Your Success is My Motivation *

Prateik Dalmia† and Emel Filiz-Ozbay‡

April 8, 2020

Abstract

We study how an individual's effort choice is impacted by feedback on her own past performance and another individual's past performance. In an effort choice problem where effort is costly but increases the chance of receiving a prize, subjects who failed in the previous period increase their effort in the next period. More interestingly, failed subjects who observe that their partner succeeded exert higher effort in the next period than failed subjects who observe that their partner also failed — behavior consistent with behindness aversion. This effect is more pronounced for female subjects than male subjects, suggesting that failing women are more motivated by the success of others than failing men. Rather than letting subjects work in isolation, we find that the highest joint effort can be achieved by matching failed and successful subjects into pairs and providing feedback about the other's performance. Our results suggest that social comparisons in independently performed and paid tasks may mitigate moral hazard.

Key Words: effort choice, moral hazard, motivation, gender

JEL Classification: D81, D91, J16

*We are grateful to the UMD-BSOS Dean's Research Initiative and the UMD-Department of Economics for funding. We thank Erkut Ozbay for helpful comments and fruitful discussions and Erika Dömötör for helping conduct the experimental sessions. Corresponding Author: Emel Filiz-Ozbay.

†University of Maryland, Economics, 3114 Tydings Hall, College Park, MD 20742, pdalmia@umd.edu

‡University of Maryland, Economics, 3114 Tydings Hall, College Park, MD 20742, efozbay@umd.edu

I INTRODUCTION

In many economically relevant decisions, agents decide how much risk to take while observing other people facing similar decisions rather than acting in isolation. In the process of weighing the costs and benefits of risk taking and figuring out her best action, a decision maker may compare their earlier performance with respect to the performance of their peers. For example, a worker may be motivated to work harder when she sees a coworker succeed in his project. Our laboratory experiments investigate how feedback about an agent's self-performance and relative performance with respect to a partner impacts their investment in a risky task. Our design isolates these effects from other externalities between agents arising directly from the production technology or compensation scheme.

We find that subjects who failed in a task are more likely to increase their effort in the next task. However, we find that the magnitude by which they increase their effort increases if they observe a partner who succeeded in the previous task. This suggests that being behind a peer motivates subjects to exert higher levels of effort.

Our environment imitates the decision problem of an agent in a standard moral hazard situation. An agent exerts costly effort in order to increase her chance of a good outcome (success on a task). Her payment scheme is contingent on a good outcome, as typically the case in a moral hazard setting: for example, a salesman's pay is proportional to his sales, a student's grade is based on her exam score, a farmer's earnings is based on the farmer's harvest, and an assistant professor's tenure depends on her publications. The degree of proximity among workers may naturally allow them to compare their outcomes with each other — a salesmen observe others' sales, students often catch wind of each other's grades, a farmer notices a neighboring farm's harvest, and assistant professors pay attention to the publications of their peers.[1] Indeed, employers may actively provide relative performance information even when they do not tie the payments of employees to each other's performances (Anderson et al., 1983). We aim to understand how such feedback about an agent's relative performance affects her effort. This question is applicable both to labor settings where co-workers might be motivated or demotivated from observing each other's outcomes, and to more general problems of performance feedback in risk-taking environments beyond those of moral hazard. For example, investment in lottery tickets might be affected by observing a peer win the jackpot.

[1] For example, observability of the performance of co-workers is present in Falk and Ichino (2006) and Mas and Moretti (2009).

In order to measure the peer effect, our main treatment (Partner Treatment) assigns each subject a partner and requires her to choose an effort level; while in our control treatment (Individual Treatment) subjects perform their tasks in isolation. In both treatments, subjects know that they will learn their own outcome (success or failure) after their tasks. In the Partner Treatment, they are informed that they will learn their partner's outcome as well. Someone who is behindness averse or status seeking may exert higher effort when she knows that she will receive feedback about her partner's performance. However, even if subjects have such a behavioral motive, they may not anticipate it. We also ask whether subjects anticipate their inequity concerns, or if these behavioral motives only affect effort after having experienced inequity in the past. To study this question, we implement a two-period design where subjects repeat the task twice. This allows us (i) to compare the behavior in the first period of the Partner and Individual Treatments to study the anticipation of peer effects and (ii) to study whether experiencing inequity with respect to a partner in the earlier task affects effort the next time around. Subjects are randomly assigned a different partner in each period so that there is no room for reputation building.

Standard economic theory rules out the co-dependence of agent objectives as long as their tasks are independently performed and paid. On the other hand, observing a peer's outcome in a previous task may generate a reference point in an agent's objective. Someone who failed while her partner succeeded may feel behind. Such an emotion may motivate her to get ahead by exerting higher effort. Alternatively, someone who succeeded while her partner failed may feel guilty over this unequal outcome and exert lower effort to avoid such outcomes. Several papers in the last decade incorporate social preferences such as inequality aversion (Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999) in a moral hazard problem and formally analyze their influence on optimal incentive contract to extract greatest effort (Bartling and Von Siemens, 2010; Bartling, 2011; Demougin et al., 2006; Englmaier and Wambach, 2010; Itoh, 2004; Neilson and Stowe, 2010). Our hypotheses are built on the predictions of a model of an inequity averse agent.[2] Our experiments aim to answer the following questions: (i) Does the anticipation of feedback about a peer's outcome motivate an agent to invest more? (ii) Does an agent revise her effort based on feedback about her prior performance? (iii) Does an agent revise her effort based on feedback about her prior relative performance with respect

[2]Since we are only interested in the behavior of agents, we omit the principal's problem of constructing an optimal incentive scheme. However, our results allow us to understand the motivations of agents better and can lead to richer models characterizing the optimal contract choice in moral hazard problems.

to a partner?

Since we only pay for one period determined randomly ex-post, an expected utility maximizer with selfish motives should view the two periods independently and choose the same effort level in each period, with or without partner. However, even in the Individual Treatment, our subjects revise their efforts based on their own prior outcomes. Subjects who failed increase their effort and subjects who succeeded decrease their effort. Hence, this rules out standard expected utility maximization and suggests some adaptive learning may take place. Furthermore, we find that this effect is more pronounced for female subjects than male subjects. This result contributes to the question in the reinforcement effects literature about whether risk preferences in future lotteries are influenced by the instance of success or failure in past lotteries (Andrade and Iyer, 2009; Dillenberger and Rozen, 2015; Imas, 2016; Langer and Weber, 2008; Liu et al., 2010; Nielsen, 2019; Smith et al., 2009; Thaler and Johnson, 1990; Tserenjigmid, 2019).[3]

Additionally, we find that subjects do not anticipate peer effects as the distribution of effort is similar in the first periods of the Individual and Partner Treatments. However, feedback about an agent's self and relative performances affects behavior in the second period of decision making. Subjects revise their effort choices differently based on their relative performance with respect to a partner. The subjects who failed while their partners' succeeded increase their second period effort more than those who failed while their partner also failed. Furthermore, this behavior is more pronounced for female subjects than male subjects.[4] This is in line with evidence that women typically react more strongly to feedback than men (Berlin and Dargnies, 2016).[5] Our results suggest that matching employees who are likely to have unequal outcomes may motivate those who are behind to exert higher effort without discouraging those who are ahead.

Many laboratory experiments (Charness et al., 2013b; Dohmen and Falk, 2011; Dohmen et al., 2011; Falk and Ichino, 2006; Kuhnen and Tymula, 2012) and other empirical studies (Bandiera et al., 2010; Breza et al., 2018; Cohn et al., 2014; Guryan et al., 2009; Mas and

[3]This literature studies environments where both the outcomes in past and future lotteries hold an impact on the agent's final payoff, thus allowing for an income effect between periods. In our environment, income effects are minimized as only one period's pay is realized, allowing us to isolate the effect of feedback. We find that feedback about failure in past lotteries decreases an agent's risk taking in future lotteries.

[4]Our results contribute to a growing experimental literature on gender differences in preferences relevant to labor market environments, such as preferences for risk, social comparison, cooperation, competition and self-promotion (see Croson and Gneezy (2009) for a summary).

[5]Xia (2018) documents that women attribute negative feedback to ability while men attribute it to luck. That is in line with the higher increase in effort by women who failed than men who failed in our experiments.

Moretti, 2009) investigate the effect of performance comparisons between individuals on effort in settings without uncertainty and risk. However, there is little evidence of the role of social comparisons on effort in labor settings where risk plays a central feature.[6] Risk is present in many real-world applications and a necessary feature of a moral hazard problem. Furthermore, incorporating uncertainty allows us to directly address the aforementioned theoretical literature.

There is a body of literature investigating peer effects in risk taking behavior over lottery choices. Those studies typically ask subjects to choose between simple lotteries and analyze the influence of peers on risk attitudes (see, for example, Linde and Sonnemans (2012), Linde and Sonnemans (2015), Rohde and Rohde (2011) and Schwerter (2019)). Our effort choice problems can be viewed as lottery choice selections, as true of any moral hazard problem with a known distribution of probabilities over effort. In the studies mentioned, subjects compare net payoffs.[7] Our research builds on this literature by adding a labor context to the decision problem. In a labor context, it becomes relevant to disentangle the influence of the outcomes of peers and their effort. We study peer effects over outcomes rather than actions, and gross payoffs rather than net payoffs. Cooper and Rege (2011) and Lahno and Serra-Garcia (2015) also give subjects feedback on peers' choices, studying whether subjects imitate their peers' actions. Certainly, there are applications relevant for both types of references. It is an open question as to how social comparisons about the realization of uncertainty affect risk taking when peers' actions are unknown.[8]

Peer effects and other forms of competition have been studied in labor economics. The tournament literature has documented evidence that rank based payments provide strong incentives for agents to exert greater effort (Lazear and Rosen, 1981; Bull et al., 1987; Hannan

[6]Instead, the literature on risk and social comparisons tends to focus on distinguishing ex-post and ex-ante fairness concerns (for example, see Brock et al. (2013) and Saito (2013)).

[7]Furthermore, the references generated by the peer in Linde and Sonnemans (2012), Rohde and Rohde (2011) and Schwerter (2019) are exogenous and known before the lottery choice, while the reference in our setting is determined endogenously.

[8]The reaction to realizations of uncertainty is also studied in tournament like settings (Gächter et al., 2018; Gill and Prowse, 2012, 2014), or settings without peer effects (as in the reinforcement effects literature). While peer effects are documented to be positive in risky tournaments, the reaction to earlier realizations of uncertainty in single person decisions are mixed — some papers have found risk taking behavior to increase after a loss (Andrade and Iyer, 2009; Langer and Weber, 2008; Verbruggen et al., 2017; Schneider et al., 2016; Clark, 2010; Smith et al., 2009) and others have found risk reduction (Shiv et al., 2006; Liu et al., 2010; Nielsen, 2019).

et al., 2008).[9] We differ from this literature as we shut down competition for a monetary prize and pay subjects independently.

The remainder of the paper is organized as follows. Section II describes the decision problem and states the predictions of the standard theory and a behavioral model of inequity aversion for this setup. Section III states the Hypotheses based on the theoretical models. Sections IV and V present the experimental design and results, respectively. Section VI discusses our findings while Section VII concludes. The instructions of the experiments are provided in Appendix C.

II A PROBLEM OF COSTLY EFFORT IN A RISKY TASK

In this section, we outline a model of an inequity averse agent in a moral hazard problem, forming the basis of our hypotheses in the next section. An agent chooses a costly effort level in an individual task, $e \geq 0$, where the unit cost of effort is $c > 0$. The task has two possible outcomes: *success* and *failure*. The outcome is a random variable and for a given effort level, e, the probability of success is denoted by $p(e) \in [0, 1]$. Assume that $p(.)$ is differentiable, increasing and concave in e and $p(0) = 0$. When the effort choice of the agent is unobservable to the principal and there are only two possible outcomes, it is well-known that the optimal payment structure to incentivize a risk averse agent to exert a positive effort level specifies a fixed payment and a bonus contingent on success (assuming that success is more desirable outcome for the principal).[10]

Let (f, b) denote the fixed payment and bonus features of the payment scheme. An agent with utility function $u(.)$ — which is strictly increasing, concave and differentiable — will choose her effort level by solving:

$$\max_{e} p(e) \;\; u(f - ce + b) + (1 - p(e))\, u(f - ce) \tag{1}$$

We enrich the standard model by allowing the agent to experience disutility or utility from inequity with respect to a partner. Suppose two agents simultaneously choose their

[9] Feedback on relative performance may have detrimental effects to the firm. For example, it may encourage people to cheat (List et al., 2001), or to sabotage other's outcomes even when there are no material benefits of high rank (Charness et al., 2013b). This might arise from a desire for status (Charness and Grosskopf, 2001), self-image concerns (Bénabou and Tirole, 2006), or a joy of outperforming others (Dohmen et al., 2011). Our setup does not provide subjects any opportunities for sabotage or cheating.

[10] See for example, Bolton and Dewatripont (2005, pp. 130-136)

effort levels (e_1, e_2). Agents can observe their partners' outcomes but not effort. Agents care not only about their own outcomes but also how it compares with their partners'. Since the compensation scheme is identical for the two agents, knowing the partner's outcome is equivalent to knowing how the partner's gross compensation. Let w_i denote the gross compensation of agent i. Hence, w_i is f when agent i 's project fails and $(f+b)$ when it succeeds. For agent i who takes agent j 's earnings as a reference, her optimization problem described in (1) becomes:

$$\max_{e_i}\ p(e_i)\, u(f - ce_i + b) + (1 - p(e_i))\, u(f - ce_i) - E\left[v\left(w_j - w_i\right)\right] \tag{2}$$

where the first bracket is the expected utility of agent i from her net monetary payoffs (as in (1)), and the second term is her expected disutility from the inequality in gross wages. Neilson and Stowe (2010) models v as follows:

$$\mathrm{v}(x) = \begin{cases} v(x) & \text{if} \quad x \geq 0 \\ \lambda v(-x) & \text{if} \quad x < 0 \end{cases}$$

where v is differentiable, weakly positive, strictly increasing and convex such that $v(0) = 0$, and λ is a constant. Note that, in this representation, agent i dislikes it when her partner's gross wage is higher than hers.[11] We call this *behindness* aversion. When agent i 's gross wage is higher than her partner's, then the sign of the disutility term in problem 2 depends on the sign of λ. A positive λ implies that agent i dislikes being ahead, and a negative λ implies that agent i enjoys being ahead — perhaps out of a preference for *competition* (MacCrimmon and Messick, 1976) or *status* (Frank, 1985). Fehr and Schmidt (1999) support behindness aversion by assuming $|\lambda| < 1$.

In our experiments, we create the benchmark described by model (1) by having individuals complete their tasks in isolation. We also have experiments where subjects receive feedback about the performance of a partner to test the predictions of model (2).

[11] Note that one may alternatively construct a model where net rather than gross earnings are compared. We study the case of gross earnings comparisons to be close to the literature. Moreover, the unobservability of a partner's effort in our experiment makes relative net earnings comparisons difficult. We leave it as an open question for future research to investigate the effect of feedback about net earnings on effort.

Given the two-outcome set-up and payment scheme (f, b), problem (2) becomes:

$$\max_{e_i} \; [p(e_i)\, u(f - ce_i + b) + (1 - p(e_i))\, u(f - ce_i)] - [p(e_j)\,(1 - p(e_i))]\, v(b) \\ - [p(e_i)\,(1 - p(e_j))]\, \lambda v(b) \tag{3}$$

where the second term is the expected disutility in the event that agent j succeeds and agent i fails; and the third term is the expected disutility in the event that agent i succeeds and agent j fails. These last two terms are absent in problem (1).

Let e_i^{Ind} and $e_i^{Partner}$ be solutions to problems (1) and (2), respectively. If $\lambda \leq 0$ and the agent is competitive or status-seeking, then in the symmetric equilibrium of problem (2), $e_i^{Partner} > e_i^{Ind}$.[12] If $\lambda > 0$ and the agent dislikes being ahead, then the model still predicts higher effort from having a partner as long as λ is sufficiently small, as may be implied by behindness aversion.[13] When $\lambda > 0$ and large, then the disutility of being ahead dominates and leads to lower effort from having a partner to avoid such guilt.

Note that the two versions of the problem above assume a one-shot environment and require that people with behavioral concerns anticipate the possible disutility from inequity when choosing effort. One may argue that this is a strong assumption. Even if people compare their outcomes with their partners' outcomes ex-post, they may be incapable of anticipating such feelings before taking an action. If this is true, then people will solve problem (1) in a one-shot decision regardless of whether they expect to receive feedback about their partner's outcome. The experienced inequity may lead to some form of adaptive learning and revision of strategies to avoid the previous outcome.

Our experiment requires subjects to choose their effort twice in a two-period problem. Hence, we can analyze whether agents can anticipate the disutility from inequity, or whether they need to experience such a situation to revise their effort choices in the second period.

One needs to be careful about possible income effects in a two-period setup. An agent with selfish utility would solve a different problem in the second period after earning a high payoff in the first period than earning a low payoff. In order to minimize such an income effect in the experiment, we pay subjects for one randomly selected period. The period which they are

[12]See Appendix A for proof.

[13]If $\lambda < \frac{p(e^{Partner})}{1-p(e^{Partner})}$, then $e^{Partner} > e_i^{Ind}$. Fehr and Schmidt (1999) argues $|\lambda| < 1$ and in our data, the average probability of success in the partner treatment given the average effort (9.5) is roughly 0.39 which makes $\frac{p(e^{Partner})}{1-p(e^{Partner})} = 0.63$.

paid for is determined randomly ex-post (with equal probability). In such an environment, an expected utility maximizer whose utility depends only on own-payment and is free from reference dependence would choose the same effort level in both periods, independent of her success or failure in the first period (see Appendix A for proof). The assumption of random payment might be violated in some applications. However, we implement such a payment structure in order to focus on the role of performance feedback.

III HYPOTHESES

Our main question is how observing a peer's success or failure affects effort in a risky task that is completed and rewarded independently?

As mentioned, a sophisticated agent with behavioral motives may anticipate the potential disutility from unequal outcomes, knowing that she will receive feedback about her partner's performance (as modeled in problem (2)). This might influence her effort. Hypothesis 1 summarizes this idea.

Hypothesis 1: *Agents choose different period 1 effort in the risky task when they anticipate feedback about a partner's performance as compared to when they do not receive (or anticipate) such feedback.*

As shown in the behavioral model, depending on the relative strength of possible emotions, agents may exert higher or lower effort when they expect feedback about a partner's outcome as compared to working in isolation. Based on the previous literature, we expect the disutility from being behind to be stronger than the disutility from being ahead (or there to be a status seeking motivation at play). Hence, if subjects are indeed sophisticated in anticipating the disutility from social comparisons, then we conjecture they exert higher effort with a partner than without.

Hypothesis 1 is about first period behavior and builds on the assumption that agents can anticipate a disutility from social comparison. However, those who do not anticipate such concerns may start thinking about social comparisons once they experience inequity, and then incorporate such behavioral concerns into their objective functions. Hence, they may start with a reference-free utility and incorporate the feedback as a reference into their utilities later. Testing such adaptation of the utility function requires examining the dynamics of effort revision with feedback.

Note that an expected utility maximizer working in isolation should pick the same effort in each period, independant of the feedback received in the first period. Hypothesis 2 below states this implication of the standard theory.

Hypothesis 2: *Agents conducting their tasks in isolation choose the same effort in periods 1 and 2.*

One may be skeptical of Hypothesis 2 if one believes that agents are unsure about their optimal effort [14] or do not fully understand the production function of their effort[15]. Such agents may revise their choice in the second period, even when working in isolation.

When paired with a partner, an agent who fails to anticipate behavioral concerns will use the objective function described in problem (1) in the first period and continue to do so in the second period as long as she receives the same outcome as her partner. However, once she experiences an unequal realization, she may start using the behavioral model described by problem (2). Our next hypothesis addresses the dynamics of reference dependent behavior in our set-up.

Hypothesis 3: *Agents who receive unequal outcomes from their partners in period 1 revise their effort differently than agents who receive the same outcomes as their partners.*

Hypotheses 1 and 3 are independent. If Hypothesis 1 is supported, and subjects anticipate social comparisons in their effort choice the first time they make the decision, they may choose the same effort the second time they make the decision. Hence, we may reject Hypothesis 3. It is also possible to support both Hypotheses because even if subjects anticipate social comparisons, the dynamic decision problem may be different than the static one as discussed

[14]Impulsive reactions and reinforcement learning strategies have been studied in the learning literature. The idea is that after some experience, people think about what might have been a better decision last time, and then adjust their behavior in that direction (Avrahami and Kareev, 2011; Hart et al., 2016; Kareev et al., 2014; Selten and Buchta, 1998; Selten and Stoecker, 1986). Subjects may also revise their effort choice in the second period due to adaptive learning. However, umlike a typical adaptive learning environment, we have a single person decision problem with objective probabilities.

[15]Bandiera et al. (2015) shows that feedback on past performance affects students' future performances. They argue that students have imperfect information on how their effort translates into test scores and that the provision of feedback might be a cost-effective means to increase students' exam performance. Note that in our setup subjects know the objective probability of success for each effort level.

in Hypothesis 2. On the other hand, if we reject Hypothesis 1, and conclude that agents cannot anticipate the disutility from social comparison, we may still (i) find support for Hypothesis 3 because once they experience a differential outcome, they may start taking into account behavioral concerns, or (ii) reject Hypothesis 3 if there are no behavioral concerns at play.

Activation of nonzero v function or nonzero parameter λ in the behavioral model may also capture heterogeneity in the population. Motivated by the literature on gender differences in risk attitudes and inequality aversion, we will analyze the data not only at the aggregate level but also by gender. As we will see, some of our findings can shed light on observed gender differences in labor settings and other applications involving the influence of feedback.

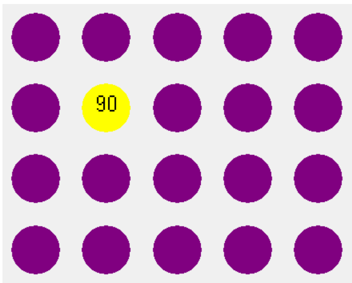
IV EXPERIMENTAL PROCEDURES

The experiments were run at the Experimental Economics Lab at the University of Maryland (EEL-UMD). All participants were undergraduate students at the University of Maryland. The data was collected in 25 sessions and there were two parts in each session. No subject participated in more than one session. Sessions lasted less than an hour. The subjects answered two identical effort choice problems in Part 1, and their risk attitude and fairness concerns were elicited in an incentivized way in Part 2. In each session, subjects were asked to sign a consent form first and given written experimental instructions (provided in Appendix C), read to them by the experimenter. The instructions for Part 2 were given after Part 1 of the experiment was completed. Before the experiment started, subjects answered a hand-written quiz testing their understanding of the decision problem. The experimenter walked around the room and checked each subject's answers, and the experiment did not start until everyone answered the quiz correctly.

The experiment was programmed in z-Tree (Fischbacher, 2007). The screen shots for the decision and feedback screens can be found in the instructions provided in Appendix C. All earnings in the experiment were denominated in Experimental Currency Units (ECUs). The final earnings of a subject were the sum of her payoffs in one randomly selected decision problems (out of two) in Part 1, one randomly selected decision problem (out of two) in Part 2, and the participation fee of $7. The payoffs in the experiment were converted to US dollars at the conversion rate of 15 ECUs = 1 USD. Cash payments were made at the conclusion of the experiment in private. The average payment was $ 13.29 (including a $7 participation fee).[16]

The decision problem in the first part of the experiment was described to the subjects as a project. A project is represented by a box with 20 balls in it as in Figure 1. One of the balls is yellow and the others are purple. The project succeeds if the yellow ball is drawn from the box in any of the subject's tries. A subject decides how many times she wants to try drawing the yellow ball with *replacement*. Each try costs 3 ECUs. In order to afford these tries, each subject is given a 60 ECUs endowment. Subjects can choose any number of tries from $\{0, 1, 2, \ldots, 20\}$.[17]

Figure 1: Projects



Project is successful if at least one yellow ball is drawn from box with replacement.

After choosing a total number of tries, the computer randomly draws one ball with replacement for each try. The computer continues drawing balls until all the tries are exhausted, regardless of the color of the ball drawn on the previous try. Each try is an independent draw where the chance of drawing a yellow ball is $\frac{1}{20}$. If the yellow ball is drawn in at least one try, the project succeeds and the subject receives 90 ECUs prize.[18] If a purple ball is drawn in all tries, then the project fails and the subject receives zero prize. Hence, if we interpret the chosen number of tries as the subject's choice of costly effort, $e \in \{0, 1, 2, \ldots, 20\}$, then the probability of success is $p(e) = 1 - \left(\frac{19}{20}\right)^e$ where $\frac{19}{20}$ is the probability of drawing a purple ball in a single draw. Note that this particular function $p(.)$ satisfies the assumptions made of the theoretical model.[19] While not presented this way, one may also interpret an effort choice

[16] Including the participation fee, participants could earn a minimum of \$ 7.13 and a maximum of \$ 21.47 during the experiment. In practice, one unlucky participant earned \$7.13 and the highest payment earned was \$20.67.

[17] There are many ways to create a moral hazard problem or risky investment problem in the laboratory. Our objective was to keep the environment easy to understand and close to the (standard and behavioral) theoretical model. We chose to have a two outcome, multi-effort decision problem as such a rich action space allows large variation in the data for a given outcome and increases our statistical power.

[18] A subject can only receive the 90 ECus once, even if the subject draws the yellow ball in more than one try.

[19] Implementing a stylized effort provision setup allows us to know and control both the cost of effort and the probability of success function. Furthermore, Dutcher et al. (2015) find identical results when comparing an experiment with stylized effort and real effort provision.

of e as opting for the following lottery where the probability of high outcome increases with e and the prizes decreases with it.[20] In this light, our findings can be generalized to other risk taking or investment problems with uncertainty where decision makers have reference dependent utilities.

In order to help subjects calculate the probability of success for each effort level, we present a table of probabilities in the printed instructions as well as the decision screens.

This methodology is also used by Sjöström et al. (2019) in a different research question.[21] We believe that our subjects might be familiar with this kind of decision problem as it is analogous to effort choices in their lives. For example, a student who is deciding how many summer internship applications to complete understands that the chance of getting an internship increases with each application, and that having at least one internship offer (like drawing at least one yellow ball) is enough to receive an internship. Furthermore, the probability of getting a summer internship increases more in the first application than the 20$^{\text{th}}$. Similarly, while preparing for an exam, a student decides how many topics to master. The more topics studied, the higher the chance of receiving a familiar exam question, but the chances of seeing a familiar question may improve more in the first hour of study than the 20$^{\text{th}}$ hour.

In each period, subjects receive a new 60 ECUs endowment and decide how much effort to select that period. We have two treatments in Part 1: the Individual and the Partner Treatments. We had 122 subjects in the Individual Treatment (8 sessions) and 261 subjects in the Partner Treatment (17 sessions).[22] The two treatments differ from each other as described below.

Individual Treatment: Each subject chooses her effort level in two periods. After each period, subjects learn their project's outcome and their net payoff. Subjects do not learn anything about the performance of others in the room.

---

[20] In theory, there should not be any difference between these alternative presentation methods, but we hope that explaining the procedure generating the probability of success and how it relates with their effort choice will improve the subjects' understanding of the decision problems. Our pilot experiments as well as the experiment of Sjöström et al. (2019) show that subjects understand this design and respond to changes (such as endowment, prize, and number of yellow balls in the box) in a rational way.

[21] They aim to understand how the motivation to exert effort varies based on whether the task is assigned by the experimenter or chosen by the subject.

[22] We collected more data for the Partner Treatment than the Individual one, in order to have enough observations for each of the four possible history in the Partner Treatment.

Partner Treatment: At the beginning of each period, each subject is randomly matched with a partner in the room. Subjects do not know the identity of their partners. After choosing an effort level, a subject learns the outcomes of both her and her partner's projects. She does not learn her partner's effort (hence, she cannot infer the partner's net payoff in that period).[23] Subjects encounter a new partner in each round to prevent reputational concerns from creating a repeated game.

Note that the only difference between the two treatments is whether a subject receives feedback about her relative performance with respect to a partner's outcome. If a subject does not consider a partner's success or failure as reference, then such feedback about the partner's performance is irrelevant information and there should be no difference in the distribution of effort between the treatments (as implied by the standard theory).

Part 2 of our experiments elicits subjects' risk attitudes and fairness preferences and does not vary across treatments. We chose two well-known and easy to administer methods that are often preferred when measures are needed for control purposes rather than the heart of the study. To elicit risk preference we used a simple methodology (also used by Eckel and Grossman (2002) and Dave et al. (2010)) where subjects asked to choose one out of six gambles of varying levels of risk.[24] To measure fairness concerns, we asked subjects to pick one of the three allocations of money between themselves and another randomly assigned subject differing in their maximization of selfish, equity and efficiency concerns (Charness and Rabin, 2002; Engelmann and Strobel, 2004). After completing Part 2, subjects self-reported their gender, SAT/ACT scores, age, gpa, major, current employment status, job experience, and two questions about their work and grading preferences.[25]

[23]Subjects in the Partner Treatment learn only about the partner's outcome but not about the partner's effort. We chose this feedback structure for two reasons: (i) In many applications of moral hazard problems, an agent's outcome but not effort is observable. For example, a researcher may see another colleague's publications on her resume but may not be able to see how hard that colleague is working. (ii) We are interested in the effect of feedback on one's effort choice. Providing feedback on outcome can create four possible history in the Partner Treatment. One of these four realizations of outcome can occur in terms of a subject's own outcome and partner's outcome, i.e. {(Fail,Fail), (Fail,Success), (Success, Fail), (Success, Success)}. However, if we had provided feedback on partner's effort choice as well, there would have been around 21x21=441 different possible realizations of history for each of the outcome pair and that would have been very hard to collect enough observations. Such a research question would require a different design with a smaller set of effort levels and we leave that as an open question for future research.

[24]See Charness et al. (2013a) for a review of prevailing methods of eliciting risk preferences.

[25]One question asked whether subjects preferred their homework to be graded based on their individual scores or on their scores relative to others in their class; and the other question asked whether they preferred to work on homework individually or in a team.

V RESULTS

Before studying subjects' choice of effort, we check to see whether there is balance between the two treatments along various demographic factors and behavioral measures collected in part 2 of the experiment. The results from this balance comparison are reported in Table B.1 in Appendix B. Except for age, which has a small yet statistically significant difference between treatments (half a year higher average age in the Partner Treatment than the Individual Treatment), there are no apparent statistical differences between treatments. Furthermore, the significance of the difference in age would disappear if we corrected for multiple hypothesis testing. We conclude that there is little relevant difference in the underlying characteristics of subjects between treatments.

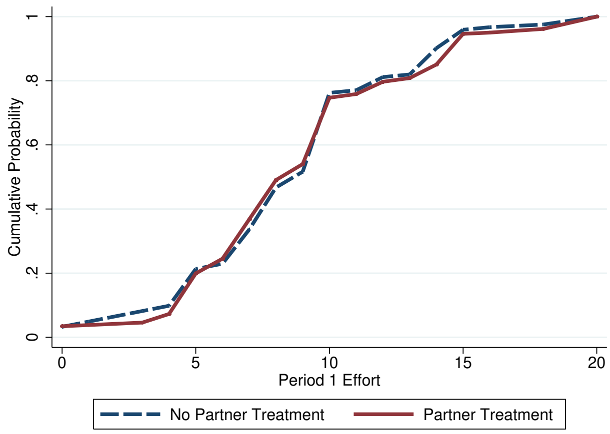
Choosing zero tries (zero effort) is the risk-free option in the decision problem. A subject with zero attempts at drawing a yellow ball will fail on the project for sure and keep the endowment of 60 ECUs. On the other hand, choosing the maximum number of tries of 20 requires a subject spending her entire endowment, receiving the prize of 90 ECUs with around 0.64 probability and receiving zero ECUs otherwise. A subject's expected return is maximized at $e = 8$. Hence, risk aversion predicts an effort choice less than 8, and extreme risk aversion predicts an effort choice of zero. We found that subjects chose an average effort level of 9.44 across both periods and treatments, with only 35% (266 out of 766 observations) of observations falling below 8. Hence, most subjects exerted more than the effort level implied by the maximization of expected return. This is unsurprising in the context of our experiment, as many other experiments conducted in labor-environment settings document that subjects take actions higher than would maximize expected returns. For example, even under a flat payment scheme where subjects are paid fixed amounts independent of effort, subjects exert positive effort (see for example Charness et al. (2013b), and Falk and Ichino (2006)). Thus, we somewhat expected to see such high risk taking in this labor context.

We start by comparing the first period effort in the two treatments.[26] Recall that if subjects are inequity averse and can anticipate their inequity aversion without any prior experience in the task, then Hypothesis 1 predicts that period 1 effort is subject to vary based on the treatment. Figure 2 shows that the cumulative distributions of period 1 effort is similar between

[26] In this section, where applicable, we report p-values from two-tailed t-tests comparing the means of relevant data samples. We leave out the p-values from Mann-Whitney U tests when they deliver similar results and report them if there are contradictory results by these two methods.

the two treatments. We fail to reject the null hypothesis that the two cumulative distributions are equal (p=0.972 in a Kolmogorov-Smirnov test). Figure B.1 in Appendix B shows that there are also no discernible differences in the period 1 effort cumulative distributions between the two treatments when broken down by gender. The average period 1 effort in the Individual and the Partner Treatments are 9.00 and 9.20, respectively, with p=0.673 in a two-tailed t-test. Table B.2 in Appendix B also shows that this result is robust to a regression analysis. We use a Two-Limit Tobit regression to account for censoring from below (at 0) and above (at 20), regressing subjects' period 2 effort on their assigned treatment, and demographic and behavioral measures from Part 3 and the end of experiment survey.[27] The partner treatment dummy is insignificant when genders are pooled or separated, indicating that neither gender anticipates inequity concerns in selecting their period 1 effort. Thus, we reject Hypothesis 1 that a subject's effort varies based on her anticipation of feedback regarding a partner's relative performance.

Figure 2: CDFs of Period 1 Effort



Result 1: *We reject Hypothesis 1. Subjects choose similar effort in the first period when they know they will receive feedback about their partner's outcome as compared to when they do not receive such feedback.*

Next, we study whether effort choices vary between periods in the Individual Treatment. If subjects do not have behavioral motives and simply maximize the expected utility of their monetary payoffs, then they should solve the optimization problem independently in the two

[27]We also control for subjects risk attitudes, fairness preferences, gpa, and relative grading preferences. The results are robust to adding in controls for additional demographic or behavioral variables as well.

periods and hence submit the same effort level in each, as argued in Hypothesis 2. Rejection of Hypothesis 2 implies that a subject's effort may be affected by the outcome in the first period. Our next set of results investigate this question.

Since only 37% of subjects in the Individual Treatment submit the same effort choice in the two periods, we reject the null hypothesis that 100% choose the same effort between periods (p=0.000 in a two-tailed t-test).[28] In our setup, the probabilities are fixed and objectively given. Thus, one may argue that the findings documented below indicate that subjects revise their objective functions, using the feedback as a reference in determining their next-period optimization. This is consistent with the finding in the reinforcement effects literature that subjects change their behavior even when the probability distribution over outcomes is relatively simple (Nielsen, 2019).

In general, female subjects are more likely to revise their second period effort than male subjects (73% and 55% respectively, p=0.033 in a two-tailed t-test).

Moreover, the direction of effort revision varies based on feedback about the first period outcome. As seen in Figure 3a (as well as its table version, Table B.3 in Appendix B), while 48% of failed subjects increased their effort, only 16% of successful subjects increased their effort (49% of them decreased their effort, and the remaining 35% did not change their effort). This difference between the upward revision rates of failed and successful subjects is statistically significant (p=0.001 in a two-tailed t-test). Additionally, the effect of failure on likelihood of upward effort revision is more pronounced for females than males. For females, the rate of upward effort revision increases from 11% given success to 58% given failure (p=0.001 in a two-tailed t-test), and for males the rate of upward effort revision increases from 21% given success to 40% given failure (p=0.139 in a two-tailed t-test).

We find that subjects are more likely to downward revise their effort if they were successful in period 1 than if they failed. As seen in Figure 3b (also in Table B.3 in Appendix B), the downward effort revision rate of successful subjects is 49%, while that of failed subjects is 14% (p=0.000 in a two-tailed t-test). [29] Here too, the effect of success on downward effort revision is more pronounced for females than males: success leads females to downward revise their

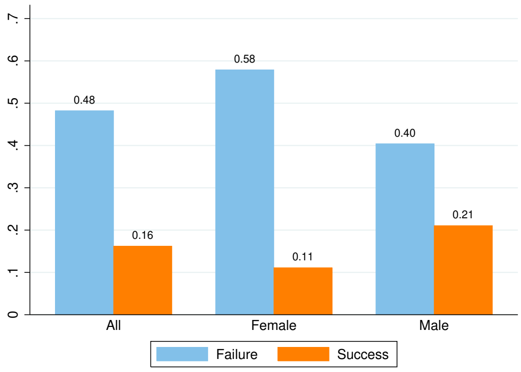
[28]Similarly, in the Partner Treatment, we observe that only 35% of subjects choose the same effort between periods.

[29]Similarly, in the Partner Treatment we find more frequent upward effort revision of failed (50%) than successful subjects (21%, p=0.000 in a two-tailed t-test); and more frequent downward effort revision of successful subjects (33%) than failed subjects (21%, p=0.031 in a two-tailed t-test). These results are reported in Table B.9a in Appendix B.

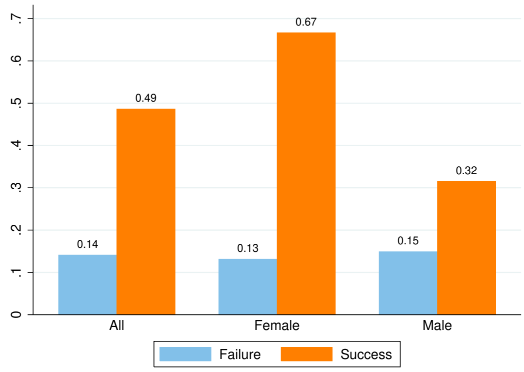
effort in 67% of instances and failure in 13% of instances (p=0.000 in a two-tailed t-test); by contrast, success leads males to downward revise their effort in 32% of instances and failure in 15% of instances (p=0.127 in a two-tailed t-test).

These results summarized by Figure 3a and 3b are supported by a regression analysis. Tables B.4a and B.4b in Appendix B report results from Probit regressions of a subject's period one outcome on the likelihood of upward and downward effort revision, respectively. Table B.4a shows that failing leads to a 33 to 34 percentage point increase in the likelihood of upward effort revision when genders are pooled, and a 51 to 54 percentage point increase in the likelihood of upward effort revision for females. However, the effect of failure on upward effort revision is statistically insignificant for males. Similarly, as seen in Table B.4b, success leads to a 28 percentage point increase in the likelihood of downward effort revision when genders are pooled, and a 46 to 47 percentage point increase in the likelihood of downward effort revision for females, while the effect of success on downward effort revision is statistically insignificant for males.[30]

Figure 3: Individual Treatment, Proportion of Subjects Revise Effort in Period 2



(a) Proportion of Subjects who Increase Effort in Period 2



(b) Proportion of Subjects who Decrease Effort in Period 2

A principal in a moral hazard problem must not only consider which direction agents revise their effort, but also *by how much*. Next, we analyze the magnitude of effort level revision between the two periods. Table 1 summarizes the mean of effort level change in two

[30]The above analysis shows that there are differential effects of failure and success on effort revision between genders. It is interesting to note that we also observe signs of gender differences when comparing the revision rates of males and females for a given history: Figure 3b shows that 67% of females and only 32% of males decreased their effort after a success (p=0.033 in a two-tailed t-test); and Figure 4a shows that 58% of females and 40% of males increased their effort after a failure (p=0.112 in a two-tailed t-test). In the Partner Treatment, failed female subjects also upward revise their effort level more frequently than failed male subjects, 63% and 37% respectively (p=0.001 in a two-tailed t-test). However, in the Partner Treatment, successful female subjects downward revise their effort at a similar rate to successful male subjects (39% and 30% respectively, p=0.335 in a two-tailed t-test). These Partner Treatment results are reported in Table B.9a in Appendix B.

periods, by period 1 outcome and gender. In comparing the period 2 effort of failed and successful subjects, one needs to control for their period 1 effort because subjects who failed exert lower period 1 effort on average (8.35) than subjects who succeeded (10.30). Thus, the two populations may have different underlying preferences. This is why Table 1 reports the mean revision in effort between periods rather than the average absolute period 2 effort level. Mean period 2 effort levels can be found in Table B.5 in Appendix B.

Table 1: Individual Treatment, Mean (Period 2 Effort – Period 1 Effort)

Period 1 Outcome	All	Female	Male
Failed	1.38***	1.55**	1.23*
	(4.28)	(4.08)	(4.48)
	n=85	n=38	n=47
Succeeded	-0.95**	-1.39*	-0.53
	(2.63)	(2.87)	(2.39)
	n=37	n=18	n=19
P-values	0.003	0.008	0.111

*** p<0.01, ** p<0.05, * p<0.1

Stars indicate whether mean is statistically different from zero in a two-tailed t-test. Standard deviations in parentheses. n=number of observations. Bottom row reports p-values from a two-tailed t-test comparing the mean effort revision of subjects who failed and succeeded in a given column.

Table 1 shows that subjects who failed increased their effort by 1.38 tries on average, and subjects who succeeded decreased their effort by 0.95 tries on average.[31] Furthermore, these two effects are more pronounced for females than males. The results are supported in Two-Limit Tobit regressions where the dependent variable is subjects' period 2 effort and the independent variables are period 1 effort and a dummy for failure in period 1 (see Table B.6 in Appendix B). When both genders are pooled, the coefficient on failure is positive and significant. When genders are separated, we see that this is driven by the fact that the coefficient is significant and positive for females, but insignificant for males.

[31]Similarly, in the Partner Treatment we find that subjects who failed increase their effort by 1.49 on average, and subjects who succeeded decrease their effort by 0.98 on average (see Table B.8 of Appendix B). Furthermore, we find that the effect of own-failure is statistically indistinguishable in the Individual and Partner Treatments. When comparing the effect of own-failure between the two treatments, we want to control for the potential impact of peer effects in the Partner Treatment. To do so, we separately compare the effect of own-failure for subjects in the Individual Treatment with that of i) subjects in the Partner Treatment whose partner succeeded and ii) subjects in the Partner Treatment whose partner failed. Table B.10 in Appendix B shows Tobit regressions of failure, being in the partner treatment, and their interaction on period 2 effort for the two groups separately (the former group in columns (1) and (2), and the latter group in columns (3) and (4)). The statistical insignificance of the interaction term indicates that the effect of own-failure is similar in the two treatments.

Result 2: *We reject Hypothesis 2. In the Individual Treatment, failed subjects increase their effort and successful subjects decrease their effort. Both the effect of failure on upward effort revision, and the effect of success on downward effort revision is more pronounced for female subjects than male subjects.*

The Partner Treatment allows us to explore how a subject's period 2 effort is affected by comparison of her period 1 outcome with that of her partner. If an agent makes no social comparison, then her period 2 effort choice in the Partner Treatment should be independent of her partner's period 1 outcome. Hence, for example, the effort choice of failed subjects should be the same regardless of their partner's success or failure.

Tables 2 and 3 report subjects' average period 2 effort and revision in effort between periods 1 and 2, respectively, by period 1 outcome and gender. As seen in Table 2, the average period 2 effort of failed subjects whose partner succeeded is higher than those whose partner also failed (10.82 versus 8.94 respectively, p=0.023 in a two-tailed t-test).[32] This observation holds when genders are either pooled or separated, applying to both females and males. Furthermore, this result is largely driven by the fact that *the magnitude*, rather than *the rate*, of upward effort revision of failed subjects is higher for those whose partner succeeded than those whose partner failed (an average increase of 2.41 and 0.88 respectively, as seen in Table 3). Indeed, 15% (10 out of 66) of failed subjects whose partner succeeded choose the maximum effort of 20 in period 2, while only 5% (5 out of 99) of failed subjects whose partner failed choose the maximum effort in period 2. Both male and female subjects increased their effort significantly after observing an outcome behind their partner's, i.e. after observing history (Fail, Success): female subjects increased their effort by 2.89, and male subjects increased their effort by 2.05 (Table 3, second row).

Subjects who succeeded in period 1 were not as sensitive to their partner's outcome as those who failed. This is potentially good news for a principal in a moral hazard problem. Among successful subjects, only male subjects whose partner succeeded, i.e. observed history (Success, Success), show a statistically significant change in effort — decreasing their effort by 2.42 (p=0.052). However, the number of observations that generates this result is not very large (n=19).

[32] At the same time, the average period 1 effort of failed subjects is similar regardless of their partner's success or failure, as one would expect (8.41 and 8.06 respectively).

Table 2: Partner Treatment, Mean Period 2 Effort

Period 1 Outcome (Self, Partner)	All	Female	Male
(Fail, Fail)	8.94	8.95	8.88
	(4.86)	(4.45)	(5.54)
	n=99	n=56	n=41
(Fail, Success)	10.82	11.43	10.37
	(5.57)	(5.52)	(5.64)
	n=66	n=28	n=38
(Success, Fail)	10.09	9.33	10.37
	(4.98)	(4.15)	(5.32)
	n=66	n=27	n=38
(Success, Success)	9.57	8.64	10.11
	(4.07)	(3.35)	(4.43)
	n=30	n=11	n=19

Standard deviations in parentheses. n=number of observations. Female and male columns exclude 3 subjects who self-reported their gender as neither female nor male. 2 of these subjects received outcome (Fail, Fail), and 1 such subject received outcome (Success, Fail).

Table 3: Partner Treatment, Mean (Period 2 Effort – Period 1 Effort)

Period 1 Outcome (Self, Partner)	All	Female	Male
(Fail, Fail)	0.88*	1.13*	0.59
	(4.73)	(4.68)	(4.94)
	n=99	n=56	n=41
(Fail, Success)	2.41***	2.89**	2.05**
	(5.20)	(4.46)	(5.72)
	n=66	n=28	n=38
(Success, Fail)	-0.58	-0.67	-0.66
	(3.28)	(3.23)	(3.27)
	n=66	n=27	n=38
(Success, Success)	-1.87**	-0.91	-2.42*
	(4.20)	(1.76)	(5.08)
	n=30	n= 11	n=19

*** p<0.01, ** p<0.05, * p<0.1

Stars indicate whether mean is statistically different from zero in a two-tailed t-test. Standard deviations in parentheses. n=number of observations. Female and male columns exclude 3 subjects who self-reported their gender as neither female nor male.

The results outlined regarding the influence of social comparison on period 2 effort are supported by regression analysis. Table 4a and 4b report Two-Limit Tobit regressions for failed and successful subjects, respectively. They show regressions of period 2 effort (censored at 0 and 20) on a dummy for receiving an unequal outcome with one's partner in period 1 and period 1 effort. For failed subjects (Table 4a), the coefficient for receiving an unequal outcome from one's partner is significant and positive, indicating that failed subjects select greater period 2 effort if their partner succeeded (columns 1-2). However, when the regression is performed separately by gender, this dummy is significant at the 5% level for females, and only significant at the 10% level for males. That said, it is positive and of similar magnitude for both genders (columns 3-6). Furthermore, when we carry out the same regressions for successful subjects (Table 4b), we find the dummy for receiving an unequal outcome from one's partner is insignificant when genders are either pooled or separated. This supports our conclusion that the period 2 effort level of successful subjects is not sensitive to their partner's outcome.

Table 4: Partner Treatment, Two-Limit Tobit of Period 2 Effort

(a) Subjects who Failed in Period 1

	(1) All	(2) All	(3) Female	(4) Female	(5) Male	(6) Male
Unequal Outcomes	1.907**	2.259***	2.346**	2.422**	1.796	2.410*
	(0.845)	(0.835)	(1.114)	(1.112)	(1.356)	(1.291)
Period 1 Effort	0.687***	0.695***	0.653***	0.676***	0.735***	0.746***
	(0.109)	(0.108)	(0.170)	(0.172)	(0.153)	(0.150)
Constant	3.351***	-0.991	3.904***	0.187	2.515	-3.541
	(0.324)	(3.120)	(1.469)	(0.389)	(1.610)	(0.519)
Controlling for Heterogeneity		Yes		Yes		Yes
Observations	165	164	84	83	79	79
Log-Likelihood	-463.9	-456.4	-237.9	-233.9	-218.6	-213.7

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

(b) Subjects who Succeeded in Period 1

	(1) All	(2) All	(3) Female	(4) Female	(5) Male	(6) Male
Unequal Outcomes	1.342	1.326	0.389	0.262	1.833	1.322
	(0.821)	(0.836)	(0.981)	(0.940)	(1.184)	(1.193)
Period 1 Effort	0.818***	0.821***	0.741***	0.736***	0.848***	0.849***
	(0.092)	(0.094)	(0.115)	(0.113)	(0.134)	(0.128)
Constant	0.189	-0.357	1.562	-0.229	-0.567	1.210
	(1.253)	(2.314)	(1.376)	(2.548)	(1.919)	(3.378)
Controlling for Heterogeneity		Yes		Yes		Yes
Observations	96	96	38	38	57	57
Log-Likelihood	-248.4	-247.2	-90.46	-88.35	-151.2	-147.8

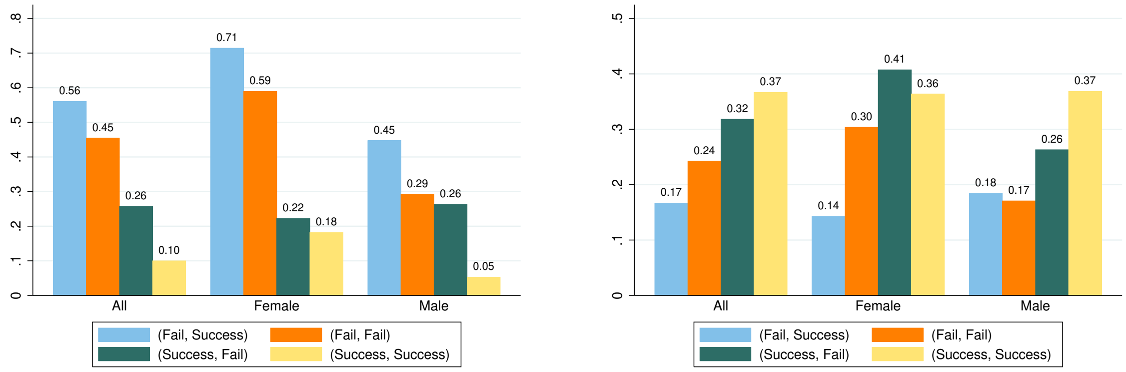
Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Dependent variable is period 2 effort. Censored from below at 0 and above at 20. Unequal outcomes is a dummy equal to one if a subject received a different outcome from their partner in period 1. Controlling for heterogeneity includes controls for risk preferences (dummy equal to 1 if the gamble chosen is either gamble 4, 5 or 6), fairness preferences (dummy for 2 of the 3 possible allocations in the fairness task), a linear term for GPA, and relative grading preferences (dummy equal to 1 if the subject is indifferent to, or strictly prefers being graded relatively to individually on a homework assignment). Female and male columns exclude 3 subjects who self-reported their gender as neither female nor male. Column (4) of Table 4a excludes 1 subject who did not report a GPA.

We argued before that the observed increase in period 2 effort of subjects with history (Fail, Success) is largely driven by the magnitude, rather than rate, of upward period 2 effort revision being influenced by inequity concerns. In other words, failed subjects are likely to increase their effort regardless of their partner's outcome but increase their effort more if they see their partner succeed. Figure 4a and 4b show the proportion of subjects who revise their effort upwards and downwards, respectively, broken down by history and gender (one can find this figure's table version in Table B.9b in Appendix B).

Figure 4: Partner Treatment, Proportion of Subjects Revise Effort in Period 2



(a) Proportion of Subjects who Increase Effort in Period 2

(b) Proportion of Subjects who Decrease Effort in Period 2

Figures 4a and 4b lend added support to our conclusion. We see that subjects who failed in period 1 are somewhat more likely to increase their effort if their partner succeeded than if their partner failed (56% versus 45%, respectively, blue and orange columns in Figure 4a). However, this difference is not statistically significant (p=0.184 in a two-tailed t-test). Additionally, there is no statistically significant difference in the upward effort revisions rates of (Fail, Success) and (Fail, Fail) subjects when genders are separated (p=0.269 for females and p=0.158 for males). Moreover, subjects who succeed in period 1 increase or decrease their effort in similar proportions regardless of their partner's success or failure (green and yellow columns in Figure 4a and 4b). We summarize our findings in period 2 of the Partner Treatment as follows.

Result 3: *We find support for Hypothesis 3 for failed subjects, but not for successful subjects.*

1. *Failed subjects select higher effort in period 2 when their outcome is behind their partner's. This effect is more pronounced for female subjects than male subjects.*
2. *Successful subjects are not sensitive to their partner's outcome.*

In short, subjects behave consistent with behindness aversion (i.e. $v \geq 0$). Recall from the discussion of the behavioral model that Hypotheses 1 and 3 might be independent. While we reject Hypothesis 1, we find support for Hypothesis 3 for failed subjects. This indicates that subjects may not be able to anticipate social comparisons in the Partner Treatment, hence, acting similarly in the first periods of the Individual and Partner Treatments. However, once they experience an outcome lower than their person of reference, they are motivated to exert higher effort. Those who are ahead of their person of reference do not seem to diminish (as suggested by "aversion to be ahead" and a positive λ in the behavioral model) or to increase their effort (as suggested by "competitiveness" or "status seeking" and a negative λ in the behavioral model).

While behindenss aversion can explain our results in the Partner Treatment, it cannot explain all of our results. Given our findings in the Individual Treatment and the literature showing that past realizations of uncertainty can affect future individual risk taking, it is fair to argue that both social comparisons and some adaptation of preferences are taking place. Our findings might be viewed as indicating that prior experiences can create a new reference in utility, triggering a revision of strategies.

VI DISCUSSION

Our findings have direct implications for incentivizing agents in a moral hazard problem. The decision problem subjects faced require them to choose a costly effort level which increases their chance of winning a prize. This is a typical moral hazard problem with an outcome contingent payment scheme. We identify two feedback dependent motivations for higher effort: (i) feedback about an agent's failure motivates higher effort, and (ii) feedback about receiving a worse outcome than a partner motivates an agent to exert even higher effort. The standard model cannot explain either of these two effects.

This suggests non-monetary incentives can be used to increase an agent's effort. A principal in a moral hazard problem may be able to extract more effort out of agents who experienced the feeling of being behind others in prior periods. Thus, a principal may consider placing a mixture of workers in close proximity or in contact who would be likely to fail and succeed, rather than letting them perform their tasks in isolation. This way, in future tasks, the principal will be able to extract greater effort out of the agents who failed when their partner succeeded. To anticipate the net effect, we need to know how such matching affects the combined effort of agents who are likely to succeed and who are likely to fail, as compared to the counterfactual where they act in isolation. Our earlier results suggest that subjects who succeeded downward revise their effort in similar amounts when their partner failed as compared to when in isolation (-0.95 and -0.58 respectively). However, there is more meaningful difference in the effort increase of subjects who failed and see a partner succeed (2.41) as compared those who failed in isolation (1.38). The net result is that the average period 2 effort of subjects who receive an outcome incongruent with their partner's in the Partner Treatment is 10.45, or 8% larger than the average period 2 effort of failed and successful subjects in the Individual Treatment (9.67).[33] Therefore, a principal could extract higher effort by putting in contact agents who are unsuccessful with those who are successful. This is consistent with the findings of Bandiera et al. (2010) in a study of the assignment of high and low ability fruit pickers to the fields of a farm.[34]

Note that our setup provides subjects individual monetary incentives rather than relative payment incentives (as in the tournaments literature) to distinguish social comparisons from competition for a monetary prize. Hence, we show that agents may choose greater effort in individually performed and rewarded tasks when they observe each other's outcomes.

Finally, note that the behindness aversion observed is independent of a reputational concern, which could provide an added non-monetary incentive. In the experiments, subjects are matched with a new partner in each period. Hence, revision of effort when behind a partner is not to prove to the partner that one can be successful too, as there is a different partner in the next period. We argue that being behind is enough to activate an agent's inequity aversion concerns and greater effort — out of the agent's desire to avoid that experience again. We conjecture that repeated interaction with the same peer would amplify the effects reported.

VII CONCLUSION

We conducted laboratory experiments to study the effects of feedback on one's own and a peer's performance on the motivation for taking a costly action. We showed that behindness aversion plays a role in the effort choice of subjects. Those who felt behind exerted higher effort while those who were ahead did not diminish their effort much. Subjects do not seem

[33]This is a rough comparison of average effort in period 2 of these two groups as it does not control for subjects' effort in period 1.

[34]This study focuses on the productivity impact of social comparisons between friends versus non-friends. On the other hand, we study the impact on effort revision of social comparison versus the case of no social comparison.

to anticipate such social comparisons, but it affects their future effort once they experience being behind. Furthermore, this effect is more pronounced for female subjects than male subjects.

Our findings have direct implications for non-monetary incentives in labor settings. For example, a student who learns about the high grades of fellow students may be motivated to work harder even when grading is not relative. A farmer may be motivated to work harder after observing the flourishing harvests of other farmers. Our results also have implications for settings beyond the labor environments we focused on here. For example, they can be applied to the purchase of lottery tickets or any risky investment, suggesting that observing others win may motivate the purchase of more lottery tickets.

Additionally, while our main research question is about social comparisons, our Individual Treatment provides interesting findings as well. We find that subjects who work in isolation revise their effort choices based on the feedback they received on their own past performances. Failed subjects (of any gender) increased their effort and successful female subjects decreased their effort. This result contributes to the literature about the impact of performance feedback (see, for example, Akın and Karagözoğlu (2017); Bandiera et al. (2015); Eriksson et al. (2009); Hannan et al. (2008); Ockenfels et al. (2014)).

References

Akin, Z. and E. Karagözoğlu (2017): "The Role of Goals and Feedback in Incentivizing Performance," *Managerial and Decision Economics*, 38, 193–211.

Anderson, D. C., C. R. Crowell, S. S. Sponsel, M. Clarke, and J. Brence (1983): "Behavior Management in the Public Accommodations Industry: A Three-project Demonstration," *Journal of Organizational Behavior Management*, 4, 33–66.

Andrade, E. B. and G. Iyer (2009): "Planned Versus Actual Betting in Sequential Gambles," *Journal of Marketing Research*, 46, 372–383.

Avrahami, J. and Y. Kareev (2011): "The Role of Impulses in Shaping Decisions," *Journal of Behavioral Decision Making*, 24, 515–529.

Bandiera, O., I. Barankay, and I. Rasul (2010): "Social Incentives in the Workplace," *The Review of Economic Studies*, 77, 417–458.

Bandiera, O., V. Larcinese, and I. Rasul (2015): "Blissful Ignorance? A Natural Experiment on the Effect of Feedback on Students' Performance," *Labour Economics*, 34, 13–25.

Bartling, B. (2011): "Relative Performance or Team Evaluation? Optimal Contracts for Other-regarding Agents," *Journal of Economic Behavior & Organization*, 79, 183–193.

Bartling, B. and F. A. Von Siemens (2010): "The Intensity of Incentives in Firms and Markets: Moral Hazard with Envious Agents," *Labour Economics*, 17, 598–607.

Bénabou, R. and J. Tirole (2006): "Incentives and Prosocial Behavior," *American Economic Review*, 96, 1652–1678.

Berlin, N. and M.-P. Dargnies (2016): "Gender Differences in Reactions to Feedback and Willingness to Compete," *Journal of Economic Behavior & Organization*, 130, 320–336.

Bolton, G. E. and A. Ockenfels (2000): "ERC: A Theory of Equity, Reciprocity, and Competition," *American Economic Review*, 90, 166–193.

Bolton, P. and M. Dewatripont (2005): *Contract Theory*, MIT press.

Breza, E., S. Kaur, and Y. Shamdasani (2018): "The Morale Effects of Pay Inequality," *The Quarterly Journal of Economics*, 133, 611–663.

Brock, J. M., A. Lange, and E. Y. Ozbay (2013): "Dictating the Risk: Experimental Evidence on Giving in Risky Environments," *American Economic Review*, 103, 415–37.

Bull, C., A. Schotter, and K. Weigelt (1987): "Tournaments and Piece Rates: An Experimental Study," *Journal of Political Economy*, 95, 1–33.

Charness, G., U. Gneezy, and A. Imas (2013a): "Experimental Methods: Eliciting Risk Preferences," *Journal of Economic Behavior & Organization*, 87, 43–51.

Charness, G. and B. Grosskopf (2001): "Relative Payoffs and Happiness: An Experimental Study," *Journal of Economic Behavior & Organization*, 45, 301–328.

Charness, G., D. Masclet, and M. C. Villeval (2013b): "The Dark Side of Competition for

Status," *Management Science*, 60, 38–55.

Charness, G. and M. Rabin (2002): "Understanding Social [references with Simple Tests," *The Quarterly Journal of Economics*, 117, 817–869.

Clark, L. (2010): "Decision-making During Gambling: An Integration of Cognitive and Psychobiological Approaches," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365, 319–330.

Cohn, A., E. Fehr, B. Herrmann, and F. Schneider (2014): "Social Comparison and Effort Provision: Evidence from a Field Experiment," *Journal of the European Economic Association*, 12, 877–898.

Cooper, D. J. and M. Rege (2011): "Misery Loves Company: Social Regret and Social Interaction Effects in Choices Under Risk and Uncertainty," *Games and Economic Behavior*, 73, 91–110.

Croson, R. and U. Gneezy (2009): "Gender Differences in Preferences," *Journal of Economic Literature*, 47, 448–74.

Dave, C., C. C. Eckel, C. A. Johnson, and C. Rojas (2010): "Eliciting Risk Preferences: When is Simple Better?" *Journal of Risk and Uncertainty*, 41, 219–243.

Demougin, D., C. Fluet, and C. Helm (2006): "Output and Wages with Inequality Averse Agents," *Canadian Journal of Economics*, 399–413.

Dillenberger, D. and K. Rozen (2015): "History-dependent Risk Attitude," *Journal of Economic Theory*, 157, 445–477.

Dohmen, T. and A. Falk (2011): "Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender," *American Economic Review*, 101, 556–90.

Dohmen, T., A. Falk, K. Fliessbach, U. Sunde, and B. Weber (2011): "Relative Versus Absolute Income, Joy of Winning, and Gender: Brain Imaging Evidence," *Journal of Public Economics*, 95, 279–285.

Dutcher, G., T. Salmon, and K. J. Saral (2015): "Is 'Real' Effort More Real?" *Working Paper*.

Eckel, C. C. and P. J. Grossman (2002): "Sex Differences and Statistical Stereotyping in Attitudes Toward Financial Risk," *Evolution and Human Behavior*, 23, 281–295.

Engelmann, D. and M. Strobel (2004): "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments," *American Economic Review*, 94, 857–869.

Englmaier, F. and A. Wambach (2010): "Optimal Incentive Contracts under Inequity Aversion," *Games and Economic Behavior*, 69, 312–328.

Eriksson, T., A. Poulsen, and M. C. Villeval (2009): "Feedback and Incentives: Experimental Evidence," *Labour Economics*, 16, 679–688.

Falk, A. and A. Ichino (2006): "Clean Evidence on Peer Effects," *Journal of Labor Economics*, 24, 39–57.

Fehr, E. and K. M. Schmidt (1999): "A Theory of Fairness, Competition, and Cooperation," *The Quarterly Journal of Economics*, 114, 817–868.

Fischbacher, U. (2007): "z-Tree: Zurich Toolbox for Ready-made Economic Experiments," *Experimental Economics*, 10, 171–178.

Frank, R. H. (1985): "The demand for Unobservable and other Nonpositional Goods," *American Economic Review*, 75, 101–116.

Gächter, S., L. Huang, and M. Sefton (2018): "Disappointment Aversion and Social Comparisons in a Real-Effort Competition," *Economic Inquiry*, 56, 1512–1525.

Gill, D. and V. Prowse (2012): "A Structural Analysis of Disappointment Aversion in a Real Effort Competition," *American Economic Review*, 102, 469–503.

——— (2014): "Gender Differences and Dynamics in Competition: The Role of Luck," *Quantitative Economics*, 5, 351–376.

Guryan, J., K. Kroft, and M. J. Notowidigdo (2009): "Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments," *American Economic Journal: Applied Economics*, 1, 34–68.

Hannan, R. L., R. Krishnan, and A. H. Newman (2008): "The Effects of Disseminating Relative Performance Feedback in Tournament and Individual Performance Compensation Plans," *The Accounting Review*, 83, 893–913.

Hart, E., Y. Kareev, and J. Avrahami (2016): "Good Times, Bad Times: Reversal of Risk Preferences," *Decision*, 3, 132.

Imas, A. (2016): "The Realization Effect: Risk-taking After Realized Versus Paper Losses," *American Economic Review*, 106, 2086–2109.

Itoh, H. (2004): "Moral Hazard and Other-regarding Preferences," *The Japanese Economic Review*, 55, 18–45.

Kareev, Y., J. Avrahami, and K. Fiedler (2014): "Strategic Interactions, Affective Reactions, and Fast Adaptations," *Journal of Experimental Psychology: General*, 143, 1112.

Kuhnen, C. M. and A. Tymula (2012): "Feedback, Self-esteem, and Performance in Organizations," *Management Science*, 58, 94–113.

Lahno, A. M. and M. Serra-Garcia (2015): "Peer Effects in Risk Taking: Envy or Conformity?" *Journal of Risk and Uncertainty*, 50, 73–95.

Langer, T. and M. Weber (2008): "Does Commitment or Feedback Influence Myopic Loss Aversion?: An Experimental Analysis," *Journal of Economic Behavior & Organization*, 67, 810–819.

Lazear, E. P. and S. Rosen (1981): "Rank-order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, 89, 841–864.

Linde, J. and J. Sonnemans (2012): "Social Comparison and Risky Choices," *Journal of Risk and Uncertainty*, 44, 45–72.

——— (2015): "Decisions Under Risk in a Social and Individual Context: The Limits of Social

Preferences?" *Journal of Behavioral and Experimental Economics*, 56, 62–71.

List, J. A., C. D. Bailey, P. J. Euzent, and T. L. Martin (2001): "Academic Economists Behaving Badly? A Survey on Three Areas of Unethical Behavior," *Economic Inquiry*, 39, 162–170.

Liu, Y.-J., C.-L. Tsai, M.-C. Wang, and N. Zhu (2010): "Prior Consequences and Subsequent Risk Taking: New Field Evidence from the Taiwan Futures Exchange," *Management Science*, 56, 606–620.

MacCrimmon, K. R. and D. M. Messick (1976): "A Framework for Social Motives," *Behavioral Science*, 21, 86–100.

Mas, A. and E. Moretti (2009): "Peers at Work," *American Economic Review*, 99, 112–45.

Neilson, W. S. and J. Stowe (2010): "Piece-rate Contracts for Other-regarding Workers," *Economic Inquiry*, 48, 575–586.

Nielsen, K. (2019): "Dynamic Risk Preferences Under Realized and Paper Outcomes," *Journal of Economic Behavior & Organization*, 161, 68–78.

Ockenfels, A., D. Sliwka, and P. Werner (2014): "Bonus Payments and Reference Point Violations," *Management Science*, 61, 1496–1513.

Rohde, I. M. and K. I. Rohde (2011): "Risk Attitudes in a Social Context," *Journal of Risk and Uncertainty*, 43, 205–225.

Saito, K. (2013): "Social Preferences Under Risk: Equality of Opportunity Versus Equality of Outcome," *American Economic Review*, 103, 3084–3101.

Schneider, S. L., S. Kauffman, and A. Ranieri (2016): "The Effects of Surrounding Positive and Negative Experiences on Risk Taking," *Judgment and Decision Making*, 11, 424.

Schwerter, F. (2019): "Social Reference Points and Risk Taking," Working Paper.

Selten, R. and J. Buchta (1998): "Experimental Sealed Bid First Price Auctions with Directly Observed Bid Functions," DV Budescu, I. Erev, R. Zwick, eds. Games and Human Behavior: Essays in Honor of Amon Rapoport.

Selten, R. and R. Stoecker (1986): "End Behavior in Sequences of Finite Prisoner's Dilemma Supergames A Learning Theory Approach," *Journal of Economic Behavior & Organization*, 7, 47–70.

Shiv, B., G. Loewenstein, and A. Bechara (2006): "The Dark Side of Emotion in Decision-making: When Individuals with Decreased Emotional Reactions Make More Advantageous Decisions," *Cognitive Brain Research*, 23, 85–92.

Sjöström, T., L. Ulku, and R. Vadovic (2019): "Free to Choose: Testing the Pure Motivation Effect of Autonomous Choice," Working Paper.

Smith, G., M. Levere, and R. Kurtzman (2009): "Poker Player Behavior after Big Wins and Big Losses," *Management Science*, 55, 1547–1555.

Thaler, R. H. and E. J. Johnson (1990): "Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice," *Management Science*, 36, 643–660.

Tserenjigmid, G. (2019): "History-Dependent Risk Aversion, the Reinforcement Effect, and Dynamic Monotonicity," Working Paper.

Verbruggen, F., C. D. Chambers, N. S. Lawrence, and I. P. McLaren (2017): "Winning and Losing: Effects on Impulsive Action," *Journal of Experimental Psychology: Human Perception and Performance*, 43, 147.

Xia, L. (2018): "Luck or Skill: How Women and Men Attribute Successes and Failures," Working Paper.

A PROOFS

A.I First Order Conditions

Here, we show that if $\lambda \leq 0$, then $e^{Partner} > e^{Ind}$ where e^{Ind} and $e^{Partner}$ are the (symmetric) equilibria of model (1) and model (3), respectively. The first order conditions of problems (1) and (3) respectively are:

$$\begin{aligned} p^{'}(e_i)\ u(f - ce_i + b) - cp(e_i)\ u^{'}(f - ce_i + b) - (1 - p(e_i))\ c\, u^{'}(f - ce_i) \\ - p^{'}(e_i)\ u(f - ce_i) = 0 \end{aligned} \tag{4}$$

$$\begin{aligned} p^{'}(e_i)\ u(f - ce_i + b) - cp(e_i)\ u^{'}(f - ce_i + b) - (1 - p(e_i))\ c\, u^{'}(f - ce_i) \\ - p^{'}(e_i)\ u(f - ce_i) + p^{'}(e_i)\ p(e_j)\, v\ (b) - p^{'}(e_i)\left(1 - p(e_j)\right)\lambda v(b) = 0 \end{aligned} \tag{5}$$

Note that the difference between the left-hand sides of the two F.O.C.s is the last two terms of the F.O.C. (5). That is, the marginal expected disutility from having a different gross payment than a partner:

$$K(e_i, e_j) = p^{'}(e_i)\ p(e_j)\ v(b) - p^{'}(e_i)\left(1 - p(e_j)\right)\lambda v(b) \tag{6}$$

If $\lambda \leq 0$, then $K\left(e^{Ind}, e^{Ind}\right) > 0$. Hence, the solution of model (1) makes the F.O.C. of model (3) positive. That implies and $e^{Partner} > e^{Ind}$.

A.II Independence of Periods

Here, we show that a selfish utility maximizer would choose the same effort in each period when she is paid for one randomly selected period. Consider the decision problem of such an agent in the second period:

$$\max_{e_X} \frac{1}{2} u(\pi_1(X)) + \frac{1}{2}\left[p(e_X)\, u(f - ce_X + b) + (1 - p(e_X))\, u(f - ce_X)\right] \tag{7}$$

where $X \in \{Success, Failure\}$ denotes the realization of her first period outcome; and based on this outcome, the first term is the utility from the net payoff π_1 earned in the first period which realizes with 50% chance; the second term is the expected utility in the second period where e_X denotes the effort choice in the second period after realization of first period outcome of X. Note that the solution to problem (7) above is independent of X, the agent's realized outcome in the first period. Moreover, the effort level that solves problem (7) also solves problem (1). The equivalence between problems (1) and (7) would be violated if the agent was instead paid for her performance in every period (unless she is risk neutral.)

B ADDITIONAL TABLES AND FIGURES

Table B.1: Treatment Balance

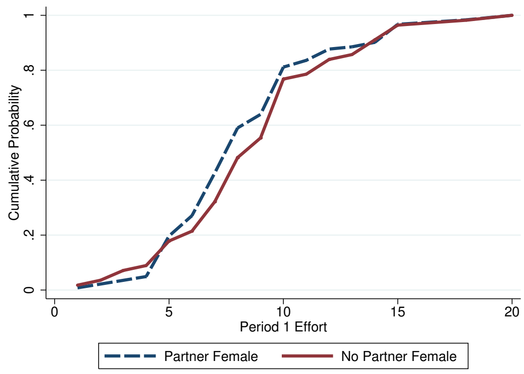
	Individual Treatment			Partner Treatment			
	n	mean	sd	n	mean	sd	Diff
Female	244	0.46	0.50	522	0.47	0.50	0.008
Age	244	19.42	1.43	522	19.93	2.08	0.509**
SAT	194	1590.62	328.30	412	1628.10	326.45	37.479
ACT	100	30.44	3.46	214	30.24	3.73	-0.197
GPA	244	3.32	0.78	520	3.29	0.61	-0.035
Math or Econ Major	244	0.16	0.36	522	0.15	0.36	-0.002
Job Experience	244	0.87	0.34	522	0.92	0.28	0.047
Currently Employed	244	0.34	0.48	522	0.43	0.50	0.085
Low Risk Aversion	244	0.57	0.50	522	0.59	0.49	0.024
Efficiency Minded	244	0.05	0.22	522	0.05	0.23	0.004
Selfish	244	0.61	0.49	522	0.63	0.48	0.014
Relative Grading Preference	244	0.34	0.48	522	0.41	0.49	0.062
Team Preference	244	0.52	0.50	522	0.56	0.50	0.039

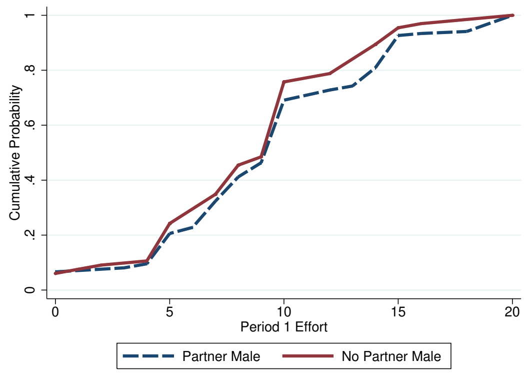
*** p<0.01, ** p<0.05, * p<0.1
n is the number of observations. Difference is the Partner Treatment mean minus the Individual Treatment mean, and p-values report whether that difference is statistically different from zero.

Demographic Variables: Female is a dummy equal to 1 if the subject self-reported as female. Math or Econ major is a dummy equal to 1 if the subject self-reported as either a math or economics major. Job experience is a dummy equal to 1 if the subject answered yes to the question "have you ever been employed in a paid job?" Currently employed is a dummy equal to 1 if the subject self-reported as currently employed either part-time or full-time.

Behavioral Variables: Low Risk Aversion is a dummy equal to 1 if the gamble chosen is either gamble 4, 5 or 6. Efficiency Minded is a dummy equal to 1 if the subject chose the efficient outcome in the fairness task. Selfish is a dummy equal to 1 if the subject chose the selfish outcome in the fairness task. Relative Grading Preference is a dummy equal to 1 if the subject is indifferent to, or strictly prefers being graded relatively to individually on a homework assignment. Team preference is a dummy equal to 1 if the subject is indifferent to or prefers working on homework on a team as compared to individually.

Figure B.1: CDFs of Period 1 Effort





Females: Fail to reject the equality of the distributions in Individual vs. Partner Treatments (p=0.771).
Males: Fail to reject the equality of the distributions in Individual vs. Partner Treatments (p=0.870).

Table B.2: Two-Limit Tobit of Period 1 Effort

	(1) All	(2) Female	(3) Male
Partner Treatment	0.192	-0.399	0.611
	(0.493)	(0.585)	(0.780)
Low Risk Aversion	0.555	-0.699	1.557*
	(0.475)	(0.546)	(0.797)
Efficiency Minded	0.619	-0.661	1.686
	(1.083)	(1.226)	(1.792)
Selfish	-0.098	0.182	-0.563
	(0.507)	(0.583)	(0.867)
GPA	-0.451	-0.014	-0.689
	(0.344)	(0.398)	(0.571)
Relative Grading Preference	-0.195	-0.210	-0.597
	(0.474)	(0.618)	(0.744)
Constant	10.261***	9.409***	10.821***
	(1.292)	(1.493)	(0.195)
Observations	382	177	202
Log-Likelihood	-1076	-475.2	-577.3

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Censored from below at 0 and above at 20. Partner treatment is a dummy equal to 1 in the Partner Treatment. Low risk aversion is a dummy equal to 1 if the gamble chosen is either gamble 4, 5 or 6. Efficiency minded is a dummy equal to 1 if the subject chose the efficient outcome in the fairness task. Selfish is a dummy equal to 1 if the subject chose the selfish outcome in the fairness task. Columns (2) and (3) exclude 3 subjects who self-reported their gender as neither female nor male. These 3 subjects were all in the Partner Treatment.

Table B.3: Individual Treatment, Proportion Revise Effort in Period 2

Period 1 Outcome	Decrease Effort			Increase Effort		
	All	Female	Male	All	Female	Male
Failed	0.14	0.13	0.15	0.48	0.58	0.40
	(0.35)	(0.34)	(0.36)	(0.50)	(0.50)	(0.50)
	n=85	n=38	n= 47	n=85	n=38	n=47
Succeeded	0.49	0.67	0.32	0.16	0.11	0.21
	(0.51)	(0.49)	(0.48)	(0.37)	(0.32)	(0.42)
	n=37	n=18	n=19	n=37	n=18	n=19
P-value	0.000	0.000	0.127	0.001	0.001	0.139

Standard deviations in parentheses. n=number of observations.

Table B.4: Individual Treatment, Probit Regressions

(a) Likelihood Increase Effort in Period 2

	(1) All	(2) All	(3) Female	(4) Female	(5) Male	(6) Male
Failed	0.327***	0.342***	0.516***	0.551***	0.182	0.227
	(0.107)	(0.111)	(0.170)	(0.177)	(0.139)	(0.153)
Period 1 Effort	-0.016	-0.016	-0.022	-0.020	-0.012	-0.015
	(0.011)	(0.011)	(0.017)	(0.018)	(0.014)	(0.015)
Controlling for Heterogeneity		Yes		Yes		Yes
Observations	122	122	56	56	66	66
Baseline Probability	0.384	0.385	0.426	0.428	0.348	0.349
Pseudo R Squared	0.088	0.108	0.183	0.214	0.037	0.094

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(b) Likelihood Decrease Effort in Period 2

	(1) All	(2) All	(3) Female	(4) Female	(5) Male	(6) Male
Success	0.284***	0.286***	0.466***	0.455***	0.139	0.109
	(0.078)	(0.083)	(0.134)	(0.139)	(0.099)	(1.578)
Period 1 Effort	0.017*	0.017*	0.027*	0.032*	0.011	0.009
	(0.009)	(0.009)	(0.016)	(0.017)	(0.011)	(0.123)
Controlling for Heterogeneity		Yes		Yes		Yes
Observations	122	122	56	56	66	66
Baseline Probability	0.244	0.243	0.302	0.301	0.196	0.196
Pseudo R Squared	0.142	0.148	0.289	0.302	0.049	0.077

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Probit marginal effects reported (calculated at means of independent variables). In Table B.4a, dependent variable is a dummy equal to 1 if subject increased effort in period 2. In Table B.4b, dependent variable is a dummy equal to 1 if subject increased effort in period 2. Failed (Succeeded) is a dummy equal to 1 if the subject failed (succeeded) in period 1. Controlling for heterogeneity includes controls for risk preferences (dummy equal to 1 if the gamble chosen is either gamble 4, 5 or 6), fairness preferences (dummy for 2 of the 3 possible allocations in the fairness task), a linear term for GPA, and relative grading preferences (dummy equal to 1 if the subject is indifferent to, or strictly prefers being graded relatively to individually on a homework assignment).

Table B.5: Individual Treatment, Mean Period 2 Effort

Period 1 Outcome	All	Female	Male
Failed	9.81	9.97	9.68
	(4.77)	(4.49)	(5.04)
	n=85	n=38	n=47
Succeeded	9.35	8.89	9.79
	(4.47)	(3.76)	(5.12)
	n=37	n=18	n=19

Standard deviation in parentheses. n = number of observations.

Table B.6: Individual Treatment, Two-Limit Tobit of Period 2 Effort

	(1) All	(2) All	(3) Female	(4) Female	(5) Male	(6) Male
Failed	1.745**	1.857**	2.400**	2.503**	1.039	1.469
	(0.795)	(0.805)	(1.052)	(1.071)	(1.179)	(1.207)
Period 1 Effort	0.754***	0.761***	0.657***	0.648***	0.825***	0.832***
	(0.090)	(0.090)	(0.131)	(0.133)	(0.125)	(0.124)
Constant	1.654	0.310	2.070	-0.472	1.518	2.295
	(1.129)	(2.139)	(1.588)	(2.834)	(1.602)	(3.488)
Controlling for Heterogeneity		Yes		Yes		Yes
Observations	122	122	56	56	66	66
Log-Likelihood	-325.8	-323.2	-148.4	-147.1	-175.7	-172.9

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Dependent variable is period 2 effort. Censored from below at 0 and above at 20. Failed is a dummy equal to one if the subject failed in period 1. Controlling for heterogeneity includes controls for risk preferences (dummy equal to 1 if the gamble chosen is either gamble 4, 5 or 6), fairness preferences (dummy for 2 of the 3 possible allocations in the fairness task), a linear term for GPA, and relative grading preferences (dummy equal to 1 if the subject is indifferent to, or strictly prefers being graded relatively to individually on a homework assignment).

Table B.7: Partner Treatment, Two-Limit Tobit of Period 2 Effort

	(1) All	(2) All	(3) Female	(4) Female	(5) Male	(6) Male
Failed	1.689***	1.898***	2.040**	2.324***	1.488	2.000**
	(0.649)	(0.655)	(0.867)	(0.876)	(0.969)	(0.965)
Period 1 Effort	0.742***	0.747***	0.713***	0.733***	0.758***	0.771***
	(0.076)	(0.076)	(0.117)	(0.116)	(0.105)	(0.102)
Constant	1.963**	0.224	2.127	-0.814	1.708	0.551
	(0.957)	(1.929)	(1.341)	(2.737)	(1.380)	(2.692)
Controlling for Heterogeneity		Yes		Yes		Yes
Observations	261	260	122	121	136	136
Log-Likelihood	-722.5	-715.5	-337.2	-332.5	-375.2	-369.2

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Dependent variable is period 2 effort. Censored from below at 0 and above at 20. Failed is a dummy equal to one if the subject failed in period 1, regardless of their partner's outcome. Controlling for heterogeneity includes controls for risk preferences (dummy equal to 1 if the gamble chosen is either gamble 4, 5 or 6), fairness preferences (dummy for 2 of the 3 possible allocations in the fairness task), a linear term for GPA, and relative grading preferences (dummy equal to 1 if the subject is indifferent to, or strictly prefers being graded relatively to individually on a homework assignment). Female and male columns exclude 3 subjects who self-reported their gender as neither female nor male. Column (4) excludes 1 subject who did not report a GPA.

Table B.8: Partner Treatment, Mean (Period 2 Effort – Period 1 Effort)

Period 1 Outcome	All	Female	Male
Failed	1.49***	1.71***	1.29**
	(4.97)	(4.66)	(5.35)
	n=165	n=84	n=79
Succeeded	-0.98***	-0.74	-1.25**
	(3.62)	(2.86)	(4.01)
	n=96	n=38	n=57
P-values	0.000	0.003	0.003

Stars indicate whether mean is statistically different from zero in a two-tailed t-test. Standard deviations in parentheses. n=number of observations. Bottom row reports p-values from a two-tailed t-test comparing the mean effort change of subjects who failed and succeeded in a given column. Female and male columns exclude 3 subjects who self-reported their gender as neither female nor male.

Table B.9: Partner Treatment, Proportion Revise Effort in Period 2

(a)

Period 1 Outcome	Decrease Effort			Increase Effort		
	All	Female	Male	All	Female	Male
Failed	0.21	0.25	0.18	0.50	0.63	0.37
	(0.41)	(0.44)	(0.38)	(0.50)	(0.49)	(0.49)
	n=165	n=84	n=79	n=165	n=84	n=79
Succeeded	0.33	0.39	0.30	0.21	0.21	0.19
	(0.47)	(0.50)	(0.46)	(0.41)	(0.41)	(0.40)
	n=96	n=38	n=57	n=96	n=38	n=57
P-values	0.031	0.106	0.099	0.000	0.000	0.028

P-values from a two-tailed t-test comparing the mean of those who failed and succeeded in relevant column. Female and male columns exclude 3 subjects who self-reported their gender as neither female nor male.

(b)

Period 1 Outcome (Self, Partner)	Decrease Effort			Increase Effort		
	All	Female	Male	All	Female	Male
(Fail, Fail)	0.24	0.30	0.17	0.45	0.59	0.29
	(0.43)	(0.46)	(0.38)	(0.50)	(0.50)	(0.46)
	n=99	n=56	n=41	n=99	n=56	n=41
(Fail, Success)	0.17	0.14	0.18	0.56	0.71	0.45
	(0.38)	(0.36)	(0.39)	(0.50)	(0.46)	(0.50)
	n=66	n=28	n=38	n=66	n=28	n=38
(Success, Fail)	0.32	0.41	0.26	0.26	0.22	0.26
	(0.47)	(0.50)	(0.45)	(0.44)	(0.42)	(0.45)
	n=66	n=27	n=38	n=66	n=27	n=38
(Success, Success)	0.37	0.36	0.37	0.10	0.18	0.05
	(0.49)	(0.50)	(0.50)	(0.31)	(0.40)	(0.23)
	n=30	n=11	n=19	n=30	n=11	n=19

Standard deviation in parentheses. n=number of observations. Female and male columns exclude 3 subjects who self-reported their gender as neither female nor male. P-values from a two-tailed t-test comparing the likelihood of increasing effort for subjects with outcome (Fail, Fail) and (Fail, Success): all genders pooled (0.184), females (0.269) and males (0.159). P-values from a two-tailed t-test comparing the likelihood of increasing effort for subjects with outcome (Success, Fail) and (Success, Success): all genders pooled (0.645), females (0.811) and males (0.422).

Table B.10: Comparing Effect of Own Failure in Partner and Individual Treatments, Two-Limit Tobit of Period 2 Effort

	(1)	(2)	(3)	(4)
	Individual Treatment or Partner Succeeded		Individual Treatment or Partner Failed	
Failed	1.934**	1.856**	1.416**	1.321**
	(0.764)	(0.775)	(0.632)	(0.638)
Failed x Partner Treat	1.258*	1.333*	-0.564	-0.576
	(0.752)	(0.767)	(0.627)	(0.630)
Period 1 Effort	0.696***	0.698***	0.779***	0.786***
	(0.078)	(0.078)	(0.064)	(0.064)
Constant	1.943*	1.924	1.766***	-0.350
	(1.004)	(2.016)	(0.783)	(1.510)
Controlling for Heterogeneity		Yes		Yes
Observations	218	218	287	286
Log-Likelihood	-600	-599.5	-773.3	-769.1

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Columns (1) and (2) include all subjects in the individual treatment, and subjects in the partner treatment whose partner succeeded in the first round. Columns (3) and (4) include all subjects in the individual treatment, and subjects in the partner treatment whose partner failed in the first round. Dependent variable is period 2 effort. Censored from below at 0 and above at 20. Failed is a dummy equal to one if the subject failed in period 1. Controlling for heterogeneity includes controls for risk preferences (dummy equal to 1 if the gamble chosen is either gamble 4, 5 or 6), fairness preferences (dummy for 2 of the 3 possible allocations in the fairness task), a linear term for GPA, and relative grading preferences (dummy equal to 1 if the subject is indifferent to, or strictly prefers being graded relatively to individually on a homework assignment). Female and male columns exclude 3 subjects who self-reported their gender as neither female nor male. Column (4) excludes 1 subject who did not report a GPA.

C INSTRUCTIONS

Individual Treatment, Part 1

Instructions

Welcome and thank you for coming today to participate in this experiment. This is an experiment in decision-making. You will receive $7 participation fee if you complete the session. In addition to that if you follow the instructions and are careful with your decisions, you can earn a significant amount of money, which will be paid to you at the end of the session.

During the experiment it is important that you do not talk to any other participants. Please either turn off your cell phones or put them on silent. If you have a question, please raise your hand, and the experimenter will answer your question. Failure to comply with these instructions means that you will be asked to leave the experiment and all your earnings will be forfeited.

The experiment will last about 60 minutes. The experiment consists of two paying parts. This is the instructions for Part 1.

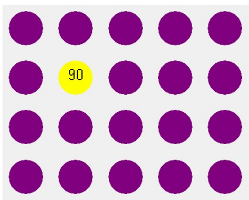
Your earnings in both parts will be calculated in Experimental Currency Units (ECUs). At the end of today's session, all your earnings will be converted to US dollars at a rate of

15 ECUs = $1.

Part 1 consists of 2 identical decision rounds.

Projects

You will begin each round with 60 ECUs. In each round, you will be given a project. The project can succeed or fail. A project is represented by a box with 20 balls in it. One of the balls is yellow and the rest are purple, as in the example below.



The project succeeds if the yellow ball is randomly drawn from the box. Note that the yellow ball has an equal chance of being drawn in a given try. You will decide how many times you want to try to make your project succeed. Each try costs you 3 ECUs. For each try, the computer will randomly draw one of the balls from the box.

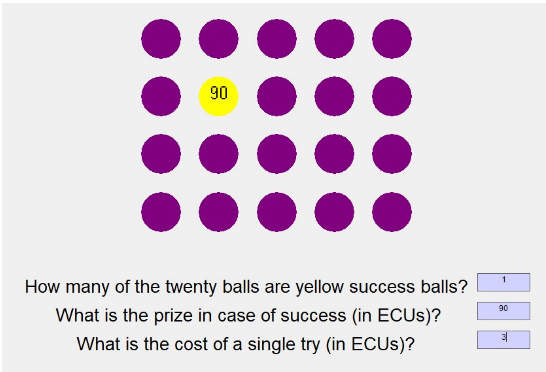
1

- If the yellow ball is selected on one or more tries, then a project succeeds. The prize for success is 90 ECUs as written on the face of the yellow ball.
- If a purple ball is selected on each and every try, then a project fails. The prize for failure is 0 ECU.

After each try, the computer replaces the ball drawn from the box with an identical colored ball. Thus, the chance of drawing the yellow ball on any given draw is the same as that in any other draw. The computer continues drawing balls until your selected number of tries is exhausted. Note that your project succeeds if at least one of the balls drawn on those tries is yellow. In other words, your project succeeds and you earn the 90 ECUs prize if the computer draws the yellow ball in one or more tries.

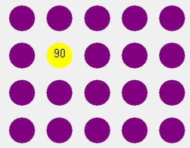
Sequence of Actions in a Round

1. First, the computer will display the project, as in the example screen below. In the fields provided below the box, you will be asked to type in the basic information about the project:
    a. Number of yellow success balls (always equals 1).
    b. Prize in case of success (always equals 90 ECUs).
    c. Cost of a single try (always equals 3 ECUs).



2. Next, you will choose your number of tries. You may choose any number from {0,1,2,3,...,20}. For every possible number of tries, the screen displays:
    a. The total cost of those tries: 3 x Number of Tries.
    b. The probability of the project's success.

See the example screen below.



Endowment: 60 ECUs

Number of Tries:	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Cost (ECUs):	0	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54	57	60
Probability of Success:	0.00%	5.00%	9.75%	14.26%	18.55%	22.62%	26.49%	30.17%	33.66%	36.98%	40.13%	43.12%	45.96%	48.67%	51.23%	53.67%	55.99%	58.19%	60.28%	62.26%	64.15%

Please Select Your Number of Tries: 6

In this example, the participant happens to have selected 6 tries.

3. After you've made your decision, the computer will display your selected number of tries and the outcome of your project. An example screen is shown below.

   In this example, the participant selects 6 tries and the project is successful. This means that the yellow ball is drawn in at least one of her 6 tries. Hence, she receives the prize of 90 ECUs, her 6 tries cost her 18 ECUs, and together with her initial endowment of 60 ECUs, her net payoff for the current round is 132 ECUs (60 + 90 - 18).

Number of tries chosen	**6**
Did the project succeed?	**YES**
Prize received (ECUs)	90
Cost of tries (ECUs)	18
Net Payoff (ECUs) (Endowment + Prize - Cost)	132

4. Finally, you will move to the next round where you will encounter a new project.

Final Earnings

Once the experiment is finished, the computer will randomly pick 1 round out of the 2 rounds that you completed. The earnings you made on that round will be your earnings for Part 1 of the experiment. Hence, you should make a careful decision in each round because it might be a paying round.

Are there any questions?

Partner Treatment, Part 1

Instructions

Welcome and thank you for coming today to participate in this experiment. This is an experiment in decision-making. You will receive $7 participation fee if you complete the session. In addition to that if you follow the instructions and are careful with your decisions, you can earn a significant amount of money, which will be paid to you at the end of the session.

During the experiment it is important that you do not talk to any other participants. Please either turn off your cell phones or put them on silent. If you have a question, please raise your hand, and the experimenter will answer your question. Failure to comply with these instructions means that you will be asked to leave the experiment and all your earnings will be forfeited.

The experiment will last about 60 minutes. The experiment consists of two paying parts. This is the instructions for Part 1.

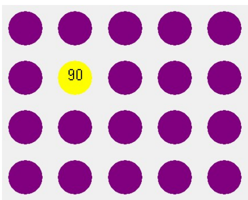
Your earnings in both parts will be calculated in Experimental Currency Units (ECUs). At the end of today's session, all your earnings will be converted to US dollars at a rate of

15 ECUs = $1.

Part 1 consists of 2 identical decision rounds. At the beginning of each round, you will be randomly paired with another participant in the room. You will be paired with a different participant in each round. These matchings are anonymous and you will not know the identities of the participants you are partnered with.

Projects

You and your partner will each begin a round with 60 ECUs. Furthermore, in each round, you and your partner will be given identical individual projects. Each project can succeed or fail. A project is represented by a box with 20 balls in it. One of the balls is yellow and the rest are purple, as in the example below.



The project succeeds if the yellow ball is randomly drawn from the box. Note that the yellow ball has an equal chance of being drawn in a given try. You will decide how many times you want to try

1

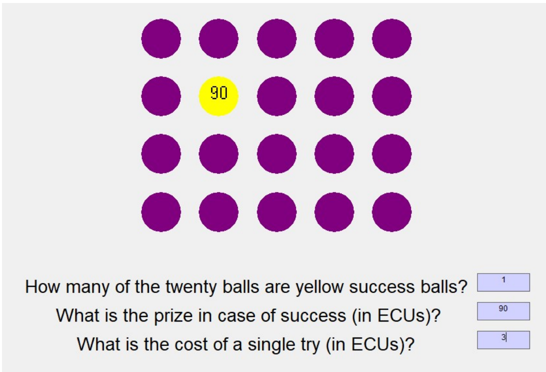
to make your project succeed. Each try costs you 3 ECUs. For each try, the computer will randomly draw one of the balls from the box.

- If the yellow ball is selected on one or more tries, then a project succeeds. The prize for success is 90 ECUs as written on the face of the yellow ball.
- If a purple ball is selected on each and every try, then a project fails. The prize for failure is 0 ECU.

After each try, the computer replaces the ball drawn from the box with an identical colored ball. Thus, the chance of drawing the yellow ball on any given draw is the same as that in any other draw. The computer continues drawing balls until your selected number of tries is exhausted. Note that your project succeeds if at least one of the balls drawn on those tries is yellow. In other words, your project succeeds and you earn the 90 ECUs prize if the computer draws the yellow ball in one or more tries.

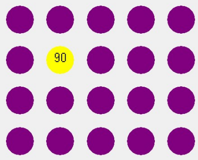
Sequence of Actions in a Round

1. First, the computer will display the project, as in the example screen below. In the fields provided below the box, you will be asked to type in the basic information about the project:
    a. Number of yellow success balls (always equals 1).
    b. Prize in case of success (always equals 90 ECUs).
    c. Cost of a single try (always equals 3 ECUs).



2. Next, you will choose your number of tries. You may choose any number from {0,1,2,3,...,20}. For every possible number of tries, the screen displays:
    a. The total cost of those tries: 3 x Number of Tries.
    b. The probability of the project's success.

See the example screen below.



Endowment: 60 ECUs

Number of Tries:	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Cost (ECUs):	0	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54	57	60
Probability of Success:	0.00%	5.00%	9.75%	14.26%	18.55%	22.62%	26.49%	30.17%	33.66%	36.98%	40.13%	43.12%	45.96%	48.67%	51.23%	53.67%	55.99%	58.19%	60.28%	62.26%	64.15%

Please Select Your Number of Tries: 6

In this example, the participant happens to have selected 6 tries.

3. After you and your partner have made your decisions, the computer will display your selected number of tries, the outcome of your project and the outcome of your partner's project. The screen will not display your partner's selected number of tries. An example screen is shown below. A similar screen will be shown on your partner's display.

   In this example, the participant selects 6 tries and the project is successful. This means that a yellow ball is drawn in at least one of her 6 tries. Hence, she receives the prize of 90 ECUs, her 6 tries cost her 18 ECUs, and together with her initial endowment of 60 ECUs, her net payoff for the current round is 132 ECUs (60 + 90 - 18).

   By contrast, the participant's partner's project is unsuccessful. This means that a purple ball is drawn in all of his draws. Note that the participant learns the outcome of his/her partner's project, but not the number of tries the partner selected.

	You	**Your Partner**
Number of tries chosen	**6**	**-----**
Did the project succeed?	**YES**	**NO**
Prize received (ECUs)	90	0
Cost of tries (ECUs)	18	-----
Net Payoff (ECUs) (Endowment + Prize - Cost)	132	-----

3

4. Finally, you will move to the next round where you will be randomly matched with a new participant in the room and encounter a new project.

Final Earnings

Once the experiment is finished, the computer will randomly pick 1 round out of the 2 rounds that you completed. The earnings you made on that round will be your earnings for Part 1 of the experiment. Hence, you should make a careful decision in each round because it might be a paying round.

Are there any questions?

Part 2 Instructions

This part of the experiment consists of two activities. Your earnings in each activity depend on your decisions and also on chance. Once you finish an activity you will not be able to go back.

The actual decision you make in each activity is up to you. There is no right or wrong answer. Just choose the one you like best.

Activity 1

In activity 1, you are asked to choose one of the following 6 gambles. The computer will flip a coin to determine your payoff from the gamble chosen as indicated in the below table. The coin turns up heads or tails with 50% chance each.

Gamble	Tails Payoff (ECUs)	Heads Payoff (ECUs)
1	28	28
2	24	36
3	20	44
4	16	52
5	12	60
6	2	70

1

Activity 2

In Activity 2, you will be randomly matched with one other participant in the room. The matching is anonymous and you will not know the identity of your partner. You are asked to choose one of the following three allocations of ECUs between you and your partner:

Allocation	You Receive (ECUs)	Your Partner Receives (ECUs)
1	35	35
2	20	60
3	50	10

Similarly, your partner will choose one of these three options. After both of you make your choices, the computer will randomly select one of you with equal chances and implement that person's decision.

For example, suppose you picked option 2 and your partner picked option 3. If you are randomly selected by the computer, your chosen option (option 2) will be implemented. Hence, you will receive 20 ECUs and your partner will receive 60 ECUs. However, if your partner's decision is randomly selected to be implemented, then you will receive 10 ECUs and your partner will receive 50 ECUs. The example described here is only for demonstration purposes, and is not meant to suggest how you should make your decision in any way.

Final Earnings

At the end of the experiment, the computer will randomly pick 1 activity out of the 2 activities that you completed in Part 2. You will receive your earnings from the randomly selected activity in Part 2 in addition to the randomly selected round in Part 1. Hence, you should make careful decisions in each activity because it might be a paying activity.

Are there any questions?