categories throughout are Darkness and No Risk. Column (2) adds omitted category Trail=0. *p<0.1; **p<0.05; ***p<0.01.

Table 4: Treatment Effects on Quitting and Breakthrough: Logistic Specifications

Support. Specification (3) in Table 4 shows that, in the absence of risk, there are no significant differences between the information mechanisms in affecting the chances of a breakthrough ($p > 0.1$ for the Daylight and both Silent Period coefficients). However, Specification (1) shows that subjects are significantly more likely to quit under Daylight compared to Darkness (0.198, $p = 0.011$). Specification (2) shows that this extra quitting can be attributed to subjects who trail their opponents: the coefficient of Daylight in Specification (2) is no longer significant when considering subjects who are not trailing (-0.001, $p > 0.1$) but the interaction between Daylight and Trail shows that this combination increases the probability of quitting by 42 p.p. (0.425, $p < 0.01$). ■

Takeaway: In the absence of risk, breakthroughs happen regardless of the information mechanism used. Subjects who trail their opponent quit when they learn this information under Daylight.

5.1.2 Infeasibility Risk

We next examine the effect of infeasibility risk and how it interacts with the information mechanism used. Proposition 2 implies Hypothesis 2.

Hypothesis 2 (Effect of Infeasibility Risk)**.** As the level of infeasibility risk increases, players quit more often and breakthroughs are made less often.

Result 2. Under Darkness, Mild Risk weakly increases the chances of quitting but has no effect on breakthroughs, whereas High Risk increases the likelihood of quitting by 29.4 p.p. and decreases the likelihood of breakthroughs by 29.6 p.p. compared to No Risk.

Support. Specification (1) in Table 4 shows that, as the level of risk changes from No to Mild, the likelihood that subjects quit weakly increases by 14.8 p.p. ($p = 0.068$). High Risk increases the probability that subjects quit by 29.4 p.p. ($p < 0.01$). Specification (3) in Table 4 shows that Mild Risk has no significant effect on breakthroughs ($p > 0.1$), but that High Risk decreases the likelihood of breakthroughs by 29.6 p.p. ($p < 0.01$). ■

Takeaway: Consistent with expectations, increasing the risk level leads to more quitting and less breakthroughs. An interesting question is whether the behavior of subjects is not a response to risk, but rather, a response to the decreased level of (expected) pay, i.e., if similar results can be obtained by considering certainty-equivalent contests with reduced levels of pay and no risk. In Appendix A, we perform a robustness check using two sessions of our pilot data that had the prize set to $4 instead of $5.

Having established how risk affects the behavior of subjects, our next hypothesis describes how the different information mechanisms modulate this behavior. As the level of risk increases and players cannot make a breakthrough, they start adjusting their beliefs about feasibility downwards. Observing, under Daylight, that their opponent is also not making a breakthrough doubles the rate at which these beliefs decline. Therefore, under High Risk and in the absence of breakthroughs, we expect to see more players quit under Daylight. Delaying this observation through introducing a silent period can help decrease quitting and increase breakthroughs.

The previous discussion, which is based on Proposition 3, implies the following hypothesis.

Hypothesis 3 (Effect of Information Mechanisms in the Presence of Risk)**.** (a) As the risk level increases, quitting becomes more likely and breakthroughs less likely under Daylight compared to Darkness. (b) When the risk level is high, the Silent Period mechanism leads to less quitting and more breakthroughs compared to either Darkness or Daylight.

The next result shows that we observe the opposite behavior from Hypothesis 3(a), and that Hypothesis 3(b) crucially depends on the duration of the silent period:

Result 3. While High Risk increases quitting by an average of 29.4 p.p. under Darkness, the Daylight mechanism reduces this likelihood by 14.3 p.p. The Short SP almost erases the effects of high risk, reducing quitting by 24.3 p.p. By contrast, the Long SP is no different from the Darkness mechanism. Likewise, High Risk decreases the chances of breakthroughs by 29.6 p.p. under Darkness, whereas the Short SP restores some of these chances, increasing them by an average of 11.6 p.p. Similar to the case when there is no risk, there are no differences between the information mechanisms when the risk level is mild.

Support. Specification (1) in Table 4 shows that the coefficient of Daylight $\times$ High Risk is equal to 0.143 ($p < 0.01$). The coefficient of Short SP $\times$ High Risk is equal to -0.243 ($p < 0.01$). For breakthroughs, Specification (3) in Table 4 shows no differences between Daylight and Darkness under High Risk. The coefficient of Short SP $\times$ High Risk is significant ($0.116, p < 0.01$). There are no differences between the information mechanisms when the level of risk is mild ($p > 0.1$ for all interactions with Mild Risk).

■

Result 3 shows that subjects behave differently from the theoretical prediction that states that, in the presence of risk, shutting down observability can reduce quitting and increase breakthroughs. This does not happen in our experiment: compared to the base case, High Risk increases quitting under Darkness compared to Daylight. In the theoretical models, social information provided by Daylight is only used to adjust beliefs about the infeasibility risk, whereas in many laboratory and field experiments subjects also use social information to infer their relative standing and the socially desirable actions (Hopkins and Kornienko, 2004; Chen et al., 2010). Under Daylight, it is plausible that a contestant stays in the game after observing that her opponent is still in the game, inferring that they have comparable skills and that it is socially desirable not to quit. While our experiment is not designed to separate these different channels, it is an interesting open question for future research.

The Short SP mechanism outperforms both mechanisms by decreasing quitting and increasing breakthroughs. We remark about the differences between Daylight and the Short SP when it comes to breakthroughs. Recall, from Result 1, that when subjects trail their opponents under Daylight, they become much more likely to quit. In fact, this behavior generalizes to ”second-movers”: if the first subject either makes a breakthrough or quits, the subject who is still in Stage A is also more likely to quit. This implies that there are feasible contests under Daylight where both subjects quit Stage A and no breakthroughs are made. The Short SP reduces the chances of this happening, since if a subject quits, this information is not immediately revealed to her opponent, who could still make a breakthrough in the interim.

Summary of Behavior The tension that exists between competition and infeasibility risk shapes the way that subjects respond to information as follows:

- In the absence of risk, breakthroughs always happen regardless of the information mechanism used. The Daylight mechanism exposes subjects to competition and therefore leads to more quitting when subjects see they are trailing their opponents.

- The Darkness mechanism, which shields subjects from competition, exposes them to high risk. This leads to excessive quitting in high risk environments under that mechanism. Being able to observe one another under Daylight or Short SP weakens or completely removes this risk effect. A Silent Period whose duration is too long resembles the Darkness mechanism, and therefore leads to similar outcomes.

In Section 5.2, we use our understanding of individual behavior to examine how a designer should select an information mechanism as a function of risk in the environment in order to maximize her payoffs. Further, because the different information mechanisms differ in their effects on whether subjects quit or continue to play (possibly for no gain), their welfare properties are different, and we explore that in Section 5.3.

5.1.3 Stage B Outcomes

We close this section by commenting on subjects' behavior in Stage B. As noted earlier, the interesting interactions in our experiment happen in Stage A, but the existence of Stage B allows us to differentiate Darkness and Daylight (since both mechanisms are identical when the contest consists of a single stage – see the discussion at the bottom of page 7) and to examine how subjects respond to learning that their competitor has completed the first stage. Specifically, whether they are encouraged by having the infeasibility risk resolved, or discouraged by trailing behind their competitors.

Recall that Stage B is always feasible, and in that sense it resembles Stage A when there is no risk, and we indeed find that the results for these two cases are similar: information has no effect on whether the first agent completes the stage or not. But unlike Stage A, subjects cannot quit Stage B if they trail because the contest is immediately over once that stage is completed. Table 11 in Appendix A is analogous to Table 4, and shows no effect of any of our experimental conditions on Stage B outcomes. This implies that completing the entire contest and obtaining the innovation heavily relies on obtaining a breakthrough and getting through to Stage B in the first place.

5.2 Designer's Objective

As noted in Section 3 and following the theoretical literature on dynamic innovation contests (e.g. Benkert and Letina (2020); Bimpikis et al. (2019)), we assume that the designer is interested in obtaining the innovation and in doing so in the shortest amount of time. As we show below, *conditional on obtaining the innovation*, the duration of the contest is similar under the different information mechanisms. Thus, the main differentiator becomes which mechanisms increase the chances that the innovation is obtained when it is indeed feasible.

	Dependent variable:		
	Contest Completion	Contest Duration (All)	Contest Duration (Complete)
	Logistic	*OLS*	*OLS*
	(1)	(2)	
Daylight	-0.133	35.260	6.301
	(0.094)	(51.191)	(45.939)
Short SP	-0.057	-42.353	-50.033
	(0.127)	(52.662)	(53.192)
Long SP	0.112	-58.520	-27.248
	(0.111)	(49.583)	(55.127)
Mild Risk	-0.111	-35.454	-37.647
	(0.093)	(44.111)	(48.269)
High Risk	-0.280^{***}	-91.681^{**}	-82.119
	(0.099)	(43.712)	(57.474)
Mild Risk × Daylight	0.0389	58.956	17.634
	(0.0970)	(78.988)	(69.649)
Mild Risk × Short SP	-0.093	108.212	133.827
	(0.186)	(69.449)	(92.390)
Mild Risk × Long SP	-0.245	151.920^{*}	61.425
	(0.278)	(82.392)	(83.604)
High Risk × Daylight	0.128^{**}	48.883	11.102
	(0.054)	(69.402)	(79.697)
High Risk × Short SP	0.184^{***}	62.325	101.938
	(0.040)	(81.295)	(91.716)
High Risk × Long SP	-0.225	229.326^{*}	119.300
	(0.281)	(120.484)	(98.584)
Baseline (Darkness + No Risk)	0.912^{***}	313.423	277.252^{***}
	(0.042)	(36.029)	(31.827)
Observations	423	423	342

Notes: The omitted categories are Darkness and No Risk. Duration for Columns (2) and (3) is in seconds. Column (2) represents treatment effects on the time for the contest to conclude (with or without having found the innovation), while Column (3) shows treatment effects on the time it takes for successful innovation. Coefficients represent average marginal effects and robust standard errors in parentheses are clustered at the group level, with *p$<$0.1; **p$<$0.05; ***p$<$0.01.

Table 5: Treatment Effects on Contest Completion and Duration In the Set of Feasible Contests

Similar to the previous section, we look for these effects in the set of feasible contests, since if a contest is infeasible the innovation is never obtained. Table 5 shows three specifications. Specification (1) shows treatment effects on the likelihood of obtaining the innovation, and provides one of the most important results of the paper. Specification (2) shows treatment effects on the duration of the contest, regardless of whether the innovation was obtained or not, and Specification (3) shows treatment effects on the time it takes to reach innovation for the set of contests where the innovation was successfully obtained.

Our main findings are as follows:

- In the presence of no or mild risk, the designer is equally well off under any of the information mechanisms.

- When the risk level is high, visibility (i.e. subjects being able to observe one another through Daylight or Silent Period) significantly improves the payoff of the designer compared to the Darkness mechanism.

- However, in these high risk environments, payoffs are sensitive to how the duration of the silent period is set. Short silent periods yield the highest payoffs, while longer ones deliver the lowest payoffs.

The above findings suggest that, when faced with a high-risk innovation task , the designer should *not* use the Darkness mechanism. If the task bears some resemblance to other tasks that have been completed in the past, then data from these tasks may be used to guide the selection of the silent period duration. But if the task is completely novel or the designer has no access to such historical data, then setting the duration of the silent period incorrectly might lead to low payoffs. In that case, the Daylight mechanism might be a more robust choice: it provides some of the advantages of visibility (increased completion rates), while avoiding the potential pitfalls that come with selecting appropriate silent period durations.

The above discussion provides a clear outline for how the designer should select her information mechanisms as a function of risk. For completeness, we combine the designer's goals (maximizing the chances of reaching innovation while reducing the time it takes to do so) into a single objective and study which mechanisms maximize her *discounted* payoff. Formally, let the designer have discount rate r and denote by τ the time at which the innovation is obtained. If the designer values the innovation at U, then her payoff can be written as

$$\text{Payoff} = \begin{cases} e^{-r\tau}U & \text{if innovation is obtained at time } \tau \\ 0 & \text{if innovation is not obtained} \end{cases}$$

(where an innovation that is not obtained can also be thought of as an innovation that arrives at time $\tau = \infty$).

We use our data to calibrate the designer's payoffs under different parameter combinations. The plots in Figure 4 display the fraction of utility U that the designer obtains under the different mechanisms, as a function of the discount factor r. We discuss each of the risk conditions in detail below.

No Risk: In the absence of risk, the innovation is equally likely to be obtained under all mechanisms (Specification (1) in Table 5 shows that the coefficient of Daylight and the two Silent Period durations are not significantly different from Darkness) and there is no difference in the time it takes to do so (Specification (3) in Table 5 again shows no significant differences among the coefficients), hence the choice of mechanism in this environment does not matter. This is also seen in the top left plot in Figure 4, where all information mechanisms produce similar discounted payoffs for the designer.

Mild Risk: Similar to the No Risk case, there are again no significant differences between the information mechanisms in whether the innovation is reached or not (coefficients and interactions with Mild Risk in Specifications (1) and (3) in Table 5 are insignificant). As the top right plot in Figure 4

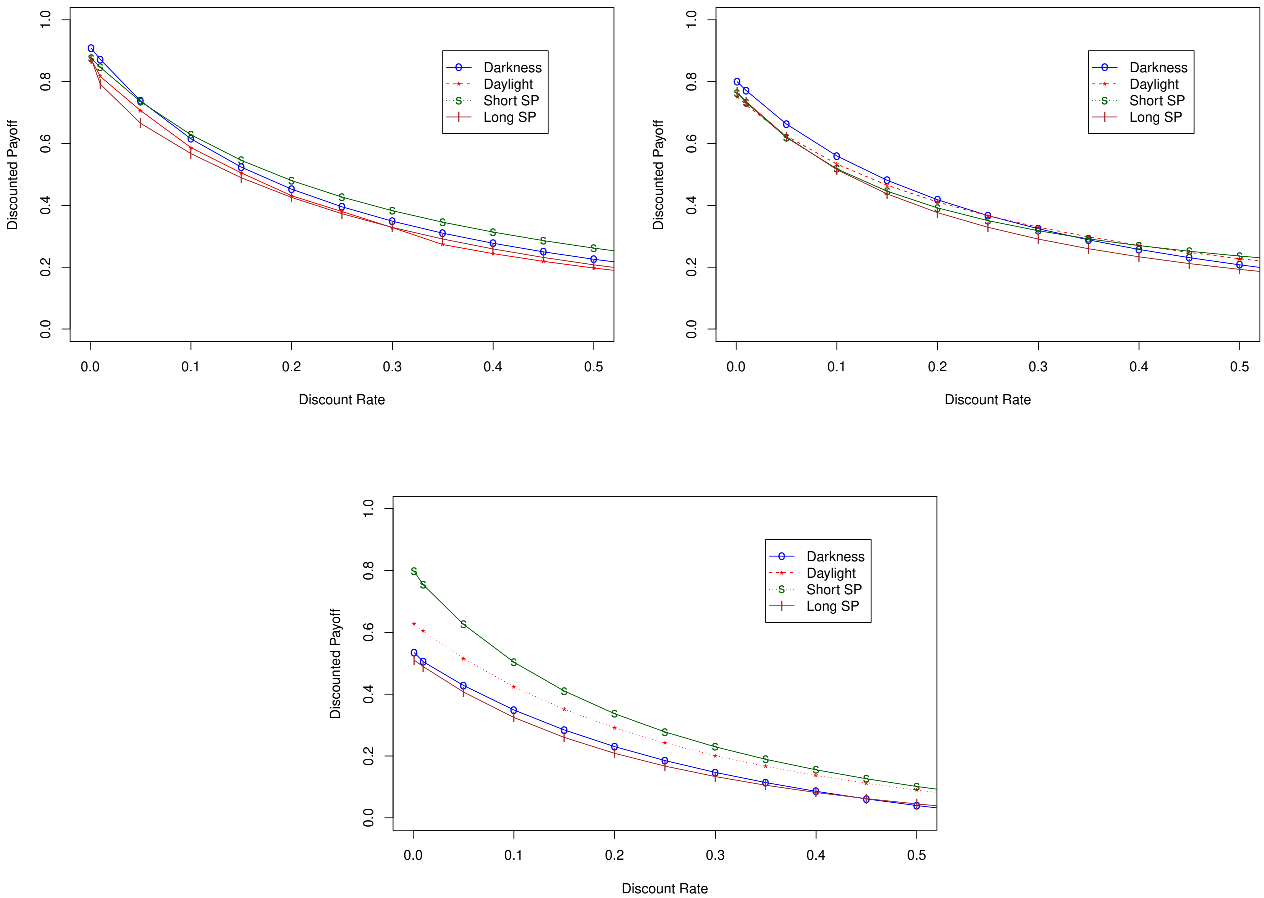

Figure 4: Designer's discounted payoffs for the three risk conditions under the different information mechanisms (as a function of the discount factor r). Top left is No Risk, top right is Mild Risk, and High Risk is at the bottom.

shows, payoffs are below their No-Risk counterparts when the discount factor is low and that gap diminishes as the designer discounts the future more.

High Risk: Finally, the innovation is not reached significantly more often under Darkness when the level of risk is high, as can be seen in Specification (1) in Table 5, where the coefficient of High Risk is both substantial and significant (0.27, $p < 0.01$). This gives the designer a utility of zero significantly more often under this mechanism. By contrast, the Daylight and Short SP mechanisms boost the chances of reaching the innovation by an average of 12.8 p.p. ($p < 0.05$) and 18.4 p.p. ($p < 0.01$), respectively. This can be seen in the bottom plot of Figure 3. Notably, the Long SP is similar to Darkness in that it provides the designer with the lowest discounted payoff.

The choice of information mechanism is therefore crucial when the level of risk in the environment is high. When that is the case, using the Darkness mechanism is decidedly worse for the designer. Selecting a longer duration for the Silent Period mechanism yields similarly poor results. The designer can obtain higher payoffs if she can pin down the optimal silent period duration, but Daylight provides a safer choice when that option is not available.

We conclude this section by noting that when the task is actually infeasible, there are again no differences between the information mechanisms when it comes to the duration of the contest, i.e., the different information mechanisms are all similar in allowing the designer to "fail fast and move on". As such, the choice of mechanism should follow the guidelines based on Table 5 and Figure 4.

5.3 Welfare and Earnings

As noted in the introduction, contests are a design where one or a few participants receive prizes and the other participants receive nothing. In that sense, contests resemble auctions in that what is optimal for the auctioneer/designer is not necessarily optimal for the bidders/players and vice-versa. While the majority of the literature on contest design focuses on effort maximization, few designs analyze the welfare effects and/or participant earnings (see Vojnović (2016) for a summary of these results in static contests.)

This tension is observed in our experiment. While the designer enjoys players exerting effort as long as the contest is active, this can be socially wasteful. For example, given that the second stage is always feasible, it might be enough that once a breakthrough is achieved, one subject continues on to finish the second stage while the other drops out, which, as we have seen in Section 5.1, is exactly what happens under the Daylight mechanism.

Table 6 examines the effects of the different treatments on the *total time spent* in the contest, i.e. the sum of the times that both subjects spend playing until the contest concludes. It can be seen that in risky environments, long silent periods induce higher total times (coefficient of Mild Risk $\times$ Long SP $= 222.774$, $p = 0.032$, and coefficient of High Risk $\times$ Long SP $= 234.087$, $p = 0.037$). Thus, not only does the Long SP deliver low payoffs to the designer, it is also socially wasteful as subjects spend the most amount of time under this treatment. In that sense, Daylight emerges again as the mechanism that achieves a good balance in high risk environments when the designer is unsure about the appropriate duration to set for the Silent Period mechanism. It increases the chances that the innovation is reached (See Specification (1) in Table 5) and does so without leading to unnecessary effort expenditure.

We next examine the effects of information mechanisms on the contestants' earnings, defined as how much money a player earns divided by the amount of time it takes to earn that amount. Recall that there are only three possible amounts to earn in each contest ($0, $1, or $5), and so the time spent to earn that money becomes a major differentiating factor between the different mechanisms.

Table 7 presents four OLS specifications, with earnings (dollar per minute) as the dependent variable, and dummies for the Daylight and the Silent Period mechanisms as the independent variables. The first column shows treatment effects on earnings for the average subject across all contests, while the rest of the columns break down earnings over the subsets of complete and incomplete contests. There are no differences in earnings for the average subject, as Specification (1) in Table 7 shows no significant coefficients. Further, Specification (2) also shows no effects of the information mechanisms

	Dependent variable:
	Total Time Spent
Daylight	33.613
	(74.893)
Short SP	−68.610
	(94.209)
Long SP	−58.856
	(88.207)
Mild Risk	−107.798
	(74.585)
High Risk	−31.423
	(75.536)
Mild Risk × Daylight	38.588
	(106.370)
Mild Risk × Long SP	222.774**
	(104.060)
Mild Risk × Short SP	168.639
	(130.911)
High Risk × Daylight	68.639
	(105.464)
High Risk × Long SP	234.087**
	(112.332)
High Risk × Short SP	135.000
	(126.139)
Baseline (Darkness + No Risk)	523.556***
	(53.412)
Observations	528

Notes: The omitted categories are Darkness and No Risk. Total Time Spent is in seconds and coefficients represent average marginal effects and robust standard errors in parentheses are clustered at the group level, with *p<0.1; **p<0.05; ***p<0.01.

Table 6: Treatment Effects on Total Time Spent in the Contest

on the earnings of those who win the contest. By contrast, the effect of information mechanisms on the earnings of those who do not win depend on how the dynamics of the contest unfold, as we discuss next.

Complete Contests We first consider the subset of complete contests. As noted, Specification (2) in Table 7 shows no significant impact of the information mechanism on winners' earnings ($p > 0.1$ for the coefficients of the Daylight and Silent Period dummies).

For those players who lost the contest, Specification (3) in Table 7 shows that they are better off under Daylight ($0.55, p < 0.01$). This is because, from Table 4, subjects who trail are much more likely to quit under this mechanism, and therefore they exercise their outside option and earn $1, instead of the $0 they are more likely to earn under the other mechanisms. In Darkness and Silent Period, subjects continue to play, unaware that they are trailing their opponent, and cannot exercise their

	Dependent variable: Earning Rate (dollar per minute)			
	All Contests	Complete Contests		Incomplete
		Winner	Loser	Contests
	(1)	(2)	(3)	(4)
Daylight	0.109	-0.112	0.550^{***}	-0.038^{**}
	(0.106)	(0.229)	(0.100)	(0.016)
Short SP	-0.018	-0.027	0.010	-0.036
	(0.129)	(0.309)	(0.122)	(0.020)
Long SP	-0.094	-0.157	0.0158	-0.054^{***}
	(0.127)	(0.309)	(0.121)	(0.019)
Baseline (Darkness)	0.872^{***}	2.243^{***}	0.174^{**}	0.217^{***}
	(0.075)	(0.161)	(0.011)	(0.011)
No. of observations	1056	342	342	372

Notes: The first column is the average wage across all subjects and contests. Columns (2) and (3) represent average earnings for winners and losers in the subset of completed contests, whereas the last column provides average earnings for contests that neither player was able to complete. The omitted category is Darkness. *p<0.1; **p<0.05; ***p<0.01.

Table 7: Effect of different information mechanisms on earnings: OLS specifications

outside option once their opponent completes the contest.

Incomplete Contests While subjects who trail their opponents are able to quit earlier in the Daylight mechanism, decreasing their time spent, they may also choose to stay if they can observe that their opponent has not made any progress. Indeed, learning about a lack of opponent progress may actually decrease earnings if Stage A is infeasible or players cannot make a breakthrough. The reason for this decline is that players spend a longer amount of time in these games (each player waiting for the other to either quit or finish the stage in case of Daylight, or waiting for the announcement that no breakthrough has been made under the Long SP) but still end up with the same \$1 prize or with \$0 (in case they do not complete the consolation game). This is reflected in Specification (4) in Table 7, which shows that in contests with no winner, earnings decrease under Daylight ($-0.038, p = 0.021$) and Long SP (-0.054, p¡0.01).

6 Discussion

This paper offers a novel experimental design to study the role of risk and information in environments characterized by strategic experimentation and dynamic competition. Infeasibility risk reduces the chances of breakthroughs by making players quit contests that are in fact feasible. We find that as the level of infeasibility risk increases, players are less likely to quit if they can observe each other or if they know that they will observe each other's progress at an appropriately-chosen future time. We discuss the implications of these finding to the choice of information mechanism and how it affects the designer's objective and participants' earnings. We also discuss implications to theory, which suggests that shutting down this information and visibility channel between players should decrease, not increase, their likelihood of quitting when the level of risk is high.

Limitations In designing our experiment we chose to focus on the factors that we consider the most salient (infeasibility risk and information provision). Other factors that impact the outcome are also important. For example:

- *Prize value*: The way participants respond to changes in risk and information is likely to also be modulated by prize value. This value was fixed at $5 in our main experimental sessions. In our experiment, subjects were more likely to persevere through risk and make the breakthrough discovery if they can observe each other, but then they would quit immediately once they realize they are trailing their opponents. As the prize value increases, it is possible that players continue to play even as they trail behind. This can be desirable if, for example, the contest has multiple (risky)stages that can benefit from having multiple competitors.

- *Information sharing*: Some contests require that breakthrough discoveries are shared. For example, The Netflix Prize required the winners of the first stage to publish a paper explaining their methodology.[10] In theory, this can be akin both players advancing to Stage B if either of them makes a breakthrough. This introduces free riding elements, although one can also imagine the player that made the breakthrough to still be advantaged through developing the know-how instead of simply receiving a solution. Clearly, this will have an effect on how players behave that can be different from what we observe in our experiment.

- *Similarity between stages*: Our experiment localized the risk inherent in the contest to the early stage, which reasonably captures the risk inherent in these environments when everyone is just starting out. There is no residual risk once that stage is complete. This means that the remainder of the contest is "easier", in the sense that there is a single problem with no infeasibility risk. Changing this by having the remainder of the contest consist of several stages can change the effect of information as subjects now are less discouraged about their opponent progressing, since they have ample time to catch up.

- *Skill asymmetry*: In our experiment, subjects were informed they are playing against opponents of similar skills. Observing an opponent making a breakthrough can signal a skill gap between players, and subjects may respond to this by quitting. Analyzing this setup when participants learn about each others' skills as well as the underlying environment is an open problem.

- *Social comparison and peer effects*: One unexpected finding is that, under high risk, our subjects are *less* likely to quit under Daylight compared to Darkness, which is the opposite of what

[10] An extension of this is to study the dynamics of team formation in contests with more than two players, which remains a problem that is not well-studied either theoretically or experimentally. For example, after intermediate prizes were announced in the NetFlix Prize competition, the winning teams were required to write reports that describe how their solution was obtained. Teams were also given the chance to merge with other teams for the rest of the contest, and in fact the team that ultimately won the competition was a combination of two previous teams. The recent experimental work of Boudreau et al. (2017) examines scientific collaborations and teams formation in a non-contest setting.

theory predicts. This result might indicate that subjects use social information provided by Daylight not only to adjust their beliefs about the infeasibility risk, but also to infer their relative standing and the socially desirable actions. It would be interesting to design new experiments to separate these different channels and to enrich theory by incorporating social comparison and peer effects into these contest models.

Appendix

A Additional Tables

Subject Demographics and Balance Checks Our subjects come from a large public university. Table 8 reports the summary statistics for our pre-treatment characteristics, broken down into the three information conditions. Note that for balancedness, we need only check the information mechanisms, as the risk level is randomized within-subject.

The first three columns in Table 8 report average values as well as standard deviations. We perform F and χ^2 tests on joint orthogonality across the treatments and report the associated p-values in the last column. The statistics in Table 8 show that the randomization yields balanced experimental groups along most characteristics. One exception is that the Daylight treatment has a higher proportion of undergraduate students compared to the other two conditions.

Characteristic	Darkness	Daylight	Silent Period	p-value
Age	21.0	22.0	20.9	0.217
	(3.55)	(6.92)	(5.59)	
Female	0.638	0.629	0.75	0.087
	(0.483)	(0.485)	(0.435)	
Undergrad	0.690	0.819	0.692	0.038
	(0.465)	(0.387)	(0.464)	
White	0.405	0.440	0.45	0.770
	(0.493)	(0.498)	(0.500)	
African American	0.06	0.06	0.075	0.871
	(0.239)	(0.239)	(0.264)	
Asian	0.388	0.405	0.342	0.583
	(0.489)	(0.493)	(0.476)	
Hispanic	0.077	0.026	0.05	0.202
	(0.269)	(0.159)	(0.219)	
Multiracial	0.043	0.051	0.058	0.869
	(0.204)	(0.222)	(0.235)	
Skill	2.7	2.73	2.52	0.178
	(0.99)	(0.952)	(0.857)	

Notes: The first three columns report average values in each experimental condition, whereas the last column reports the p-values testing the joint orthogonality across treatments. Standard deviations are provided in parentheses.

Table 8: Balancedness Checks: Subject Demographics and Skills

Effect of Risk on Quitting As noted in Section 5.1.2, we test whether subjects quit because of risk or because of reduced expected pay. Our main experiment used a prize value of \$5, while our pilot experiments had a prize value of \$4 (we increased the prize value to attract more participation for our main study). We pool the data from both the pilot and main experiments to test whether quitting is different between the following two settings:

- Setting A: this is our setting in the paper where the prize value is \$5. We used the subset of contests that were feasible but subjects were told that there is a 20% chance of infeasibility, leading to an expected winning payoff of \$4.

- Setting B: this is the certainty-equivalent setting, where the prize is \$4 and subjects know the problem is feasible, i.e. there is no infeasibility risk.

We let Risk be a dummy variable that is equal to 1 for Setting A and 0 for Setting B. We find a significant difference in quitting between the two settings: subjects are 21.6% more likely to quit in Setting A compared to the certainty-equivalent Setting B (logistic regression, $p < 0.001$), establishing that the reason subjects quit in our main experiment is the risk attached to Setting A.

	Quit
Risk	0.216^{***}
	(0.0513)
Observations	312
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 9: Effect of Risk on Quitting

Results With a Single Silent Period Dummy Table 10 provides the same results as in Table 4 but using only a single dummy variable that comprises all Silent Period treatments. The same effects identified in the main text still exist, albeit with weaker and noisier estimates, owing to the fact that the different durations induce different behaviors (as noted in Table 4, under High Risk, Short SP decreases quitting and increases breakthroughs, whereas both of these outcomes are negatively affected when using the Long SP mechanism.)

	Dependent variable:		
	Quit	Quit	Breakthrough
	(1)	(2)	
Daylight	0.204**	−0.001	0.006
	(0.082)	(0.078)	(0.078)
Silent Period	0.078	0.021	−0.032
	(0.088)	(0.074)	(0.0816)
Mild Risk	0.151*	0.139*	−0.089
	(0.084)	(0.071)	(0.0909)
High Risk	0.295***	0.269***	−0.265**
	(0.08)	(0.074)	(0.112)
Trail		−0.089	
		(0.061)	
Trail × Daylight		0.440***	
		(0.085)	
Trail × Silent Period		0.138	
		(0.103)	
Mild Risk × Daylight	−0.099	−0.077	−0.005
	(0.070)	(0.073)	(0.104)
Mild Risk × Silent Period	−0.021	−0.016	0.040
	(0.090)	(0.085)	(0.068)
High Risk × Daylight	−0.143**	−0.113*	0.054
	(0.051)	(0.058)	(0.061)
High Risk × Silent Period	−0.170***	−0.161***	0.084**
	(0.0579)	(0.050)	(0.041)
Observations	846	846	432

Notes: The omitted categories are Darkness, No Risk and No Trail, respectively. Coefficients represent average marginal effects and robust standard errors in parentheses are clustered at the group level, with *p<0.1; **p<0.05; ***p<0.01.

Table 10: Quitting and Breakthroughs – Single Silent Period Treatment Dummy

Stage B Outcomes Following the discussion in Section 5.1.3, because Stage B resembles the first stage when there is no risk, we see no effect of any of the experimental conditions on quitting or completing the stage. This can be seen in Table 11, where none of the treatments show significant effects.

	Dependent variable:	
	Quit (Subject Level)	Complete (Stage Level)
	(1)	(2)
Daylight	0.104	-0.135
	(0.076)	(0.090)
Short SP	0.015	0.070
	(0.082)	(0.086)
Long SP	-0.069	0.064
	(0.075)	(0.095)
High Risk	-0.007	0.055
	(0.091)	(0.112)
Mild Risk	0.002	-0.028
	(0.079)	(0.096)
Mild Risk × Daylight	0.017	0.029
	(0.100)	(0.094)
Mild Risk × Short SP	0.179	-0.387
	(0.232)	(0.332)
Mild Risk × Long SP	0.341	-0.251
	(0.294)	(0.294)
High Risk × Daylight	-0.007	-0.026
	(0.104)	(0.113)
High Risk × Short SP	0.010	-0.171
	(0.146)	(0.376)
High Risk × Long SP	0.364	-0.430
	(0.337)	(0.378)
Observations	520	385

Notes: The omitted categories are Darkness, No Risk and No Trail, respectively. Coefficients represent average marginal effects and robust standard errors in parentheses are clustered at the group level, with *p<0.1; **p<0.05; ***p<0.01.

Table 11: Regression Results for Stage B

B Experiment Instructions

The instructions used in the the sessions are reprinted on the next page.

Experiment Instructions

IMPORTANT: PLEASE TURN OFF OR MUTE YOUR PHONES.

Welcome! This is an experiment in the economics of decision making. In this experiment, you will be asked to solve knapsack games. The amount of money you earn will depend on the decisions you make and on the decisions other people make. Please do not communicate with others during the experiment. If you have questions at any point during the experiment, raise your hand and the experimenter will help you.

This experiment includes two phases, the training phase and the contest phase.

1. In the **training phase**, you will play six different knapsack games individually.
2. In the **contest phase**: you will play six different knapsack games in three contests. In each contest, you will compete against another player in solving two knapsack games. The player who finishes first wins.

You will also fill out a short survey after the experiment is over.

You will get paid after completing both phases and filling out the survey.

Knapsack Game Description

Imagine you are going on a trip and have only one suitcase. You cannot fit all the things you would like to take with you, so you try to fit the most important things. This is the knapsack game. In the game, you have a knapsack with limited weight capacity. There are a number of items available and each item has a weight and a value. You cannot fit all the items in the knapsack because their total weight is more than the maximum weight capacity of the knapsack. You will be given a target value and your goal is to find a subset of items that will fit into the knapsack whose total value is at least equal to the target value.

Payment

Your total payment is the sum of what you earn in the training phase and in the contest phase.

In the **training phase**, you will be paid $1 for each game in which you can successfully fit items adding up to the target value (or above) into the knapsack. If you are unable to reach the target value, you can skip to the next game.

In the **contest phase**, there is a final prize of $5.0 for each contest. The player who is able to finish the games first will earn the prize. The other player will get nothing. Finishing a game in the contest phase means being able to fit items in the knapsack whose total value is at least equal to the target value.

Training Phase

The items, along with their dollar values and weights, are displayed in a pool on top of the page. Clicking on an item adds it to the knapsack (if it fits). Clicking on an item in the knapsack removes it and returns it back to the pool of items not in the knapsack. You can add and remove items as many times as you like before you submit your solution.

If you succeed in fitting the target value (or higher) into the knapsack, your answer will be submitted automatically and you will go on to the next game. You can also click the "Submit" button at any time to go on to the next game even if you were unable to fit the target value into the knapsack.

Training games are always **feasible**, that is, the target value can always be reached.

You will have a total of 30 minutes to finish the training phase. After this, this phase will automatically stop and we will move to the contest phase.

When you are finished with the training phase, you will enter the **waiting page** until everyone is finished before moving to the contest phase.

Feel free to refer to the experimental instructions at any time during the experiment.

We encourage you to earn as much cash as you can. Are there any questions?

Now we will start the training phase.

Contest Phase

Based on your performance in the training phase, you will be placed in a **group of four players with similar skill levels**. Before you enter a contest, you will go to the **waiting page** until everyone in your group is in the waiting page, ready to start the next contest.

The contest phase consists of **three contests**. In each contest, you will compete with a different player in your group. Each contest consists of two sequential knapsack games. Each game is finished if you can fit the target value (or higher) into the knapsack. You cannot play Game 2 unless you finish Game 1. The player who finishes Game 2 first wins a prize of $5.0.

Uncertainty:

Recall that a **feasible** game is one where the target value can always be reached. While all training games you have played so far are feasible, Game 1 in each contest can either be feasible or not. An **infeasible** game is one where the target value cannot be reached, no matter how hard you try. More precisely, in an infeasible game, there does not exist a subset of items whose values add up to the target value.

Game 2 is always feasible. This means that if you have finished Game 1, you can go on to try and finish Game 2. Naturally, the contest cannot be completed (and no one can win a prize) if Game 1 is infeasible. You will not be told whether Game 1 is feasible or not. Instead, before you start the contest, you will be told that the chance Game 1 is feasible is 100% (definitely feasible), or 80%, or 60%.

Before each contest, you will see one of the following sentences displayed on the screen:

[For the next contest, the chance that Game 1 is feasible is 100%.]

[For the next contest, the chance that Game 1 is feasible is 80%.]

[For the next contest, the chance that Game 1 is feasible is 60%.]

The Quit Option: At any point in the contest, unless your opponent has already won, you will have an option to quit by pressing the "**Quit**" button. If you press the button, you will be presented with a feasible knapsack game that you can try to solve (with no competition). If you finish it, you will get paid $1.0, similar to the training phase.

Note that if the other player wins, the contest will be over and you will not have the chance to quit and play the feasible game.

At any time during a contest, you will be immediately notified if your opponent finishes Game 1 or quits. This means that, if you do not receive a notification, your opponent is still trying to solve Game 1.

Are there any questions?

Review Questions: To make sure that everyone understands the instructions, you will be asked a number of review questions. When everyone is finished with these questions, we will go through the answers together.

Feel free to refer to the experimental instructions before you answer any question. Each correct answer is worth 20 cents, and will be added to your total earnings.

3

1. How many people are there in your group?
    a. 3;
    b. 4;
    c. 12.

2. In the contest phase, are you going to play against the same opponent more than once?
    a. Yes;
    b. No.

3. True or false: An infeasible game can be solved if I try hard enough.
    a. True;
    b. False.

4. True or false: In a contest, you are told that "the chance that Game 1 is feasible is 100%." This means that the target value can be reached.
    a. True;
    b. False.

5. In a contest, you are told that "the chance that Game 1 is feasible is 60%." Which of the following statements is correct?
    a. The likelihood that Game 1 is infeasible is 40%.
    b. The likelihood that the target value can be reached is 40%.

C Knapsack Games

In this appendix, we present the knapsack games used in our experiment and their solutions.

C.1 Training Games

The training games are designed by the authors and progress with increasing difficulty level. These games are designed and sequenced so that subjects can get used to both the interface and practice solving these games.

Game 1:
Values = [3, 6, 9, 12, 15, 18, 21]
Weights = [1, 4, 7, 10, 13, 16, 19]
Capacity = [43]
Solution = [0, 1, 0, 1, 1, 1, 0]

Game 2:
Values = [15, 100, 90, 60, 40, 15, 10, 1]
Weights = [2, 20, 20, 30, 40, 30, 60, 10]
Capacity = [102]
Solution = [1, 1, 1, 1, 0, 1, 0, 0]

Game 3:
Values = [70, 20, 39, 37, 7, 5, 10]
weights = [31, 10, 20, 19, 4, 3, 6]
Capacity = [50]
Solution = [1, 0, 0, 1, 0, 0, 0]

Game 4:
Values = [350, 400, 450, 20, 70, 8, 5, 5]
weights = [25, 35, 45, 5, 25, 3, 2, 2]
Capacity = [104]
Solution = [1, 0, 1, 1, 1, 0, 1, 1]

Game 5:
Values = [37, 72, 106, 32, 45, 71, 23, 44, 85, 62]
Weights = [50, 820, 700, 46, 220, 530, 107, 180, 435, 360]
Capacity = [1500]
Solution = [1, 0, 0, 1, 1, 0, 1, 1, 1, 1]

Game 6:
Values = [2, 3, 4, 5, 6, 9, 8, 7, 6, 5, 8, 9]
Weights = [3, 4, 6, 3, 5, 13, 6, 9, 2, 4, 7, 7]
Capacity = [14]
Solution = [0, 0, 0, 1, 1, 0, 0, 0, 1, 1, 0, 0]

C.2 Contest Games

The contest games are from Meloso et al. (2009) (page 2 of the Supplementary Material).

Contest 1:
Game 1:
Values = [500, 350, 505, 505, 640, 435, 465, 50, 220, 170]
weights = [750, 406, 564, 595, 803, 489, 641, 177, 330, 252]
Capacity = [1900]
Solution = [0, 0, 1, 1, 0, 1, 0, 0, 0, 1]

Game 2:
Values = [31, 141, 46, 30, 74, 105, 119, 160, 59, 71]
weights = [21, 97, 32, 21, 52, 75, 86, 116, 43, 54]
Capacity = [265]
Solution = [0, 1, 0, 0, 1, 0, 0, 1, 0, 0]

Contest 2:
Game 1:
Values = [15, 14, 3, 3, 10, 9, 28, 28, 31, 25, 24, 1]
weights = [129, 144, 77, 77, 66, 60, 184, 184, 229, 184, 219, 72]
Capacity = [850]
Solution = [0, 0, 0, 0, 1, 0, 1, 1, 1, 1, 0, 0]
Game 2:
Values = [300, 350, 400, 450, 47, 20, 8, 70, 5, 5]
weights = [205, 252, 352, 447, 114, 50, 28, 251, 19, 20]
Capacity = [1044]
Solution = [1, 0, 1, 1, 0, 0, 0, 0, 1, 1]

Contest 3:
Game 1:
Values = [37, 72, 106, 32, 45, 71, 23, 44, 85, 62]
weights = [50, 820, 700, 46, 220, 530, 107, 180, 435, 360]

Capacity = [1500]
Solution = [1, 0, 0, 1, 1, 0, 1, 1, 1, 1]

Game 2:
Values = [201, 84, 113, 303, 227, 251, 129, 147, 86, 127, 144, 167]
weights = [192, 80, 106, 288, 212, 240, 121, 140, 82, 120, 137, 160]
Solution = [1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1]

References

Archak, Nikolay (2010), "Money, glory and cheap talk: Analyzing strategic behavior of contestants in simultaneous crowdsourcing contests on topcoder.com." In *Proceedings of the 19th international conference on World Wide Web*, Raleigh, North Carolina.

Arthur, W Brian and Wolfgang Polak (2006), "The evolution of technology within a simple computer model." *Complexity*, 11, 23–31.

Benkert, Jean-Michel and Igor Letina (2020), "Designing dynamic research contests." *American economic journal: Microeconomics*, 12, 270–89.

Bimpikis, Kostas, Shayan Ehsani, and Mohamed Mostagir (2019), "Designing dynamic contests." *Operations Research*, 67, 339–356.

Bolton, Patrick and Christopher Harris (1999), "Strategic experimentation." *Econometrica*, 67, 349–374.

Boosey, Luke, Philip Brookins, and Dmitry Ryvkin (2017), "Contests with group size uncertainty: Experimental evidence." *Games and Economic Behavior*, 105, 212–229.

Boosey, Luke, Philip Brookins, and Dmitry Ryvkin (2019), "Contests between groups of unknown size." *Games and Economic Behavior*, 113, 756–769.

Boosey, Luke, Philip Brookins, and Dmitry Ryvkin (2020), "Information disclosure in contests with endogenous entry: An experiment." *Management Science*.

Bothner, Matthew S, Jeong-han Kang, and Toby E Stuart (2007), "Competitive crowding and risk taking in a tournament: Evidence from nascar racing." *Administrative Science Quarterly*, 52, 208–247.

Boudreau, Kevin J, Tom Brady, Ina Ganguli, Patrick Gaule, Eva Guinan, Anthony Hollenberg, and Karim R Lakhani (2017), "A field experiment on search costs and the formation of scientific collaborations." *Review of Economics and Statistics*, 99, 565–576.

Boudreau, Kevin J. and Karim 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, URL `https://www.sciencedirect.com/science/article/pii/S0048733314001425`.

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

Cadsby, C Bram, Fei Song, Jim Engle-Warnick, and Tony Fang (2019), "Invoking social comparison to improve performance by ranking employees: The moderating effects of public ranking, rank pay, and individual risk attitude." *Journal of Economic Psychology*, 72, 64–79.

Casas-Arce, Pablo and F Asis Martinez-Jerez (2009), "Relative performance compensation, contests, and dynamic incentives." *Management Science*, 55, 1306–1320.

Cason, Timothy N., William A. Masters, and Roman M. Sheremeta (2020), "Winner-take-all and proportional-prize contests: Theory and experimental results." *Journal of Economic Behavior & Organization*, 175, 314–327, URL `https://www.sciencedirect.com/science/article/pii/S0167268118300234`.

Chaudoin, Stephen and Jonathan Woon (2018), "How hard to fight? cross-player effects and strategic sophistication in an asymmetric contest experiment." *The Journal of Politics*, 80, 585–600.

Chen, Yan, F. Maxwell Harper, Joseph Konstan, and Sherry Xin Li (2010), "Social comparisons and contributions to online communities: A field experiment on movielens." *The American Economic Review*, 100, 1358–1398, URL `http://www.jstor.org/stable/27871259`.

Chen, Yan and Tayfun Sönmez (2006), "School choice: An experimental study." *Journal of Economic Theory*, 127, 202–231.

Choi, Jay P (1991), "Dynamic r&d competition under" hazard rate" uncertainty." *The RAND Journal of Economics*, 596–610.

Dasgupta, Partha and Eric Maskin (1987), "The simple economics of research portfolios." *The Economic Journal*, 97, 581–595.

Dechenaux, Emmanuel, Dan Kovenock, and Roman M Sheremeta (2015), "A survey of experimental research on contests, all-pay auctions and tournaments." *Experimental Economics*, 18, 609–669.

Ederer, Florian and Ernst Fehr (2007), "Deception and incentives: How dishonesty undermines effort provision.", URL `https://ssrn.com/abstract=1071602`. University of Zurich Institute for Empirical Research in Economics Working Paper No. 341.

El Emam, Khaled, Luk Arbuckle, Gunes Koru, Benjamin Eze, Lisa Gaudette, Emilio Neri, Sean Rose, Jeremy Howard, and Jonathan Gluck (2012), "De-identification methods for open health data: The case of the heritage health prize claims dataset." *J Med Internet Res*, 14, e33, URL `http://www.jmir.org/2012/1/e33/`.

Eriksson, Tor, Anders Poulsen, and Marie Claire Villeval (2009a), "Feedback and incentives: Experimental evidence." *Labour Economics*, 16, 679 – 688, URL `http://www.sciencedirect.com/science/article/pii/S0927537109000980`. European Association of Labour Economists 20th annual conference University of Amsterdam, Amsterdam, The Netherlands 18–20 September 2008.

Eriksson, Tor, Anders Poulsen, and Marie Claire Villeval (2009b), "Feedback and incentives: Experimental evidence." *Labour Economics*, 16, 679–688.

Fershtman, Chaim and Uri Gneezy (2011), "The tradeoff between performance and quitting in high power tournaments." *Journal of the European Economic Association*, 9, 318–336.

Franco, Juan Pablo, Nitin Yadav, Peter Bossaerts, and Carsten Murawski (2018), "Where the really hard choices are: A general framework to quantify decision difficulty." *BioRxiv*, 405449.

Friedman, Daniel and Shyam Sunder (1994), *Experimental Methods: A Primer for Economists*. Cambridge University Press, New York, NY.

Gauriot, Romain and Lionel Page (2019), "Does success breed success? a quasi-experiment on strategic momentum in dynamic contests." *The Economic Journal*, 129, 3107–3136.

Genakos, Christos and Mario Pagliero (2012), "Interim rank, risk taking, and performance in dynamic tournaments." *Journal of Political Economy*, 120, 782–813.

Genakos, Christos, Mario Pagliero, and Eleni Garbi (2015), "When pressure sinks performance: Evidence from diving competitions." *Economics Letters*, 132, 5–8.

Gill, David and Victoria Prowse (2012), "A structural analysis of disappointment aversion in a real effort competition." *American Economic Review*, 102, 469–503.

Greiner, Ben (2015), "Subject pool recruitment procedures: organizing experiments with orsee." *Journal of the Economic Science Association*, 1, 114–125, URL `http://dx.doi.org/10.1007/s40881-015-0004-4`.

Halac, Marina, Navin Kartik, and Qingmin Liu (2017), "Contests for experimentation." *Journal of Political Economy*, 125, 1523–1569.

Harris, Christopher and John Vickers (1987), "Racing with uncertainty." *The Review of Economic Studies*, 54, 1–21.

Hoelzemann, Johannes and Nicolas Klein (2018), "Bandits in the lab."

Hopkins, Ed and Tatiana Kornienko (2004), "Running to keep in the same place: Consumer choice as a game of status." *The American Economic Review*, 94, 1085–1107.

Hörner, Johannes and Andrzej Skrzypacz (2016), "Learning, experimentation and information design."

Jeppesen, Lars Bo and Karim R. Lakhani (2010), "Marginality and problem solving effectiveness in broadcast search." *Organization Science*, 21, 1016–1033.

Kagel, John H. and Alvin E. Roth (2000), "The dynamics of reorganization in matching markets: A laboratory experiment motivated by a natural experiment." *Quarterly Journal of Economics*, 115, 201–235.

Keller, Godfrey, Sven Rady, and Martin Cripps (2005), "Strategic experimentation with exponential bandits." *Econometrica*, 73, 39–68.

Koren, Yehuda (2009), "The bellkor solution ot the netflix grand prize.", URL `https://www.asc.ohio-state.edu/statistics/dmsl/GrandPrize2009_BPC_BellKor.pdf`. Netflix grand prize documentation.

Kremer, Ilan, Yishay Mansour, and Motty Perry (2014), "Implementing the "wisdom of the crowd"." *Journal of Political Economy*, 122, 988–1012.

Mago, Shakun D, Roman M Sheremeta, and Andrew Yates (2013), "Best-of-three contest experiments: Strategic versus psychological momentum." *International Journal of Industrial Organization*, 31, 287–296.

Malueg, David A and Shunichi O Tsutsui (1997), "Dynamic r&d competition with learning." *The RAND Journal of Economics*, 751–772.

Meloso, Debrah, Jernej Copic, and Peter Bossaerts (2009), "Promoting intellectual discovery: patents versus markets." *Science*, 323, 1335–1339.

Mihm, Jürgen and Jochen Schlapp (2019), "Sourcing innovation: On feedback in contests." *Management science*, 65, 559–576.

Murawski, Carsten and Peter Bossaerts (2016), "How humans solve complex problems: The case of the knapsack problem." *Scientific reports*, 6, 1–10.

Nature (2013), "First proof that infinitely many prime numbers come in pairs." `https://www.nature.com/news/first-proof-that-infinitely-many-prime-numbers-come-in-pairs-1.12989`.

Papanastasiou, Yiangos, Kostas Bimpikis, and Nicos Savva (2018), "Crowdsourcing exploration." *Management Science*, 64, 1727–1746.

Roth, Alvin E. (2002), "The economist as engineer: Game theory, experimentation, and computation as tools for design economics." *Econometrica*, 70, 1341–1378, URL `http://dx.doi.org/10.1111/1468-0262.00335`.

Sheremeta, Roman M. (2018), "Experimental research on contests.", URL `http://dx.doi.org/10.2139/ssrn.3260263`. Available at SSRN: https://ssrn.com/abstract=3260263.

Song, Hummy, Anita L Tucker, Karen L Murrell, and David R Vinson (2017), "Closing the productivity gap: Improving worker productivity through public relative performance feedback and validation of best practices." *Management Science*, 64, 2628–2649.

Taylor, Curtis R. (1995), "Digging for golden carrots: an analysis of research tournaments." *AER*, 85, 872–890.

Terwiesch, Christian and Yi Xu (2008), "Innovation contests, open innovation, and multiagent problem solving." *Management science*, 54, 1529–1543.

Vojnović, Milan (2016), *Contest theory: Incentive mechanisms and ranking methods*. Cambridge University Press.

Wooten, Joel O. and Karl T. Ulrich (2017), "Idea generation and the role of feedback: Evidence from field experiments with innovation tournaments." *Production and Operations Management*, 26, 80–99, URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/poms.12613`.

Youn, Hyejin, Deborah Strumsky, Luis MA Bettencourt, and José Lobo (2015), "Invention as a combinatorial process: evidence from us patents." *Journal of the Royal Society interface*, 12, 20150272.

Zizzo, Daniel John (2002), "Racing with uncertainty: a patent race experiment." *International Journal of Industrial Organization*, 20, 877–902.